Principles and **E**lements

of

POWER ELECTRONICS

Devices, Drivers, Applications, and Passive Components

Barry W Williams
B.Sc., Dipl.Eng., B.Eng., M.Eng.Sc., Ph.D., D.I.C.

Professor of Electrical Engineering
University of Strathclyde
Glasgow

Published by *Barry W Williams*
ISBN 978-0-9553384-0-3

bww

Table of Contents

7

247

Driving Transistors and Thyristors

8

267

Protecting Diodes, Transistors, and Thyristors

9

Switching-aid Circuits with Energy Recovery

10

Device Series and Parallel Operation, Protection, and Interference

11

Naturally Commutating AC to DC Converters - Uncontrolled Rectifiers

13

537

AC Voltage Regulators

14

613

DC Choppers

15

661

DC to AC Inverters - Switched Mode

16

719

DC to AC Inverters - Resonant Mode

17

739

DC to DC Converters - Switched Mode

20

FACTS Devices and Custom Controllers

21

Inverter Grid Connection for Embedded Generation

22

Energy Sources and Storage - Primary Sources

23

Energy Sources and Storage - Secondary Sources

28

1293

Contactors and relays

PREFACE

The book is in four parts.

Part 1 covers power semiconductor switching devices, their static and dynamic electrical and thermal characteristics and properties. Part 2 describes device driving and protection, while Part 3 presents a number of generic applications. The final part, Part 4, introduces capacitors, magnetic components, resistors, and dc relays and their characteristics relevant to power electronic applications.

The 156 non-trivial worked examples cover the key issues in power electronics.

BWW
April 2010

CHAPTER 1

Basic Semiconductor Physics and Technology

The majority of power electronic circuits utilise power semiconductor switching devices which *ideally* present infinite resistance when off, zero resistance when on, and switch instantaneously between those two states. It is necessary for the power electronics engineer to have a general appreciation of the semiconductor physics aspects applicable to power switching devices so as to be able to understand the vocabulary and the non-ideal device electrical phenomena. To this end, it is only necessary to attempt a qualitative description of switching devices and the relation between their geometry, material parameters, and physical operating mechanisms.

Typical power switching devices such as diodes, thyristors, and transistors are based on a monocrystalline group IV silicon semiconductor structure or a group IV polytype, silicon carbide. These semiconductor materials are distinguished by having a specific electrical conductivity, σ, somewhere between that of good conductors (>10^{20} free electron density) and that of good insulators (<10^{3} free electron density). Silicon is less expensive, more widely used, and a more versatile processing material than silicon carbide, thus the electrical characteristics and processing properties of silicon are considered first, in more detail.

In pure silicon at equilibrium, the number of *electrons* is equal to the number of *holes*. The silicon is called *intrinsic* and the electrons are considered as negative charge-carriers. Holes and electrons both contribute to conduction, although holes have less mobility due to the covalent bonding. Electron-hole pairs are continually being *generated* by thermal ionization and in order to preserve equilibrium previously generated pairs *recombine*. The intrinsic carrier concentrations n_i are equal, small (1.4×10^{10} /cc), and highly dependent on temperature. In order to fabricate a power-switching device, it is necessary to increase greatly the free hole or electron population. This is achieved by deliberately doping the silicon, by adding specific impurities called *dopants*. The doped silicon is subsequently called *extrinsic* and as the concentration of dopant N_c increases, the resistivity $\rho = 1/\sigma$ decreases.

n-type:- Silicon doped with group V elements, such as As, Sb or P, will be rich in electrons compared to holes. Four of the five valence electrons of the group V dopant will take part in the covalent bonding with the neighbouring silicon atoms, while the fifth electron is only weakly attached and is relatively 'free'. The semi-conductor is called *n-type* because of its free negative charge-carriers. A group V dopant is called a *donor,* having donated an electron for conduction. The resultant electron impurity concentration is denoted by N_D - the donor concentration.

p-type:- If silicon is doped with atoms from group III, such as B, Aℓ, Ga or In, which have three valence electrons, the covalent bonds in the silicon involving the dopant will have one covalent-bonded electron missing. The impurity atom can accept an electron because of the available thermal energy. The dopant is thus called an *acceptor*, which is ionised with a net positive charge. Silicon doped with acceptors is rich in holes and is therefore called *p-type.* The resultant hole impurity concentration is denoted by N_A - the acceptor concentration.

Electrons in n-type silicon and holes in p-type are called *majority carriers,* while holes in n-type and electrons in p-type are called *minority carriers*. In a given silicon material, at equilibrium, the product of the majority and minority carrier concentration is a constant:

$$p_o \times n_o = n_i^2 \tag{1.1}$$

where p_o and n_o are the hole and electron equilibrium carrier concentrations.

Therefore, the majority and minority concentrations are given by:

$$\text{For an } n\text{-type } n_o = N_D \text{ therefore } p_o = \frac{n_i^2}{N_D} \text{ and}$$
$$\text{For a } p\text{-type } p_o = N_A \text{ therefore } n_o = \frac{n_i^2}{N_A} \tag{1.2}$$

These equations show that the number of minority carriers decreases as the doping level increases. The resistivity, ρ, of doped silicon is

$$\rho = \frac{1}{\sigma} = \frac{1}{q\left(\mu_n n + \mu_p p\right)} \tag{1.3}$$

where: $\sigma = 1/\rho$ = conductivity, $\Omega^{-1}.cm^{-1}$
$\rho = 1/\sigma$ = resistivity, $\Omega.cm$
μ_n = electron mobility, $cm^2/V\text{-}s$
μ_p = hole mobility, $cm^2/V\text{-}s$
q = electron charge, 1.602×10^{-19} C
n = electron concentration, cm^{-3}
p = hole concentration, cm^{-3}

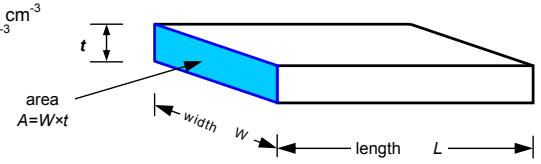


Figure 1.1. *Elemental doped silicon.*

Resistance of semiconductor materials is usually expressed in terms of sheet resistance R_s, which is related to resistance as follows. The impurity depth x_j, mobility μ, and impurity distribution $N(x)$ are related to sheet resistance by

$$R_s = \frac{1}{q\int_0^{x_j} \mu N(x)dx} \qquad \Omega/\text{square} \tag{1.4}$$

The average resistivity is $\bar{\rho} = R_s x_j$ and given a length L and width w, as defined in figure 1.1, the resistance is given by

$$R = \rho\frac{L}{A} = \frac{\rho}{t}\frac{L}{w} = R_s\frac{L}{w} \qquad \Omega \tag{1.5}$$

For consecutive n-doped profiles, the resistance can be estimated by treating each layer independently:

$$R_{total}^{-1} = R_1^{-1} + R_2^{-1} + .. = \frac{wt_1}{L\rho_1} + \frac{wt_1}{L\rho_1} + .. = \frac{w}{L}\left(\frac{t_1}{\rho_1} + \frac{t_1}{\rho_1} + ..\right) = \frac{w}{L}\left(R_{s1}^{-1} + R_{s2}^{-1} + ..\right) = \sum_{i=1} q\mu_{ni}N_{Di}t_i \tag{1.6}$$

Example 1.1: ***Resistance of homogeneously doped silicon***

Silicon doped with phosphorous ($N_D = 10^{17}/cm^3$) measures 100μm by 10μm by 1μm. Calculate the sheet resistance and resistance between opposite faces, assuming the electron mobility at this doping level is μ_n = 720cm^2/ V-s. Doping to produce a p-type material has a hole mobility of 40% that for electrons. Recalculate sheet resistance and resistance values.

Solution

From equation (1.3), the resistivity, ρ, of doped silicon is

$$\rho = \frac{1}{\sigma} = \frac{1}{q\left(\mu_n n + \mu_p p\right)}$$

Since $n >> p$ in the n-type silicon

$$\rho = \frac{1}{q\mu_n n} = \frac{1}{1.6\times10^{-19}\times720\times10^{17}} = 0.086\Omega\text{cm}$$

For a length of 100μm, the resistance is

$$R = \rho \times \frac{Length}{Area} = \rho \times \frac{L}{w \times t} = 0.086 \times \frac{100 \times 10^{-4}}{10 \times 10^{-4} \times 1 \times 10^{-4}} = 8.6\text{k}\Omega$$

From equation (1.5) the sheet resistance is given by

$$R_s = R\frac{W}{L} = 8.6\text{k}\Omega \times \frac{10 \times 10^{-4}}{100 \times 10^{-4}} = 860\,\Omega/\text{square}$$

If the length is assumed to be one of the shorter dimensions, then for a length 10μm or 1μm, the resistance is 86Ω or 0.86Ω, respectively, while the sheet resistance possibilities, depending on the thickness reference axis, are 86 Ω/square and 8.6 Ω/square.

For a p-type material, the 40% decrease in mobility of holes μ_p increases resistivity by a factor of 1/0.4 = 2.5. Each aspect resistance therefore increases by a factor 2.5, viz., increases to 21.5kΩ, 215Ω, and 2.15Ω for lengths 100μm, 10μm, and 1μm, respectively. From equation (1.4) the sheet resistances are increased to 2.15kΩ/square, 215Ω/square, and 21.5Ω/square.

♣

The carrier concentration equilibrium can be significantly changed by irradiation by photons, the application of an electric field or by heat. Such carrier injection mechanisms create *excess carriers.*

If n-type silicon is irradiated by photons with enough energy to ionise the valence electrons, electron-hole pairs are generated. There is already an abundance of majority electrons in the n-type silicon, thus the photon-generated excess minority holes are of more relative and detectable importance. If the light source is removed, the time constant associated with recombination, or decay of excess minority carriers, is called the *minority carrier hole lifetime,* τ_h. For a p-type silicon, exposed to light, excess minority electrons are generated and after the source is removed, decay at a rate called the *minority carrier electron lifetime,* τ_e. The minority carrier lifetime is often called the *recombination lifetime.*

A difficulty faced by manufacturers of high-voltage, large-area semiconductor devices is that of obtaining uniformity of n-type phosphorus doping throughout the usual high-resistivity silicon starting material. Normal crystal growing (by liquid encapsulated, contactless, Czochralski crystal growth – see section 1.19.3i) and doping techniques give no better than ±10 per cent fluctuation around the wanted resistivity at the required low concentration levels (<10^{14} /cc). Final device electrical properties will therefore vary widely in all lattice directions. Tolerances better than ±1 per cent in resistivity and homogeneous distribution of phosphorus can be attained by neutron radiation, commonly called *neutron transmutation doping,* NTD. The neutron irradiation flux transmutes silicon atoms first into a silicon isotope with a short 2.62-hour half-lifetime, which then decays into phosphorus. Subsequent thermal annealing removes any crystal damage caused by the irradiation. Neutrons can penetrate over 100mm into silicon, thus large silicon crystals can be processed using the NTD technique.

A p-n junction is the location in a semiconductor where the doping changes from p to n while the monocrystalline lattice continues undisturbed. A bipolar diode is thus created, which forms the basis of any bipolar semiconductor device.

The donor-acceptor doping junction is formed by any one of a number of process techniques, namely alloying, diffusion, epitaxy, ion implantation or the metallization for ohmic contacts.

Power semiconductor device processing involves most of the following range of process possibilities.

- ***Deposition*** is any process that grows, coats, or otherwise transfers a material onto the wafer. Available technologies consist of physical vapour deposition (PVD), chemical vapour deposition (CVD), electrochemical deposition, molecular beam epitaxy (MBE) and atomic layer deposition among others.
- ***Removal*** processes are any that remove material from the wafer either in bulk or selective form and consist primarily of etch processes, both wet etching and dry etching such as reactive ion etch.
- ***Patterning*** covers the series of processes that shape or alter the existing shape of the deposited materials and is generally referred to as lithography. In conventional lithography, the wafer is coated with a chemical called a *photoresist*. The photoresist is exposed by a 'stepper', a machine that focuses, aligns, and moves the mask, exposing select portions of the wafer to short wavelength UV light. The unexposed regions are washed away by a developer solution. After etching or other processing, the remaining photoresist is removed by oxygen plasma ashing or stripping.
- ***Modification*** of electrical properties consists of doping transistor sources and drains in diffusion furnaces and by ion implantation. These doping processes are followed by furnace annealing or in advanced devices, by rapid thermal annealing which serve to activate the implanted dopants. Modification of electrical properties extends to reduction of dielectric constant in low-k insulating materials via exposure to ultraviolet light.

1.1 Processes forming and involved in forming semiconductor devices

1.1.1 Alloying

At the desired region on an n-type wafer, a small amount of p-type impurity is deposited. The wafer is then heated in an inert atmosphere and a thin film of melt forms on the interface. On gradual, slow cooling, a continuous crystalline structure results, having a step or abrupt pn junction as shown in figure 1.2. This process is not employed to form modern p-n junctions but can be used at the metallisation stage of wafer fabrication.

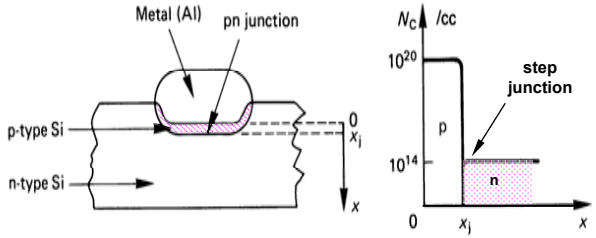


Figure 1.2. *N-Si to Aℓ metal alloy junction: (a) cross-section where x_j is the junction depth below the metal-semiconductor boundary and (b) doping profile of the formed step junction.*

1.1.2 Diffusion

Diffusion, the movement of a chemical species from an area of high concentration to an area of lower concentration, is one of the two major processes by which chemical dopants are introduced into a semiconductor (the other process being ion implantation). The controlled diffusion of dopants into silicon to alter the type and level of conductivity of semiconductor materials is the foundation of forming a p-n junction and formation of devices during wafer fabrication, as shown in figure 1.3. It is used to form bases, emitters, and resistors in bipolar devices, as well as drains and sources in MOS devices. It is also used to dope polysilicon layers.

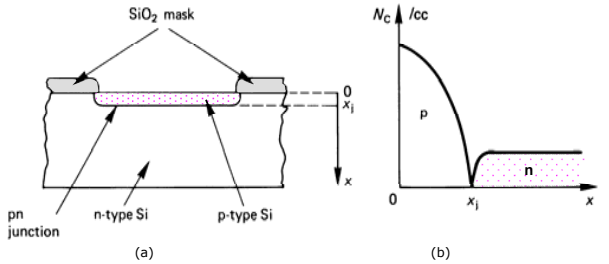


Figure 1.3. *Diffused pn junction: (a) cross-section where x_j is the junction depth below the silicon surface and (b) doping concentration profile.*

The mathematics that govern the mass transport phenomena of diffusion are based on two concepts.

First Concept

Whenever an impurity concentration gradient, ∂C/∂x, exists in a finite volume of a matrix substance (the silicon substrate in this context), the impurity material has a natural tendency to move in order to distribute itself more evenly within the matrix and decrease the concentration gradient.

Given time, this flow of impurities eventually results in homogeneity within the matrix, causing the net flow of impurities to stop. The mathematics of this transport mechanism is based on the flux of material across a given plane is proportional to the concentration gradient across the plane. That is:

$$J = -D\frac{\partial N(x,t)}{\partial x} \tag{1.7}$$

where J is the flux,
$D = \mu kT$ is the diffusion constant or diffusivity for the material that is diffusing the solvent, m/s,
$\partial N(x,t)/\partial x$ is the concentration gradient. k is Boltzmann's constant and μ is ionic mobility.
The diffusion constant of a material is also referred to as *diffusion coefficient* or *diffusivity* and is related to mobility by $D = \mu kT$. It is expressed in units of length^2/time, such as μm^2/hour. The negative sign of the right side of the equation indicates that the impurities flow to the lower concentration.

Second Concept
Equation (1.7) does not account for the fact that the gradient and local concentration of the impurities in a finite volume of material decreases with an increase in time, an aspect that is important to diffusion processes.
The flux J_1 of impurities entering a section of a material with a concentration gradient is different from the flux J_2 of impurities leaving the same section. From the law of conservation of matter, the difference between J_1 and J_2 must result in a change in the concentration of impurities within the section, assuming that no impurities are formed or consumed in the section.
The second concept states that the change in impurity concentration over time is equal to the change in local diffusion flux, or

$$\frac{\partial N(x,t)}{\partial t} = -\frac{\partial J}{\partial x}$$

or, from the first concept, equation (1.7)

$$\frac{\partial N(x,t)}{\partial t} = \frac{\partial\left(D\frac{\partial N(x,t)}{\partial x}\right)}{\partial x} \tag{1.8}$$

If the diffusion coefficient is independent of position, such as when the impurity concentration is low, then the second concept may be simplified to:

$$\frac{\partial N(x,t)}{\partial t} = D\frac{\partial^2 N(x,t)}{\partial x^2} \tag{1.9}$$

There are two major ways by which to deposit impurities into a substance by thermal diffusion. In the first method, known as *predeposition*, a flux of impurities continuously arrives at the surface of the substrate such that the concentration gradient of the impurity remains constant at the surface of the substrate, as shown in figure 1.4b. In the second method, known as redistribution or *drive-in* diffusion, a thin layer of the impurity material is deposited on the substrate. In this case, the impurity gradient at the surface of the substrate decreases with time, as shown in figure 1.4c.

The semiconductor diffusion process is usually performed in two steps: predeposition and then drive-in.
During *predeposition*, the impurity dopant is added to the wafer n-type silicon substrate. Predeposition is done in a diffusion furnace at temperatures around 1000 to 1250ºC. The dopant is introduced into the furnace, and may be in the form of a gas, solid, or liquid. Gaseous dopants are mixed with an inert carrier gas, such as nitrogen or argon, and introduced into the furnace. Solid dopants are often applied in a powder form. The solid is heated and a stream of carrier gas moves the dopant into the furnace. Liquid sources are used by bubbling an inert carrier gas through the liquid dopant, and the gas saturated with the liquid is added to the furnace. This compound breaks down as a result of the high temperature, and is slowly diffused into the substrate. The maximum impurity concentration occurs at the surface, tailing off towards the inside.
The wafers are then put into a second furnace at higher temperatures (about 1300ºC) to *drive-in* the dopant. The drive-in process usually occurs in an oxidizing atmosphere so that a protective layer of SiO_2 is grown over the diffused layer.

Table 1.1: Dopants and chemical reactions

Dopant state	Dopant type	dopant	chemistry
gas	p-type	diborane B_2H_6	$B_2H_6 + 3O_2 \rightarrow B_2O_3 + 3H_2O$
	n-type	arsine AsH_3 phosphine PH_3	$2PH_3 + 4O_2 \rightarrow B_2O_5 + 3H_2O$
liquid	p-type	BBr_3	$4BBr_3 + 3O_2 \rightarrow 2B_2O_3 + 6Br_2$
	n-type	$AsC\ell_3$, $P0C\ell_3$	$4P0C\ell_3 + 3O_2 \rightarrow 2P_2O_5 + 6C\ell_2$
solid	p-type	BN, B_2O_3	$2B_2O_3 + 3Si_4 \rightarrow 4B + 3SiO_2$
	n-type	As_2O_3, P_2O_5	$2As_2O_3 + 3Si \rightarrow 4As + 3SiO_2$ $2P_2O_5 + 5Si_4 \rightarrow 4P + 5SiO_2$

Typical dopants and silicon chemical reactions are shown in Table 1.1, while common diffusion coefficients and activation energies, referenced to 0K, are shown in Table 1.2.
The diffusion process is the only junction forming technique that is not applicable to silicon carbide wafer processing.

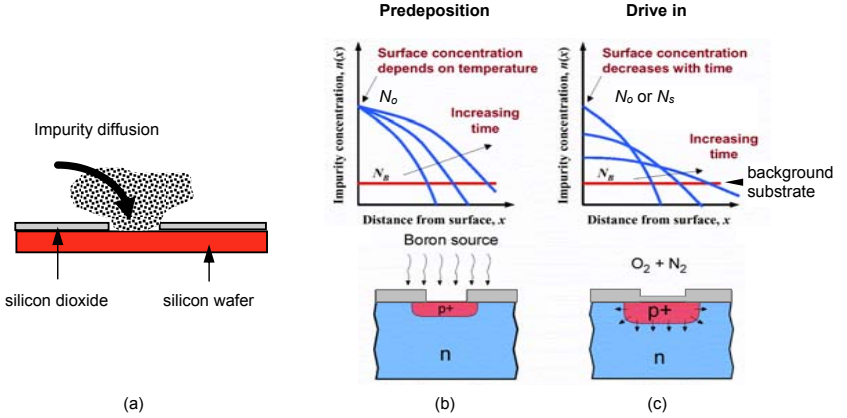


Figure 1.4. *Diffusion processes:*
(a) pictorial representation of mechanism; (b) predeposition diffusion; and (c) drive in diffusion.

The doping profile is mathematically defined and is varied by controlling the vapour mixture concentration, the furnace temperature, and time of diffusion.
If the source concentration is continuously replenished – ***predeposition dose***, thus maintained constant, the surface concentration is $N(0,t) = N_s$, and the initial concentration is $N(x,0)=0$, then the doping profile is given by a complementary error function, *erfc*.

$$N(x,t) = N_s\, erfc\left(\frac{x}{2\sqrt{Dt}}\right)$$

$$= N_s\left(1 - erf\left(\frac{x}{2\sqrt{Dt}}\right)\right) = N_s\left(1 - erf\,(u)\right) = N_s\left(1 - \frac{2}{\sqrt{\pi}}\int_0^u e^{-v^2}dv\right) \tag{1.10}$$

where $u = \dfrac{x}{2\sqrt{Dt}} = \dfrac{x}{\text{difussion length}}$

The area under the diffusion profile is the total amount of dopant diffused into the wafer:

$$Q(t) = \int_o^\infty N(x,t)\,dx = \frac{2}{\sqrt{\pi}}N_s\sqrt{Dt} = 1.13N_s\sqrt{Dt} \tag{1.11}$$

The junction depth is where the doping profile $N(x,t)$ equals the background doping N_B level, that is

$$N(x_j,t) = N_s\, erfc\left(\frac{x_j}{2\sqrt{Dt}}\right) = N_B$$

Rearranging, gives the junction depth x_j as

$$x_j = 2\sqrt{Dt} \times erfc^{-1}\frac{N_B}{N_s} \tag{1.12}$$

If natural depletion of dopant occurs **– *drive in***, that is the initial dose S at the surface is not replenished, then the profile is an exponential function, which gives a Gaussian diffusion distribution.

$$N(x,t) = \frac{S}{\sqrt{\pi Dt}}e^{-\frac{x^2}{4Dt}} \tag{1.13}$$

where

$$\int_o^\infty N(x,t)\,dx = S = \text{non-replenished inital surface dose} \tag{1.14}$$

The diffusion length, $x = 2\sqrt{Dt}$, is an approximate measure of how far the dopant has diffused, which is the distance from the surface to where the concentration has fallen to 1/e.

The surface concentration, which is not replenished ($dN(0,t)/dx = 0$), diminishes with time, according to

$$N(0,t) = \frac{S}{\sqrt{\pi Dt}} \tag{1.15}$$

The junction depth is where the doping profile $N(x,t)$ equals the background doping N_B level, that is

$$N(x_j,t) = \frac{S}{\sqrt{\pi Dt}} e^{-\frac{x^2}{4Dt}} = N_B$$

Rearranging, gives the junction depth x_j as

$$x_j = 2\sqrt{Dt} \times \left(\ell n \frac{S}{\sqrt{\pi Dt}\, N_B} \right)^{½} = 2\sqrt{Dt} \times \left(\ell n \frac{N(0,t)}{N_B} \right)^{½} \tag{1.16}$$

In both processing cases, $N(\infty,t) = 0$.

Diffusivity D varies with temperature according to

$$D = D_o \, e^{-\frac{E_a}{kT}} \tag{1.17}$$

where D_o = diffusion coefficient (in cm^2/s) extrapolated to infinite temperature
E_a = activation or threshold energy in eV, which is not particularly temperature dependant.

Table 1.2: Typical diffusion coefficients and activation energies at 0K

Element		D_o	E_a
	0K	cm^2/s	eV
boron	B	1.0	3.50
phosphorous	P	4.7	3.68
antimony	At	4.58	3.88
arsenic	As	9.17	3.99
indium	In	1.20	3.50

Example 1.2: ***Constant Surface Concentration diffusion - predepostion***

For a constant-source boron diffusion into n-type $10^{15}\,cm^{-3}$ silicon at 1000°C, the surface concentration is maintained at $10^{19}\,cm^{-3}$ and the diffusion time is 1 hour. Find

i. Total amount of dopant diffused, $Q(t)$ and the gradient at $x = 0$ and
ii. The gradient and location (junction depth) where the dopant concentration reaches $10^{15}\,cm^{-3}$.

Solution

Using data for boron in Table 1.2, equation (1.17) gives the diffusion coefficient of boron at 1000°C as

$$D = D_o e^{-\frac{E_a}{kT}} = 24e^{-\frac{3.05}{8.614\times10^{-5}\times1273}} = 1.39\times10^{-14}\,cm^2/s$$

so the diffusion length is

$$\sqrt{Dt} = \sqrt{1.39\times10^{-14}\times3600} = 7.07\times10^{-6}\,cm$$

i. The area under the diffusion profile from equation (1.11) is

$$Q(t) = 1.13 N_s \sqrt{Dt} = 1.13\times10^{19}\times7.07\times10^{-6} = 8.0\times10^{13}\,cm^{-2}$$

$$\left.\frac{dN}{dx}\right|_{x=0} = -\frac{N_s}{\sqrt{\pi Dt}} = -\frac{10^{19}}{\sqrt{\pi}\times7.07\times10^{-6}} = -7.98\times10^{23}\,cm^{-4}$$

ii. From equation (1.10) rearranged, when $N_B = 10^{15}\,cm^{-3}$, x_j is given by

$$x_j = 2\sqrt{Dt}\times erfc^{-1}\left(\frac{N_B}{N_s}\right) = 2\sqrt{Dt}\times erfc^{-1}\left(\frac{10^{15}}{10^{19}}\right)$$

$$= 2\times7.07\times10^{-6}\times2.75 = 0.389\mu m$$

$$\left.\frac{dN}{dx}\right|_{x=0.389\mu m} = -\frac{N_s}{\sqrt{\pi Dt}} e^{-\frac{x^2}{4Dt}} = -4.0\times10^{20}\,cm^{-4}$$

♣

Example 1.3: ***Constant Total Dopant diffusion – drive in - 1***

Arsenic was pre-deposited by arsine gas, and the resulting dopant per unit area was $10^{14}\,cm^{-2}$. How long would it take to drive the arsenic in to $x_j = 1\,\mu m$? Assume a background doping of $N_{sub} = 10^{15}\,cm^{-3}$, and a drive-in temperature of 1200°C. For As, assume $D_o = 24\,cm^2/s$, and $E_a = 4.08$ eV at 1200°C.

Solution

From equation (1.17) the diffusion coefficient for arsenic at 1200°C is

$$D = D_o e^{-\frac{E_a}{kT}} = 24e^{-\frac{4.08}{8.614\times10^{-5}\times1473}} = 2.602\times10^{-13}$$

Rearranging equation (1.13) gives

$$x_j^2 = 10^{-8} = 4Dt \; \ell n\left(\frac{S}{N_B\sqrt{\pi Dt}}\right) = 1.04\times10^{-12} t \; \ell n\left(\frac{1.106\times10^5}{\sqrt{t}}\right)$$

That is

$$t \times \ell n\, t - 23.22t + 19230 = 0$$

An iterative solution gives t=1191.7s or approximately 19.9 minutes.

♣

Example 1.4: ***Constant Total Dopant diffusion – drive in - 2***

An arsenic constant-dose diffusion is performed with an initial dose of S=$10^{14}\,cm^{-2}$. The diffusion temperature is 1100°C for 2 hours. The starting wafer had a p-type substrate background doping of $10^{17}\,cm^{-3}$. Find the concentration of the As at the surface and find the junction depth.

Solution

From Table 1.2

$$D = D_o e^{-\frac{E_a}{kT}} = 9.17e^{-\frac{3.99}{8.614\times10^{-5}\times1100+273}} = 2.07\times10^{-14}\,cm^2/s$$

Then the diffusion length is

$$\sqrt{Dt} = \sqrt{2.07\times10^{-14}\times7200} = 1.22\times10^{-5}\,cm$$

The surface concentration is

$$\left.\frac{dN}{dx}\right|_{x=0} = N_o = \frac{N_s}{\sqrt{\pi Dt}} = \frac{10^{18}}{\sqrt{\pi}\times1.22\times10^{-5}} = 4.6\times10^{18}\,cm^{-3}$$

From equation (1.13) rearranged, the junction depth for Gaussian diffusions is

$$x_j = 2\sqrt{Dt}\; \ell n\left(\frac{N_o}{N_B}\right)^{½}$$

$$= 2\times1.22\times10^{-5}\,cm\; \ell n\left(\frac{4.6\times10^{18}\,cm^{-3}}{10^{17}\,cm^{-3}}\right)^{½}$$

$$= 0.467\mu m$$

♣

1.1.3 Epitaxy growth - deposition

Epitaxy or epitaxial growth is the process of depositing a non-volatile, thin solid layer typically 0.5 to 100 µm, of single crystal material over a single crystal substrate, usually through chemical vapour deposition (CVD). The semiconductor deposited film is often the same material as the substrate, and the process is known as homoepitaxy, or simply, epi, as with silicon deposition on a silicon substrate. If the substrate is an ordered semiconductor crystal (that is mono-silicon, gallium arsenide), the process continues building on the substrate with the same crystallographic orientation, with the substrate acting as a seed for the deposition. If an amorphous/polycrystalline substrate surface is used, the film will also be amorphous or polycrystalline. A key feature of epitaxy is that a lightly doped layer of epitaxial silicon can be grown on top of a heavily doped silicon substrate, thus creating a layer of differing conductivity that can serve as an insulating layer or intrinsic buffer region.
The chemical vapour deposition CVD (see section 1.2.1) of silicon epitaxy occurs in an epitaxial reactor that consists of a quartz induction heated reaction chamber into which a susceptor is placed. The susceptor provides two features:

- mechanical support for the wafers and
- an environment with uniform thermal distribution.

The technological method of introducing reactant gases with only the substrates heated inside a reactor is called Vapour Phase Epitaxy, a schematic of which is shown in figure 1.5.
A possible fabrication process is as follows. A pre-cleaned, polished, almost perfect silicon crystal surface acts as a substrate for subsequent deposition. Usually hydrogen chloride is first used to etch the wafers. The pre-doped silicon is heated to about 1150°C in a quartz reactor tube at atmospheric pressure. A hydrogen gas flow carrying a compound of silicon such as silicon tetrachloride $SiC\ell_4$ or silane SiH_4 is passed over the hot substrate surface, and silicon atoms are deposited, growing a new continuous lattice. If phosphine (PH_3) arsine (AsH_3) or diborane (B_2H_6) is included in the silicon compound carrier gas flow of H_2 and N_2, a layer of the required type and resistivity occurs. Up to 100µm of doped silicon can be grown on substrates for power devices at a high growth rate of about 1 µm/min at 1200°C. A very low crystalline fault rate is essential if uniform electrical properties are to be attained. Selective deposition, depending on the surface masking of the substrate, is possible.

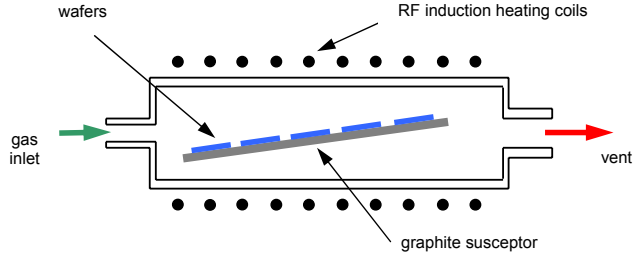


Figure 1.5. *Typical cold-wall vapour phase epitaxial reactor.*

There are four major chemical sources of silicon for epitaxial deposition:

- silane, SiH_4
- silicon tetrachloride, $SiC\ell_4$;
- trichlorosilane, $SiHC\ell_3$; and
- dichlorosilane, $SiH_2C\ell_2$.

Chemical reactions equations can describe the growth of epitaxial layers. Each of the chemical sources mentioned can be described by an over-all reaction equation that shows how the vapour phase reactants form the silicon epitaxial film. For example, the over-all pyrolytic reaction for silicon epitaxy by silane decomposition reaction is:

$$SiH_4 \rightarrow Si + 2H_2 \qquad (1000°C \text{ to } 1100°C) \tag{1.18}$$

Hydrogen reduction of trichlorosilane is

$$SiHC\ell_3 + H_2 \rightarrow Si + 3HC\ell \tag{1.19}$$

Reduction of dichlorosilane is

$$SiH_2C\ell_2 \rightarrow Si + 2HC\ell \tag{1.20}$$

However, such over-all reaction equations do not describe the complete CVD process in regard to how the gas phase reactants interact or how the epi species are adsorbed on the substrate surface. For instance, the over-all reaction for the hydrogen reduction of silicon tetrachloride $SiC\ell_4$ to form a silicon epitaxial layer is as follows:

$$SiC\ell_4 + 2H_2 \rightarrow Si + 4\,HC\ell \qquad (1150°C \text{ to } 1300°C) \tag{1.21}$$

Yet, the intermediate chemical species such as $SiHC\ell_3$ and $SiH_2C\ell_2$ are present during the silicon epitaxial growth:

$$SiC\ell_4 + H_2 \leftrightarrow SiHC\ell_3 + HC\ell$$
$$SiHC\ell_3 + H_2 \leftrightarrow SiH_2C\ell_2 + HC\ell$$
$$SiH_2C\ell_2 \leftrightarrow SiC\ell_2 + H_2$$
$$SiHC\ell_3 \leftrightarrow SiC\ell_2 + HC\ell$$
$$SiC\ell_2 + H_2 \leftrightarrow Si + 2HC\ell$$

These equations confirm that even if a given process is described by a single over-all reaction, the process is actually a combination of many simultaneous chemical reactions.
The growth rate of an epitaxial layer depends on several factors:

- the chemical sources;
- the deposition temperature; and
- the mole fraction of reactants.

Silicon epitaxy improves the performance of bipolar devices. By growing a lightly doped epi-layer over a heavily-doped silicon substrate, a higher breakdown voltage across the collector-substrate junction is achieved while maintaining low collector resistance. Lower collector resistance allows a higher operating speed with the same current. Epitaxy is also used in IC fabrication. By fabricating a CMOS device on a thin (3 to 7 microns) lightly doped epi layer grown over a heavily-doped substrate, latch-up occurrence is minimized – a phenomena applicable to power devices such as the MOSFET and IGBT.
As well as improving the performance of devices, epitaxy also allows better control of doping concentrations of the devices. The layer can also be made oxygen and carbon free. The disadvantages of epitaxy include higher cost of wafer fabrication, additional process complexities, and problems associated with defects in the epi-layer.

1.1.4 Ion-implantation and damage annealing

Ion Implantation is the process of depositing chemical dopant species (atoms stripped of electrons) into a substrate by directly bombarding the substrate with high-energy ions of the chemical being deposited, as shown in figure 1.6.
Diffusion and ion implant are the two major processes by which chemical species or dopants are introduced into a semiconductor such as silicon to form electronic structures. The advantage of ion implant over diffusion is its more precise control for depositing dopant atoms into the substrate (10^{11} to 10^{18} cm^{-2}), giving excellent doping level uniformity and production repeatability.
The implanted profile shown in figure 1.6b, where two junctions may be formed, can be approximated by a Gaussian distribution function:

$$N(x) = \frac{S}{\sqrt{2\pi}\,\sigma_p} e^{-\frac{(x-R_p)^2}{2\sigma_p^2}} = N_p\, e^{-\frac{(x-R_p)^2}{2\sigma_p^2}} \tag{1.22}$$

where S is the ion dose per unit area, cm^{-2}
σ_p is the symmetrical standard deviation in the projected range of the implanted ions, cm

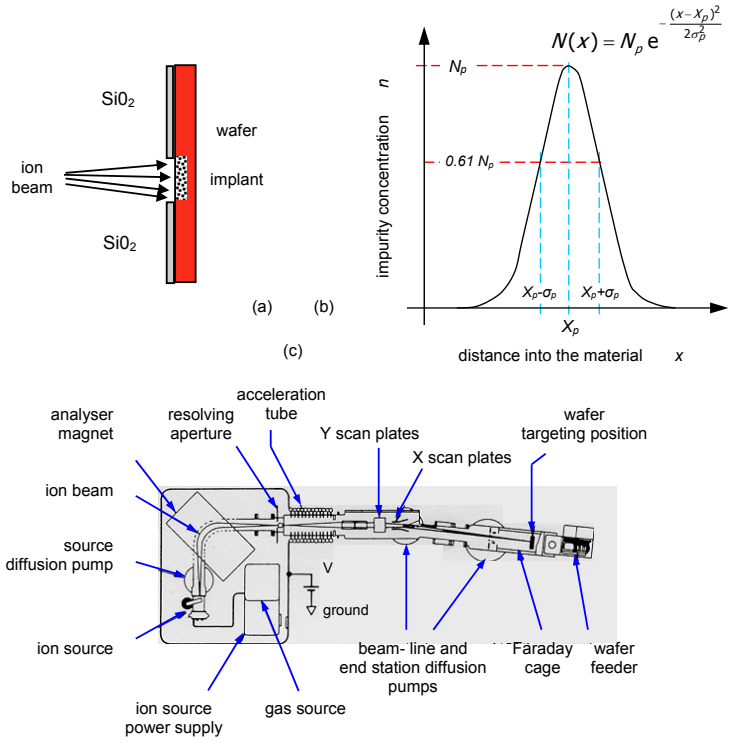


Figure 1.6. *Ion implantation:*
(a) pictorial representation of mechanism; (b) implanted ion distribution; and (c) implanting system.

The depth of average or mean projected range (peak) is at X_p along the axis of incidence, where the maximum concentration occurs.

$$S = \frac{1}{q}\int_o^t I_{beam}(t')dt' = \frac{\text{ion beam current (A)}}{q} \times \text{implant time} \Big/ \text{implant area} \qquad (1.23)$$

The point where the diffused impurity profiles intersects the background concentration N_B is the metallurgical junction depth, x_j, where the net impurity concentration is zero. From equation (1.22)

$$N_B = N_p\, e^{-\frac{(x_j - X_p)^2}{2\sigma_p^2}}$$
$$x_j = X_p \pm \sigma_p\sqrt{2\ell n\frac{N_p}{N_B}} \qquad (1.24)$$

where $N_p = S/\sqrt{2\pi}\,\sigma_p$.

Doping, which is the primary purpose of ion implanting, is used to alter the type and level of conductivity of semiconductor materials. It is used to form bases, emitters, and resistors in bipolar devices, as well as drains and sources in MOS devices. It is also used to dope polysilicon layers.

Typically, a gaseous dopant is ionized by electric discharge or by heat from a hot filament. The ions are separated using an electromagnetic field that bends the positively-charged particles to a selected band. This ion band is then passed through a high-current accelerator. The high-velocity beam of ions is focused on the wafer, causing the dopant ions to strike the wafer surface and penetrate. Sometimes a mask is used to implant a designated pattern on the wafer. As with diffusion, ion implantation allows the formation of junctions by changing the conductivity characteristics of precise regions in the wafer.

The basic procedure for ion implantation into silicon is as follows:

Ion impurities (B, P or As) are vaporised and accelerated by an electric field in a vacuum at high keV energies at the pre-doped silicon substrate, which is at room temperature. The ions penetrate the lattice to less than a few microns, typically 1μm at about ½ MeV. The resultant implanted doping profile is Gaussian, with the smaller ion like boron, penetrating deeper.

These high-energy atoms enter the crystal lattice and lose their energy by colliding with some silicon atoms before finally coming to rest at some depth. Adjusting the acceleration energy controls the average deposition depth of the impurity atoms. Heat treatment is subsequently used to anneal or repair the crystal lattice disturbances caused by the atomic collisions.

Every implanted ion collides with several target atoms before it comes to rest. Such collisions may involve the nucleus of the target atom or one of its electrons. The total power of a target to stop an ion, or its total stopping power S, is the sum of the stopping power of the nucleus and the stopping power of the electron. Stopping power is described as the energy loss of the ion per unit path length of the ion.

Implantation energies are typically 10keV to 1MeV, giving ion distributions with depths of 10 nm to 10 μm from doses vary from 10^{12} ions/cm^2 for threshold voltage adjustment in MOSFETs to 10^{18} ions/cm^2 for formation of buried insulating layers.

The damage caused by atomic collisions and bombardment during high-energy ion implantation changes the material structure therefore electrical characteristics of the target substrate. Many target atoms are displaced, creating deep electron and hole traps which capture mobile carriers and increase resistivity. Annealing is therefore needed to repair the lattice damage and put dopant atoms in substitutional sites where they can be electrically active again.

Silicon damage caused by ion implantation includes:

- the formation of crystal defects such as Frenkel defects, vacancies, di-vacancies, higher-order vacancies, and interstitials;
- the creation of local zones of amorphous material within the supposedly crystalline structure; and
- formation of continuous amorphous layers as the localized amorphous regions grow and overlap.

The first two damage types are categorized as 'primary crystalline damage'. Restoring the ion-implanted substrate to its pre-implant condition requires the substrate being subjected to a reparative thermal process known as *annealing*.

Ion implantation damage annealing has five major components:

- electrical activation of the implanted impurities;
- primary crystalline damage annealing;
- annealing of continuous amorphous layers;
- dynamic annealing; and
- diffusion of implanted impurities.

Annealing is conducted in a neutral environment, such as in Ar or a N_2 atmosphere in a stack furnace.

Electrical activation of the implanted impurities refers to the process of increasing the electrical activity of newly implanted impurity atoms during annealing, which usually do not occupy substitutional sites after

being implanted. Temperatures up to 500°C remove trapping defects, releasing carriers to the valence or conduction bands in the process. Electrical activity decreases again at 500 to 600°C, because of the formation of dislocations. Beyond 600°C, electrical activation increases until a peak at 800 to 1000°C.

In summary, primary crystalline damage annealing consists of:

- recombination of vacancies and self-interstitials in the low temperature range, up to 500°C;
- formation of dislocations at 500 to 600°C which can capture impurity atoms; and
- dissolution of these dislocations at 900 to 1000°C.

Annealing of the continuous amorphous layers that extend to the surface occur by solid-phase epitaxy between 500 to 600°C. The crystalline substrate beneath the amorphous layers initiates the recrystallization of the amorphous layers, with the regrowth proceeding towards the substrate surface. Factors affecting the recrystallization rate include crystal orientation and the implanted impurities. Amorphous layers that do not extend to the surface anneal differently, with the solid-phase epitaxy occurring at both amorphous-single crystal interfaces and the regrowth interfaces meeting below the surface.

Dynamic annealing effects refers to the healing of implant damage while the implantation process is occurring. This takes place because the heat applied to the wafer during implantation makes the point defects more mobile.

Diffusion of implanted impurities relates to the mass transport of implanted species across a concentration gradient within an implanted layer during the annealing process. The presence of implant damage makes this diffusion process more complex than what occurs in an undamaged single-crystal substrate. Diffusion of implanted impurities during annealing degrades devices that have shallow junctions or narrow base and emitter regions if the thermal processing is not rapid enough, particularly in the case of boron ion implantation.

Example 1.5: Ion implantation

For a 100 keV boron implant with a dose of S=5×10^{14} cm^{-2}, calculate

i. the peak concentration if this concentration occurs at a depth of X_p = 0.31 μm and the ion implant standard deviation is σ_p = 0.07 μm,
ii. the junction depth, if the substrate phosphorus background doping level is 10^{15} /cm^3, and
iii. the surface concentration.

Solution

i. From equation (1.22)

$$N(x) = \frac{S}{\sqrt{2\pi}\,\sigma_p}\, e^{-\frac{(x-X_p)^2}{2\sigma_p^2}}$$

Differentiation gives

$$\frac{dn}{dx} = -\frac{S}{\sqrt{2\pi}\sigma_p}\frac{2(x-X_p)}{2\sigma_p^2}\, e^{-\frac{(x-X_p)^2}{2\sigma_p^2}} = 0$$

which confirms that the maximum concentration occurs when $x = X_p$.

Substitution into equation (1.22) gives the concentration $N(x = X_p = 0.31\ \mu m) = 2.85\times10^{18}$ cm^{-3}.

ii. The junction depth is given by equation (1.24), that is

$$x_j = X_p \pm \sigma_p\sqrt{2\ell n\frac{N_p}{N_B}}$$
$$= 0.31 \pm 0.07\sqrt{2\ell n\frac{2.85\times10^{18}}{10^{15}}} = 0.31 \pm 0.28\mu m$$

Two junctions are created, at 0.03 μm and 0.59 μm.

iii. Since the ion implant has formed two junctions within the n-substrate, the surface concentration is dominated by the background doping level of the substrate, 10^{15} /cm^3. The surface ion implant doping is given by equation (1.22)

$$N(x=0) = \frac{S}{\sqrt{2\pi}\sigma_p}\, e^{-\frac{X_p^2}{2\sigma_p^2}} = 2.85\times10^{18}\times e^{-\frac{0.31^2}{2\times0.07^2}} = 1.57\times10^{14}/\text{cm}^3$$

The n-type surface concentration is 10^{15} /cm^3 – 1.57×10^{14} /cm^3 = 8.4×10^{14} /cm^3.

♣

1.2 Thin Film Deposition

A thin film is a layer with a high surface-to-volume ratio. Thin films are extensively used to apply dopants and sealants to wafers and microelectronic parts, and can be a resistor, a conductor, an insulator, or a semiconductor. Thin films can be deposited with a thickness of between a few nanometres to about 100μm. The film can subsequently be locally etched using processes described in the Lithography and Etching sections of this chapter, sections 1.5 and 1.6, respectively.
Thins films behave differently from bulk materials of the same chemical composition in several ways. Thin films are sensitive to surface properties while bulk materials generally are not. Thin films are also more sensitive to thermo-mechanical stresses. Thin film integrity is influenced by the quality of its adhesion to and conformal coverage of the underlying layer, residual or intrinsic stresses after deposition, and the presence of surface imperfections such as pinholes.
The adhesion of a thin film to the substrate or underlying layer is paramount to ensuring thin film reliability. A thin film that is initially adhering to the underlying layer may lift off after the device is subjected to thermo-mechanical stresses. Reliable thin film adhesion depends on the cleanliness of the surface upon which it is deposited. Optimum substrate roughness also affects thin film adhesion. An ultra-smooth substrate decreases adhesion tendency. A rough substrate on the other hand can result in coating defects, which can also lead to thin film adhesion failures.
Regardless of the deposition process, thin films always have an intrinsic stress that can be either tensile or compressive. High residual stresses can lead to adhesion problems, corrosion, cracking, and deviations in electrical properties. Thus, proper deposition is critical to minimize intrinsic stresses in thin films.

Deposition technology is classified into two reaction types, viz. chemical and physical:

i. Deposition that results because of a chemical reaction:
- Chemical Vapour Deposition (CVD)
- Electrodeposition
- Epitaxy
- Thermal oxidation

These processes exploit the creation of solid materials directly from chemical reactions in gas and/or liquid compositions or with the substrate material. The solid material is usually not the only product formed by the reaction. By-products can include gases, liquids and other solids.

ii. Deposition that results because of a physical reaction:
- Physical Vapour Deposition (PVD)
- Casting

Common for these processes is that the material deposited is physically moved onto the substrate. In other words, there is no chemical reaction that forms the material on the substrate. This is not completely correct for casting processes, though it is more convenient to classify them as such.

Whether the process is physical or chemical, the processing deposition reactor uses either:
- A *cold wall* system, where the heating process uses radio frequency or infra red heating, while
- A *hot wall* system uses a thermal heating resistive element or series of elements forming heating zones.

1.2.1 Chemical Vapour Deposition (CVD)

A fluid precursor undergoes a chemical change at a solid surface, leaving a solid layer.
In this process, the substrate is placed inside a reactor to which a number of gases are supplied, as shown in figure 1.7. The fundamental principle of the process is that a chemical reaction takes place between the source gases. The product of that reaction is a solid material that condenses on all surfaces inside the reactor.
CVD is capable of producing thick, dense, ductile, and good adhesive coatings on metals and non-metals such as glass and plastic. In contrast to PVD coating in the 'line of sight', CVD can simultaneously coat all surfaces of the substrate. The thin films from chemical deposition techniques tend to be conformal, rather than directional.
CVD processes are used to produce a thin film with good step coverage. A variety of materials can be deposited, however, some form hazardous by-products during processing. The quality of the material varies from process to process, however generally a higher process temperature yields a material with higher quality and fewer defects. They are generally not suitable for mixtures of materials. CVD processing is not possible for some materials; there simply is no suitable chemical reaction.

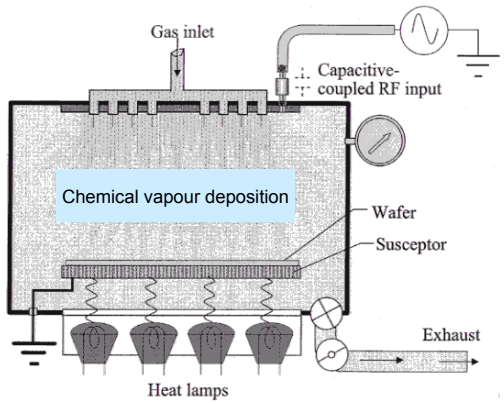


Figure 1.7. *Typical CVD processing reactor system.*

Chemical deposition is categorized by the phase of the precursor:

- Plating relies on liquid precursors, often a solution of water with a salt of the metal to be deposited. Some plating processes are driven only by reagents in the solution (usually for noble metals), but the most important process is electroplating. It was not commonly used in semiconductor processing, but has resurfaced with the use of chemical-mechanical polishing techniques.
  1. Conventional CVD coating processing requires a metal compound that will volatilize at a low temperature and decompose to a metal when it contacts the substrate at higher temperature. An example of CVD is the nickel carbonyl ($NiC0_4$) coating as thick as 2.5 mm on glass windows and containers to make them explosion or shatter resistant.
  2. Diamond CVD coating processing is used to increase the surface hardness of cutting tools. The process is performed at the temperatures higher than 700ºC which softens most tool steels. Thus, the application of diamond CVD is limited to materials which do not soften at this temperature, such as cemented carbides.
  3. Plasma-assisted CVD coating processing is performed at lower temperature than diamond CVD coating. Diamond coatings or silicon carbide barrier coatings are applied on plastic films and semiconductors, including sub-¼μm semiconductors.

- *Chemical solution deposition* uses a liquid precursor, usually a solution of organometallic powders dissolved in an organic solvent. This is a relatively inexpensive, simple thin film process that is able to produce stoichiometrically accurate crystalline phases.

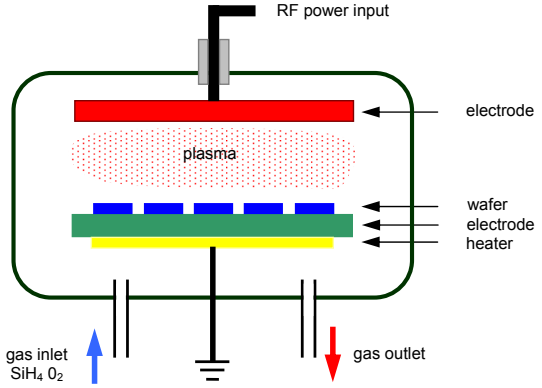


Figure 1.8. *Typical PECVD reactor.*

- Chemical vapour deposition generally uses a gas-phase precursor, often a halide or hydride of the element to be deposited. In the case of metal-organic CVD, an organometallic gas is used.

  The two most important CVD technologies are *Low Pressure CVD* (LPCVD) and *Plasma Enhanced CVD* (PECVD). The key features are:

  1. The LPCVD process produces layers with uniformity of thickness and material characteristics. The main processing problems are the high deposition temperature, greater than 600°C, and the relatively slow deposition rate. The PECVD process can operate at lower temperatures, down to 300° C, due to the extra energy supplied to the gas molecules by the ionised vapour precursor, or plasma in the reactor. However, the quality of the films tend to be inferior to processes running at higher temperatures. PECVD relies on electromagnetic means (electric current, microwave excitation), rather than a chemical reaction, to produce a plasma, as shown in figure 1.8.
  2. Most PECVD deposition systems can only deposit the material on one side of the wafers, on 1 to 4 wafers at a time. LPCVD systems deposit films on both sides of at least 25 wafers, simultaneously. A schematic diagram of a typical LPCVD reactor is shown in figure 1.9. PECVD films are conformal and deposited at lower temperatures than for LPCVD, although the film is not stoichiometric, prone to cracking and peeling, with by-products formed.

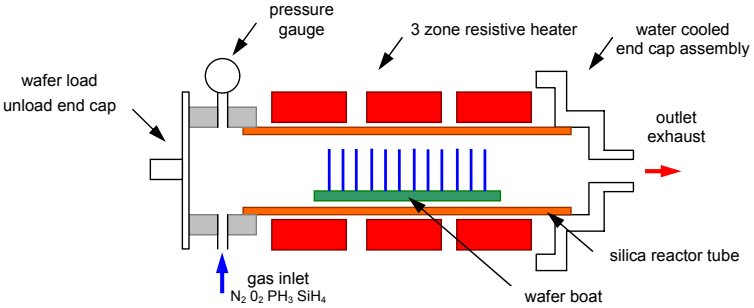


Figure 1.9. *Typical horizontal hot-wall LPCVD reactor.*

CVD is accomplished by placing the substrate wafers in a reactor chamber and heating them to a specific temperature. Controlled amounts of silicon or nitride source gases, usually carried by either nitrogen and/or hydrogen, are added to the reactor. Dopant gases may also be added if desired. A reaction between the source gases and the wafer occurs, thereby depositing the desired layer. Reaction temperatures between 500 to 1100ºC and pressures ranging from atmospheric to low pressure are used, depending on the specific deposition performed. Heating is usually accomplished with radio frequency, infrared, or thermal resistance heating. Common source gases include silane SiH_4, silicon tetrachloride $SiC\ell_4$, ammonia NH_3, and nitrous oxide N_2O. Some dopant gases that are used include arsine AsH_3, phosphine PH_3, and diborane B_2H_6. The major categories of silicon CVD are shown in the following equations.

LPCVD Atmospheric or low pressure

Medium temperature (600 to 1100°C)

Silicon Nitride, Si_3N_4: H_2 carrier gas (900 to 1100°C)

$3\ SiH_4 + 4\ NH_3 \rightarrow Si_3N_4 + 12\ H_2$

Poly Silicon, Poly-Si, H_2 carrier gas (850 to 1000°C), N_2 carrier gas (600 to 700°C)

$SiH_4 + Heat \rightarrow Si + 2\ H_2$

Silicon Dioxide, $Si0_2$:

$SiH_4 + 4\ C0_2 \rightarrow Si0_2 + 4\ C0 + 2\ H_20$ N_2 carrier gas (500 to 900°C)

$2\ H_2 + SiC\ell_4 + C0_2 \rightarrow Si0_2 + 4\ HC\ell$ H_2 carrier gas (800 to 1000°C)

$SiH_4 + C0 \rightarrow Si0_2 + 2\ H_2$ H_2 carrier gas (600 to 900°C)

Low Temperature (< 600°C)

Silicon Dioxide, $Si0_2$ or p-doped $Si0_2$, N_2 carrier gas (< 600°C)

$SiH_4 + 2\ 0_2 + Dopant \rightarrow Si0_2 + 2\ H_20$

Silicon Nitride, Si_3N_4, N_2 carrier gas (600 to 700°C)

$3\ SiH_4 + 4\ NH_3 \rightarrow Si_3N_4 + 12\ H_2$

$3\ SiH_4 + 2\ N_2O \rightarrow Si_3N_4 + 4\ H_2 + 2\ H_2O$

PECVD Low Temperature Plasma Enhance (passivation) (< 600°C), RF or reactive sputtering

Silicon Dioxide, $Si0_2$:

$SiH_4 + 2\ 0_2 \rightarrow Si0_2 + 2\ H_20$ $SiH_4 + 0_2 \rightarrow Si0_2 + 2\ H_2$

Silicon Nitride, Si_3N_4:

$3\ SiH_4 + 4\ NH_3 \rightarrow Si_3N_4 + 12\ H_2$

$3\ SiH_4 + 2\ N_20 \rightarrow Si_3N_4 + 4\ H_2 + 2\ H_20$

1.2.2 Physical Vapour deposition (PVD)

Physical deposition uses mechanical or thermodynamic means to produce a thin film of solid. Physical deposition covers a number of deposition technologies in which material is released from a source and transferred to the substrate. Physical deposition coatings involve atom-by-atom, molecule-by-molecule, or ion deposition of various materials on solid substrates in vacuum systems. Since most engineering materials are held together by relatively high energies, and chemical reactions are not used to store these energies, physical deposition systems tend to require a low-pressure vapour environment to function properly.

The material to be deposited is placed in an energetic, entropic environment, so that particles of material escape its surface. Facing this source is a cooler surface which draws energy from these particles as they arrive, allowing them to form a solid layer. The system process is in a vacuum deposition chamber, to allow the particles to travel freely. Since particles tend to follow a straight path, films deposited by physical means are commonly directional, rather than conformal.

PVD comprises the standard technologies for deposition of metals. It is more common than CVD for metals since it can be performed with lower process risk and cheaper materials costs. The film quality is inferior to CVD, which for metals means higher resistivity and for insulators more defects and traps. The step coverage is also not as good as with CVD.

The choice of deposition method (specifically evaporation versus sputtering) may be arbitrary, and may depend more on what technology is available for the specific material.

Physical deposition includes:

- A *thermal evaporator* uses an electric resistance heater to melt the material and raise its vapour pressure to a useful range, where it starts to boil and evaporate. An atomic cloud is formed by the evaporation of the coating metal in a vacuum environment to coat all the surfaces in the line of sight between the substrate and the target (source). The vacuum allows the vapour to reach the substrate without reacting with or scattering against other gas-phase atoms in the chamber, and reduces the incorporation of impurities from the residual gas in the vacuum chamber. Only materials with a higher vapour pressure than the heating element can be deposited without contamination of the film. It is often used in producing thin, ½ µm, decorative shiny coatings on plastic parts. The thin coating, however, is fragile and wears poorly. The thermal evaporation process can also coat a thick, 1 mm, layer of heat-resistant materials, such as MCrAℓY - a metal, chromium, aluminium, and yttrium alloy, on jet engine parts. Molecular beam epitaxy is a particular sophisticated form of thermal evaporation. A schematic diagram of a typical system for e-beam evaporation is shown in figure 1.10.

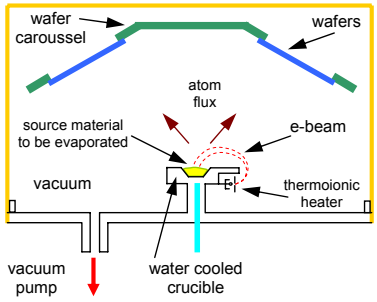


Figure 1.10. *Typical system for e-beam evaporation of materials.*

The principle is the same for all evaporation technologies, only the method used to the heat (evaporate) the source material differs. There are two popular evaporation technologies, which are e-beam evaporation and resistive evaporation, each referring to the heating method.

- An *electron beam evaporator* fires a high-energy beam from an electron gun to boil a small spot of material; since the heating is not uniform but local, lower vapour pressure materials can be deposited. The beam is usually bent through a 270° angle in order to ensure that the gun filament is not directly exposed to the evaporant flux. Typical deposition rates for electron beam evaporation range from 1 to 10 nm/s.
- In *resistive evaporation*, a tungsten boat, containing the source material, is heated electrically with a high current to make the material evaporate. Many materials are restrictive in terms of what evaporation method can be used (that is, aluminium is quite difficult to evaporate using resistive heating), which typically relates to the phase transition properties of that material.

- *Sputtering* relies on a plasma (usually a noble gas, such as Argon) to knock material from a *target* or *source*, a few atoms at a time, at a much lower temperature than with evaporation. The coatings, such as ceramics, metal alloys, organic and inorganic compounds, involve connecting the work-piece and the substance to a high-voltage DC power supply in an argon vacuum system at 10^{-2} to 10^{-3} mmHg. The gas plasma is established between the substrate (work-piece) and the target (donor) and transposes the sputtered-off target ionised atoms to the surface of the substrate. Because the target is kept at a relatively low temperature, unlike evaporation, this is a flexible deposition technique. It is especially useful for compounds or mixtures, where different components would otherwise tend to evaporate at different rates. Sputtering's step coverage is virtually conformal, producing thin, less than 3 µm, hard thin-film coatings; for example, titanium nitride (TiN) which is harder than the hardest metal. Sputtering is widely applied on cutting tools, injection moulding tools, and common tools such as punches and dies, to increase wear resistance and service life. When the substrate is non-conductive, for example, a polymer, radio-frequency (RF) sputtering is used. A schematic diagram of a typical RF sputtering system is shown in figure 1.11a. As for evaporation, the same basic principle applies to all sputtering technologies. The differences typically relate to the manner in which the ion bombardment of the target is realized. Magnetron splutter disposition, figure 1.11b, is used to deposit Aℓ, titanium, and tungsten, although CVD is difficult for alloys, Aℓ-Cu-Si.
- *Pulsed laser deposition* systems work by an ablation process. Pulses of focused laser light vaporize the surface of the target material and convert it to plasma; this plasma usually reverts to a gas before it reaches the substrate.
- *Cathodic Arc Deposition* or *Arc-PVD* is a kind of ion beam deposition where an electrical arc is created that literally blasts ions from the cathode. The arc has an extremely high power density resulting in a high level of ionization (30 to 100%), multiply charged ions, neutral particles, clusters, and macro-particles (droplets). If a reactive gas is introduced during the evaporation process, dissociation, ionization, and excitation occurs during interaction with the ion flux and a compound film is deposited.

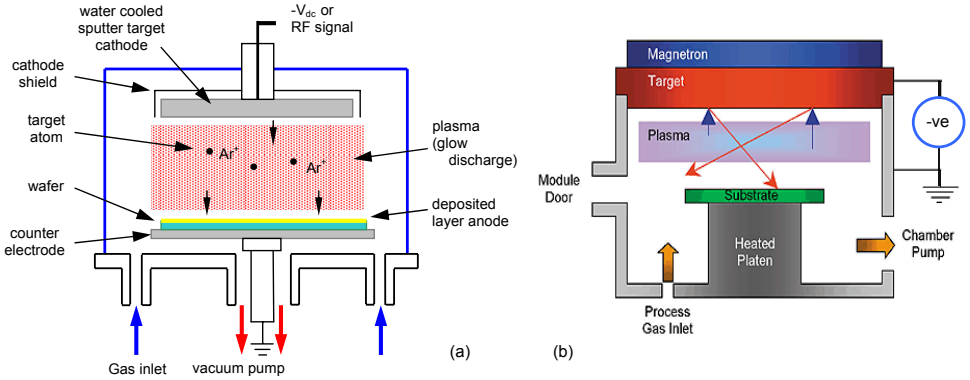


Figure 1.11. *Typical sputtering systems: (a) RF (ac) plasma and (b) dc plasma magnetron.*

1.3 Thermal oxidation and the masking process

The oxide of silicon, silica, or ***silicon dioxide*** (SiO_2), is an important planar processing ingredient and a widely used dielectric in semiconductor manufacturing because it facilitates stable insulation and

conformal passivation layers, with a high electric field breakdown strength of 10MV/cm, a resistivity of up to 10^{20} Ωcm, and a 9eV energy band gap. It is a useful and convenient deposition process, used many times on the silicon wafer surface during device fabrication. Besides the passivation and glass layer deposited over the surface of the die to protect it from mechanical damage and corrosion, dielectric layers are also used for isolating components or structures in the active circuit from each other, and as dielectric structures for MOS transistors and capacitors. Silicon dioxide is used extensively as an insulating barrier between the gate metal and channel of insulated gate semiconductor switching devices.

The formation of $Si0_2$ on a silicon surface is accomplished through a process called thermal oxidation, which is a technique that uses high temperatures, usually between 700 to 1300°C, to promote the growth rate of oxide layers. The thermal oxidation of $Si0_2$ consists of exposing the silicon substrate to a rich oxidizing atmosphere of oxidisers, 0_2 or H_20, at elevated temperature, producing oxide films with thicknesses ranging from 6nm to 1µm. Oxidation of silicon is not difficult, since silicon naturally tends to form a stable oxide even at room temperature, provided an oxidizing environment is present. The elevated temperature used in thermal oxidation accelerates the oxidation process, resulting in thicker oxide layers per unit of time. This process affords control over the thickness and properties of the $Si0_2$ layer.

The silicon wafers placed in a furnace containing oxygen gas for three to four hours at 1000°C to 1200°C, form a surface oxide layer of $Si0_2$, usually less than 1µm thick. Wet oxidation, with water added, is about 20 times faster (100nm to 120nm per hour) than dry oxidation (14nm to 25nm per hour) but the oxide quality is lower. The wafer is effectively encapsulated by silica glass, which prevents penetration by impurity atoms, except gallium atoms. Selective diffusions or implanting are made in the silicon by opening windows through the oxide by selective etching with hydrofluoric HF acid following a photo-resist lithography masking process - see section 1.5.

The oxidation furnace (or diffusion furnace, since oxidation is a diffusion process involving oxidant species), provides the heat needed to elevate the oxidizing temperature and the typical furnace consists of:

- a heating system;
- a temperature measurement and control system;
- fused quartz process tubes where the wafers undergo oxidation;
- a system for moving process gases into and out of the process tubes; and
- a loading station used for loading (or unloading) wafers into (or from) the process tubes.

The heating system usually consists of several heating coils that control the temperature around the furnace tubes. The wafers are placed in quartz glassware called boats, which are supported by fused silica paddles inside the process tube. A boat can contain many wafers, typically 50 or more. The oxidizing agent (oxygen or steam) then enters the process tube through its source end, subsequently diffusing to the wafers where the oxidation occurs. A schematic diagram of a typical wafer oxidation furnace is shown in figure 1.12.

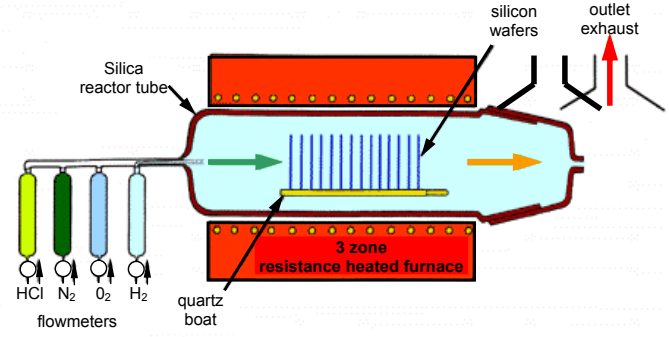


Figure 1.12. *Typical wafer oxidation furnace.*

Depending on the oxidant species used, namely 0_2 or H_20, the thermal oxidation of $Si0_2$ may either be in the form of dry oxidation, wherein the oxidant is 0_2 or wet oxidation, wherein the oxidant is H_20. The reactions for dry and wet oxidation are characterised by:

In *dry oxidation*, the oxidising agent is oxygen, 0_2:

$$Si_{(s)} + O_{2(g)} \rightarrow SiO_{2(s)} \qquad (1.25)$$

During dry oxidation, the silicon wafer reacts with the ambient oxygen (and hydrogen chloride at near atmospheric pressure), forming a layer of silicon dioxide on its surface, usually less than 100nm thick.

In *wet oxidation*, used for thick oxides, the oxidising agent is water vapour, H_20:

$$Si_{(s)} + 2H_2O_{(g)} \rightarrow SiO_{2(s)} + 2H_{2(g)} \quad (1.26)$$

Hydrogen and oxygen gases are introduced into a torch chamber where they react to form water molecules, which are then made to enter the reactor under high pressure where they diffuse toward the wafers. The water molecules react with the silicon to produce the oxide and hydrogen gas byproduct.

The oxidation reactions occur at the Si-$Si0_2$ interface, that is, silicon at the interface is consumed as oxidation takes place. As the oxide grows, the Si-$Si0_2$ interface moves into the silicon substrate. Consequently, the Si-$Si0_2$ interface will always be below the original Si wafer surface. $Si0_2$ formation therefore proceeds in two directions relative to the original wafer surface. Oxidation is the only deposition technology which actually consumes some of the substrate as it proceeds. The amount of silicon consumed by silicon dioxide formation is predictable from the relative densities and molecular weights of Si and $Si0_2$. The thickness of silicon consumed is 44% of the final thickness of the oxide formed, thus an oxide that is 100nm thick will consume about 44nm of silicon from the substrate.
Oxidation processes that have short durations (and during the first 50nm of oxide growth), may be modelled by a Linear Growth Law equation: $x_o = C\,(t + \tau)$, where x_o is the growing oxide thickness, C is the linear rate constant, t is the oxidation time, and τ is the initial time displacement to account for the initial oxide layer in situ at the start of the oxidation process. As the process proceeds the oxide growth rate decreases.

For oxidation processes that have long durations (where the oxide thickness reaches 100nm), the rate of oxide formation may be modelled by a Parabolic Growth Law equation: $x_o^2 = B \times t$, where x_o is the growing oxide thickness, B is the parabolic rate constant, and t is the oxidation time. This equation shows that the oxide thickness grown is proportional to the square root of the oxidizing time, which confirms that the oxide growth is hampered as the oxide thickness increases. This is because the oxidizing species diffusion rate decreases as it has to travel a greater distance through the oxide to the Si-$Si0_2$ interface as the oxide layer thickens.

Together the Linear and Parabolic Growth equations are known as the Linear-Parabolic Model. This oxide growth model is accurate over a wide temperature range (700 to 1,300°C), oxide thicknesses (20nm to 2µm), and oxidant partial pressures (0.2 to 2.5 atmospheres). An increase in pressure increases the oxide growth rate, but importantly, allows the temperature to be decreased for a given growth rate. For every 10 atmospheres of pressure, the temperature can be reduced by 30°C.

Oxide growth is accelerated by an increase in oxidation time, oxidation temperature, or oxidation pressure. Other factors that affect thermal oxidation growth rate for $Si0_2$ include:

- the crystallographic orientation of the wafer;
- the wafer's doping level;
- the presence of halogen impurities in the gas phase;
- the presence of plasma during growth; and
- the presence of a photon flux during growth.

Other uses for dielectric layers include:

- masking for diffusion and ion implant processes;
- diffusion from doped oxides;
- overcoating of doped films to prevent dopant loss;
- gettering of impurities (see section 1.12); and
- mechanical and chemical protection.

There are other commonly-used dielectric and isolation materials besides $Si0_2$.
Silicon dioxide doped with phosphorus (commonly referred to as P-glass, phospho-silicate glass, or PSG) is used because it inhibits sodium impurity diffusion and exhibits a smooth topography. Adding boron to PSG results in boro-phospho-silicate glass. BPSG, flows at lower temperatures than PSG, 850°C to 950°C for BPSG as opposed to 950°C to1100°C for PSG.
Polysilicon $Si0_2$ with enough oxygen content is also semi-insulating and is used in circuit and surface junction passivation. Alternately, silicon nitride is an excellent moisture barrier while stoichiometric silicon nitride is used in oxidation masks and for MOS gate dielectric. These dielectric layers are usually deposited by sputtering or chemical vapour deposition (CVD). The layer material deposited depends on the processing reactants.

The oxidising process is restricted to materials that can be oxidized, and only films that are oxides or nitrides of that material are possible. Silicon nitride, like silicon dioxide, is an amorphous insulating material that is an excellent moisture and contamination barrier, highly resistant to diffusion, not prone to delamination or cracking, and forms a progressive conformal layer on silicon. The oxidant is pure

ammonia gas NH_3 or an ammonia plasma. Although vastly superior to silicon dioxide, it has a much higher dielectric constant 7.5 as opposed to 3.85 for silicon dioxide, so is not favoured for power device insulated gate oxide structures. The disadvantages of silicon nitride are thermal related, namely higher processing temperatures are needed (950 to 1200C) and the thermal expansion of silicon nitride is twice that of silicon dioxide. The relative properties of silicon dioxide, $Si0_2$, and silicon nitride, Si_3N_4, at 300K, are shown in Table 1.3.

Table 1.3: Properties of silicon dioxide (SiO_2) and silicon nitride (Si_3N_4) at 300K

Properties		$Si0_2$	Si_3N_4
Structure		amorphous	amorphous
Melting Point	°C	≈ 1600	-
Density	g/cm^3	2.2	3.1
Refractive Index		1.46	2.05
Dielectric Constant		3.9	7.5
Dielectric Strength	V/cm	10^7	10^7
Infrared Absorption Band	µm	9.3	11.5 - 12.0
Energy Gap at 300K	eV	9	≈ 5.0
Linear Coefficient of Thermal Expansion, ΔL/L/ΔT	1/K	5 x 10^{-7}	-
Thermal Conductivity at 300 K λ	W/cm-K	0.014	-
DC Resistivity at 25°C	Ohm-cm	10^{14} - 10^{16}	≈ 10^{14}
DC Resistivity at 500°C	Ohm-cm	-	2 x 10^{13}
Etch Rate in Buffered HF	nm/min	100	½ - 1

1.4 Polysilicon deposition

Polysilicon Deposition is the process of depositing a thin-film layer of polycrystalline silicon on a semiconductor wafer, similar to epi-deposition but processed at much lower temperatures.

Polysilicon, poly-Si is compatible with high temperature processing and interfaces with thermal $Si0_2$. One of its primary uses is as gate electrode material for metal-oxide type devices because it is more reliable than Aℓ. A polysilicon gate's electrical conductivity may be enhanced by depositing a metal (such as tungsten) or a metal silicide (such as tungsten silicide) over the gate. It can also be deposited conformally over steep topography. Heavily-doped poly thin-films are used in bipolar emitter structures. Lightly-doped poly-Si films may be employed as a resistor, a conductor, or as an ohmic contact for shallow junctions, with the desired electrical conductivity attained by doping the polysilicon material.
Polysilicon deposition is achieved by thermal decomposition, called pyrolysis of silane, SiH_4, inside a low-pressure reactor at a temperature of 580 to 650°C, with the deposition rate exponentially increasing with temperature. This pyrolysis process involves the basic reaction: $SiH_{4\,(g)} \rightarrow Si_{(s)} + 2H_{2\,(g)}$.
There are two common low-pressure processes for depositing polysilicon layers:

- using 100% silane at a pressure of 25 to 130 Pa; and
- using 20 to 30% silane (diluted in nitrogen) at the same total pressure.

Both of these processes can deposit polysilicon on 10 to 200 wafers per run, at a rate of 10 to 20 nm/min and with thickness uniformities of ± 5%. The critical process variables for polysilicon deposition include temperature, pressure, silane concentration, and dopant concentration. Wafer spacing and load size have only minor effects on the deposition process.
The rate of polysilicon deposition R_d increases rapidly with temperature, since it follows the Arrhenius equation:

$$R_d = R_{do} e^{-qE_a / kT} \quad (1.27)$$

where R_d is the deposition rate,
E_a is the activation energy in electron volts, eV,
T is the absolute temperature in degrees Kelvin, K,
k is the Boltzmann constant, q is the electron charge, and
R_{do} is a constant.
The activation energy for polysilicon deposition is about 1.7eV.
Based on equation (1.27), the rate of polysilicon deposition increases as the deposition temperature increases. There will be a minimum temperature at which the rate of deposition becomes faster than the rate at which unreacted silane arrives at the surface. Beyond this temperature, the deposition rate can no longer increase with temperature, since it is now hampered by the lack of silane from which the

polysilicon is being generated. Such a reaction is *mass-transport-limited*. When a polysilicon deposition process becomes mass-transport-limited, the reaction rate is dependent on reactant concentration, reactor geometry, and gas flow.

When the rate at which polysilicon deposition occurs is slower than the rate at which unreacted silane arrives, the deposition is *surface-reaction-limited*. A deposition process is then primarily dependent on reactant concentration and reaction temperature. Surface-reaction-limited processes result in excellent thickness uniformity and step coverage. A plot of the logarithm of deposition rate against the reciprocal of the absolute temperature in the surface-reaction-limited region is a straight line with slope $-qE_a/k$.

At reduced pressures, polysilicon deposition below 575°C is too slow to be practical. Above 650°C, poor deposition uniformity and excessive roughness are encountered due to unwanted gas-phase reactions and silane depletion. Pressure can be varied inside a low-pressure reactor either by changing the pumping speed or changing the inlet gas flow into the reactor. If the inlet gas is composed of both silane and nitrogen, the inlet gas flow, and hence the reactor pressure, may be varied either by changing the nitrogen flow with a constant silane flow, or changing both the nitrogen and silane flow to change the total gas flow while keeping the gas ratio constant.

Polysilicon doping, if needed, is also performed during the deposition process. The electrical characteristics of a poly-Si thin film depends on its doping. As in single-crystal silicon, heavier doping results in lower resistivity. Poly-Si is more resistive than single-crystal silicon for any given level of doping mainly because the grain boundaries in poly-Si hamper carrier mobility. Common dopants for polysilicon include arsenic, phosphorus, and boron. Polysilicon is usually deposited undoped, with the dopants introduced after deposition.

There are three ways to dope polysilicon, namely, diffusion, ion implantation, and in-situ doping. Diffusion doping consists of depositing a heavily-doped silicon glass over the undoped polysilicon. This glass will serve as the source of dopant for the poly-Si. Dopant diffusion takes place at a high temperature of 900 to 1000°C and attains the lowest resistivities. Resistance is reduced if silicides (WSi_2, $TaSi_2$, $CoSi_2$, etc.) are used and minimised (for minimum *R-C* delay) if a metal gate is used for both the gate and interconnections. Ion implant is more precise in terms of dopant concentration control and involves directly bombarding the poly-Si layer with high-energy ions. In-situ doping consists of adding dopant gases, such as phosphine, arsine or diborane, to the CVD reactant gases during the epi polysilicon deposition process. Adding phosphine or arsine results in slower deposition, while adding diborane increases the deposition rate. The deposition thickness uniformity degrades when dopants are added during deposition.

1.5. Lithography – optical and electron

The fabrication of features on silicon wafers requires that several different layers, each with a different specific pattern, be deposited on the surface one at a time, and that doping of the active regions be done in controlled amounts over small regions of precise areas. The various patterns used in depositing layers and doping regions on the substrate are defined by a process called *lithography*. One important aspect of lithography is *photoresist* processing, which is the process of covering areas that either need to be subsequently removed or retained with a light sensitive film - the photoresist. The process of material removal following a photolithographic process is known as *etching*.
Photoresist layers have two basic functions:

- precise pattern formation; and
- protection of the substrate from chemical attack during the etch process.

Performance metrics

- Resolution: minimum feature dimension that can be transferred with high fidelity to a resist film.
- Registration: how accurately patterns on successive masks can be aligned (or overlaid) with respect to previously defined patterns.
- Throughput: number of wafers that can be exposed/unit time for a given mask level.

Photoresist materials consist of three components:

- a matrix material (also known as resin), which provides body and binder for the photoresist;
- the inhibitor (also referred to as sensitizer), which is the photoactive ingredient; and
- the solvent, which keeps the resist liquid until it is applied to the substrate.

The lithography process consists of the following steps.

- Dehydration and priming;
- A layer of photoresist material is first spin-coated on the surface of the wafer;
- Soft baking;

- The resist layer is then selectively exposed to radiation such as ultraviolet light, electrons, or X-rays, with the exposed areas defined by the exposure tool, mask, or computer data.
- The photoresist layer is subjected to photo-development which removes unwanted areas of the resist-layer, exposing the corresponding areas of the underlying layer. Depending on the resist type, the development stage may remove either the exposed or unexposed areas. The areas with no resist material left on them are then subjected to additive or subtractive processes, allowing the selective deposition or removal of material on the substrate. Hard bake.
- Post-development inspection.

During development, unwanted areas in the photoresist are dissolved by the developer. When the exposed areas become soluble in the developer, a positive image of the mask pattern is produced on the resist. Such a resist is therefore called a *positive photoresist*. *Negative photoresist* layers result in negative images of the mask pattern, where the exposed areas are not soluble in the developer. Wafer fabrication may employ both positive and negative photoresists, although positive resists offer higher resolution capabilities.

1 Dehydration and priming

Prior to the application of resist to a wafer, the wafer must be free of moisture and contaminants, both of which cause a multitude of resist processing problems. Dehydration baking is performed to eliminate any moisture adsorbed by substrate surfaces, since hydrated substrates result in adhesion failures. The bake is usually performed at between 400°C to 800°C. Convection ovens are used for baking up to 400°C, while furnace tubes are used for baking up to 800°C. After dehydration baking, the wafer is coated with a pre-resist priming layer of hexamethyldisilazane with glycol-ether solvent designed to enhance the adhesion properties of the wafer. Resist coating must follow the priming, within an hour.

2 Spin Casting

Resist coating, or the process of producing a uniform, adherent, and defect-free resist film of the correct thickness over the wafer, is usually performed by *spin-coating*. In the spin casting process the material to be deposited is dissolved in liquid form in a solvent. The viscous material (a polymer like UV positive resist ortho-diazoketone) is applied to the substrate centre by spraying or spinning. Most spin-coating processes reach speeds of 2000 to 7000 rpm for a duration of 20 to 60 seconds. The thickness that can be cast on a substrate ranges from a single monolayer of molecules, by adhesion promotion, to tens of micrometres. The control on film thickness, typically 350nm to 2µm for UV exposed photoresists, can be sustained within ±10%. Thickness of the photoresist is given by

$$t = c\, S_{\%s} \left(\frac{\nu}{\omega^2 r^2} \right)^{1/3} \tag{1.28}$$

where t = thickness
c = constant
$S_{\%s}$ = fraction of solids
ν = viscosity
ω = angular velocity
r = radius

The spin casting process is illustrated in figure 1.13. Other materials such as polyimide and spin-on glass can be applied by casting. Once the solvent is evaporated, a thin film of the material remains on the substrate.

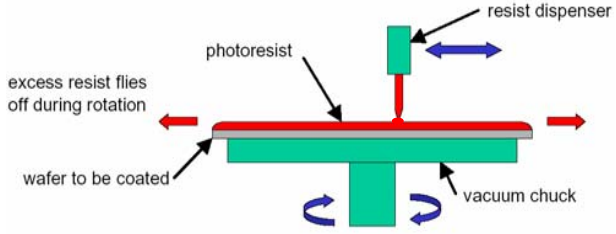


Figure 1.13. *The spin casting process as used for photoresist in photolithography.*

3 Soft baking

The soft bake causes the photoresist to cure and the remaining solvents to evaporate. Soft baking is performed using an oven (for example, convection, IR, hot plate), sometimes in a N_2 atmosphere. Soft-bake ovens provide well-controlled and uniformly distributed temperatures and a bake environment with a high degree of cleanliness. The temperature range for soft baking is between 80 to 100°C for 5 to 30 minutes, while the exposure time is established based on the heating method used and the resulting properties of the soft-baked resist.

4 Alignment and Exposure

After photoresist coating and soft baking, the wafer undergoes exposure to some form of radiation that produces a pattern image on the resist. A photomask is aligned and placed on the coated wafer with precision instruments. Two exposure forms are common.

- Optical lithography exposure which uses intense mercury arc lamp UV light and
- Electron beam lithography exposure which uses a focussed laser-type beam.

Each use specific resists, positive and negative resists, which react when exposed.
The pattern is formed on the wafer using a mask, which defines which areas of the resist surface will be exposed to radiation and those that will be covered. The chemical properties of the resist regions impinged by radiation change in a manner that depends on the type of resist used. Irradiated regions of positive photoresists become more soluble in the developer, so positive resists form a positive image of the mask on the wafer. Negative resists form a negative image of the mask on the wafer because the exposed regions become insoluble in the developer.

5 Development

Development, which is the process step that follows resist exposure, is done to leave behind the correct resist pattern on the wafer which will serve as the physical mask that covers wafer areas to be protected from chemical attack during subsequent etching, implantation, lift-off, etc. The wafers are developed with aqueous alkaline solutions of either sodium hydroxide or potassium hydroxide. Metal free developers such as tetramethylammonium hydroxide avoid sodium metal contamination. The developer is applied by either immersion, spraying, atomization or puddle developing, causing the un-polymerized areas of the photoresist to be dissolved and removed. A good development process has a duration of less than a minute, results in minimum pattern distortion or swelling, keeps the original film thickness of protected areas intact, and faithfully recreates the intended pattern.
Regardless of the developer method used, it is followed by thorough solvent rinsing with acetone and drying to ensure that the development action does not continue after the developer has been removed from the wafer surface.
After developing an additional baking process or *hard bake* is performed, at 120 to 180°C for 20 to 30 minutes, to harden the remaining photoresist to an enamel like finish. The photoresist is then ready to protect the underlying SiO_2 during etching, etc.

6 Inspection

Post-development inspection ensures that the resist processing steps have produced the desired results. This is performed using an optical microscope, although scanning electron microscope SEM and laser-based systems are also used. Items that this inspection step check include:

- use of the correct mask;
- resist film quality;
- adequate image definition;
- dimensions of critical features;
- defects and their densities; and
- pattern registration.

1.5.1 Optical Lithography

Optical Lithography refers to a lithographic process that uses visible or ultraviolet (300 to 600nm) light to form patterns on the photoresist through printing. Printing is the process of projecting the image of the patterns onto the wafer surface using a light source and a photo mask. There are three types of printing:

- contact,
- proximity, and
- projection printing, each of which will be described below.

Printers or aligners are the equipment used for printing.

Patterned masks, usually composed of glass or chromium, are used during printing to cover areas of the photoresist layer that should not be exposed to light. Development of the photoresist in a developer solution after exposure to light produces a resist pattern on the wafer, which defines which wafer areas are exposed for material deposition or removal.

Both negative and positive photoresist are used.
Positive resists also have two major components:

- a resin, ortho-diazoketone with xylene solvent; and
- a photoactive compound dissolved in a solvent, sodium or potassium hydroxide.

The photoactive compound in its initial state is an inhibitor of dissolution. Once this photoactive dissolution inhibitor is reacted with light, the resin becomes soluble in the developer.
Negative photoresists consist of two parts:

- a chemically inert, azide based, aliphatic polyisoprene rubber with xylene solvent; and
- a photoactive agent, xylene.

When exposed to light, the photoactive agent reacts with the rubber, promoting cross-linking between the rubber molecules that make them less soluble in the developer. Such cross-linking is inhibited by oxygen, so this light exposure process is usually performed in a nitrogen atmosphere.

A disadvantage of negative resists is that exposed portions swell as the unexposed areas are dissolved by the developer. This swelling, which is simply volume increase due to the penetration of the developer solution into the resist material, results in distortion of the pattern features. This swelling phenomenon limits the resolution of negative resist processes. The unexposed regions of positive resists do not exhibit swelling and distortion to the same extent as the exposed regions of negative resists. This allows positive resists to attain better image resolution.

i. *Contact printing* refers to the light exposure process wherein the photomask is pressed against the resist-covered wafer with pressure. This pressure is typically in the range of 0.05 to 0.3 atmospheres. Light with a wavelength of about 400nm is used in contact printing.
Contact printing is capable of attaining resolutions of less than 1μm. However, the contact between the mask and the resist diminishes the uniformity of the attainable resolution across the wafer. To alleviate this problem, masks used in contact printing must be thin and flexible to allow better contact over the whole wafer.
Contact printing also results in defects in both the masks used and the wafers, necessitating the regular replacement of masks (whether thick or thin) after a certain amount of use. Mask defects include pinholes, scratches, intrusions, and star fractures. Despite these drawbacks, contact printing is widely used. Good contact printing processes achieve resolutions of ¼μm or better.

ii. *Proximity printing* is another optical lithography technique. It involves no contact between the mask and the wafer, hence masks used with this technique have longer useful lives than those used in contact printing. During proximity printing, the mask is usually only 10 to 50 μm from the wafer.
Minimum line width (or critical dimension):

$$cd \cong \sqrt{\lambda g}$$

where λ = wavelength and g = gap
The resolution achieved by proximity printing is poorer than that of contact printing. This is due to the diffraction of light caused by its passing through slits that make up the pattern in the mask, and traverse across the gap between the mask and the wafer.
This type of diffraction is Fresnel diffraction, or near-field diffraction, since it results from the small gap between the mask and the wafer. Proximity printing resolution is improved by reducing the gap between the mask and the wafer and by using light of shorter wavelengths.

iii. *Projection printing* is the third technique used in optical lithography. It also involves no contact between the mask and the wafer. In fact, this technique employs a large gap between the mask and the wafer, such that Fresnel diffraction is no longer involved. Instead, far-field diffraction occurs, which is known as Fraunhofer diffraction.
Projection printing is the technique employed by most modern optical lithography equipment. Projection printers use a precision objective lens between the mask and the wafer, which collects diffracted light from the mask and projects it onto the wafer. The capability of a lens to collect and project diffracted light onto the wafer is measured by its numerical aperture, *NA*. The *NA* values of lenses used in projection printers typically range from 0.16 to 0.40.
The resolution achieved by projection printers depends on the wavelength and coherence of the incident light and the *NA* of the lens. The resolution achievable by a lens is governed by Rayleigh's criterion, which defines the minimum distance between two images for them to be resolvable. Thus, for any given *NA*, there exists a minimum resolvable dimension. Line resolution is given by:

$$\ell_m = c\frac{\lambda}{NA}$$

where c is a process dependent factor, typically 0.6 to 0.8, and
NA = numerical aperture, which is

$$NA = \overline{n}\sin\theta$$

where $\overline{n}$ is the index of refraction.

Using a lens with a higher *NA* results in better image resolution, but the penalty is that the depth of focus of a lens is inversely proportional to the square of the *NA*, so improving the resolution by increasing the *NA* reduces the depth of focus of the system. Poor depth of focus causes some points on the wafer to be out of focus, since no wafer surface is perfectly flat. Thus, projection printing aligner design compromises between resolution and depth of focus.

1.5.2 Electron Lithography

Electron Beam Lithography (EBL) refers to a lithographic process that uses a focused beam of electrons to form the patterns needed for material deposition on (or removal from) the wafer, in contrast with optical lithography which uses light for the same purpose. Electron lithography offers higher patterning resolution than optical lithography because of the shorter wavelength (sub 100nm) of the 10 to 50 keV electrons employed.

Since a small-diameter focused beam of electrons can be scanned over a surface, an EBL system does not use masks (unlike optical lithography, which uses photomasks to project the patterns). An EBL system draws the pattern over the resist wafer using the electron beam as its drawing pen. Thus, EBL systems produce the resist pattern in a sequential manner, making it slow compared to optical systems.

A typical EBL system consists of the following parts:

- an electron gun or electron source that supplies the electrons;
- an electron column that shapes and focuses the electron beam;
- a mechanical stage that positions the wafer under the electron beam;
- a wafer handling system that automatically feeds wafers to the system and unloads them after processing; and
- a computer system that controls the equipment.

The resolution of optical lithography is limited by diffraction, which is not a problem for electron lithography. This is because of the short wavelengths of the electrons in the energy range used by EBL systems. However, the resolution of an electron lithography system is constrained by electron scattering in the resist and by various aberrations in its electron optics. Resolution is as low as 10 to 25nm.

Just like optical lithography, electron lithography also uses negative (copolymer-ethyl-acrylate) and positive (polymethylmethacrylate) 100nm thick resists, which in this case are referred to as electron beam resists (or e-beam resists). E-beam resists are e-beam-sensitive materials that are used to cover the wafer according to the defined pattern. The same polymer based photo resists are used for x-ray (½ to 5nm wavelengths) exposure.
Positive electron resists produce an image that is the same as the pattern drawn by the e-beam (positive image), while negative ones produce the inverse image of the pattern drawn (negative image). Positive resists undergo bond breaking when exposed to electron bombardment, while negative resists form bonds or cross-links between polymer chains under the same situation. As a result, areas of the positive resist that are exposed to electrons become more soluble in the developer solution, while the exposed areas of the negative resist become less soluble. The positive resists form positive images - because its electron-exposed areas result in exposed areas on the wafer after dissolving in the developer. In the case of negative resists, the electron-exposed areas become the unexposed areas on the wafer, forming a negative image.

The resolution achievable with any resist is limited by two major factors:

- the tendency of the resist to swell in the developer solution and
- electron scattering within the resist.

Resist swelling occurs as the developer (isopropyl alcohol) penetrates the resist material. The resulting volume increase can distort the pattern, such that close adjacent lines merge. Resist contraction after the resist has undergone swelling can also occur during rinsing. However, this contraction is often not enough to bring the resist back to its intended form, so the swelling distortion remains even after rinsing. Unfortunately, a swelling/contraction cycle weakens the adhesion of the smaller features of the resist to the substrate, which can create undulations in narrow lines. Reducing resist thickness decreases the resolution-limiting effects of swelling and contraction.
When electrons strike a material, they penetrate the material and lose energy from atomic collisions. These collisions can cause the striking electrons to scatter, termed *scattering*. The scattering of electrons may be backward (or back-scattering, wherein electrons bounce back), but it is often forward through small angles with respect to the original trajectory. During electron beam lithography, scattering occurs as the electron beam interacts with the resist and substrate atoms. This electron scattering has two major effects:

- it broadens the diameter of the incident electron beam as it penetrates the resist and substrate; and
- it gives the resist an unintended extra doses of electron exposure as back-scattered electrons from the substrate bounce back into the resist.

Thus, scattering effects during e-beam lithography result in wider images than what can be ideally produced from the e-beam diameter, degrading the resolution of the EBL system. Closely spaced adjacent lines can add electron exposure to each other, a phenomenon known as the *proximity effect*.
The advantages of EBL are:

- Generation of submicron resist geometries
- Highly automated and precisely controlled operation
- Greater depth of focus than that available from optical lithography
- Direct patterning onto the wafer without using a mask

EBL disadvantages are

- Low throughput
- Expensive resists
- Proximity effect: backscattering of electrons irradiates adjacent regions and limits minimum spacing between features

1.6 Etching

In wafer fabrication, etching refers to a process by which material is removed from the wafer, that is, either from the silicon substrate itself or from any film or layer of material on the wafer.

The rate at which the etching process occurs is known as the *etch rate*. The etching process is said to be *isotropic* if it proceeds in all wafer directions at the same rate. If it proceeds in only one direction, then it is completely *anisotropic*. Each case is illustrated in figure 1.14.

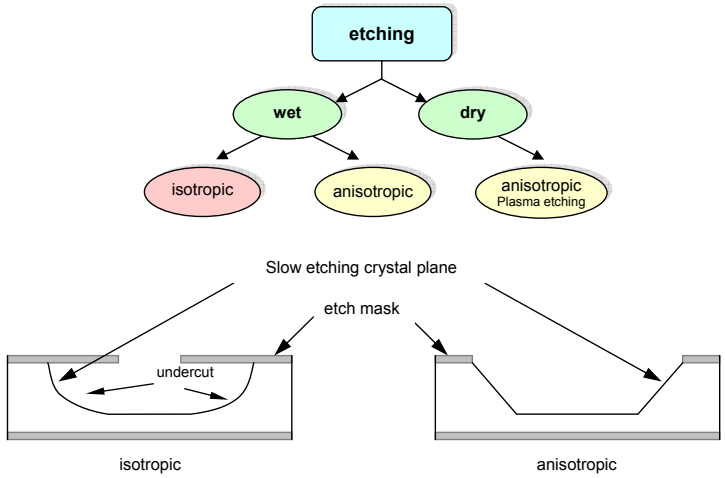


Figure 1.14. *Difference between anisotropic and isotropic wet etching.*

There are two types of etching processes:

- *Wet etching* where the material is dissolved when immersed in a chemical solution and
- *Dry etching* where the material is sputtered or dissolved using reactive ions or a vapour phase etchant.

Since etching processes generally fall between being completely isotropic and completely anisotropic, an etching process needs to be described in terms of its level of isotropy. Wet etching, or etching with the use of chemicals, is generally isotropic. On the other hand, dry etching processes that employ reactive plasmas are generally anisotropic.

1.6.1 Wet Chemical Etching

Wet etching is an etching process that utilizes liquid chemicals or etchants to remove materials from the wafer, usually in specific patterns defined by photoresist masks on the wafer. Materials not covered by these masks are etched away by the chemicals while those covered by the masks remain intact. These masks are deposited on the wafer in an earlier wafer fabrication 'lithography' step, as in section 1.5.

This is the simplest etching technology. A simple wet etching process may consist of dissolution of the material to be removed in a liquid solvent, without changing the chemical nature of the dissolved material. A wet etching process usually involves one or more chemical reactions that consume the original reactants and produce new species. Typical SiO_2 etch rates rates are 30nm/min.

A basic wet etching process involves three steps:

- diffusion of the etchant to the surface for removal;
- reaction between the etchant and the material being removed; and
- diffusion of the reaction by-products from the reacted surface.

Reduction-oxidation (redox) reactions are commonly encountered in wafer fabrication wet etching processes, that is, an oxide of the material to be etched is first formed, which is then dissolved, leading to the formation of new oxide, which is again dissolved, and so on until the material is consumed.
Wet etching is generally isotropic, that is, it proceeds in all directions at the same rate. An etching process that is not isotropic is referred to as 'anisotropic.' In semiconductor fabrication, a high degree of anisotropy is desired in etching because it results in a more faithfully copy of the mask pattern, since only the material not directly under the mask is attacked by the etchant. Isotropic etchants, on the other hand, etch away a portion of material that is directly under the mask (usually in the shape of a quarter-circle), since its horizontal etching rate is the same as its vertical rate, as shown in figure 1.10. When an isotropic etchant eats away a portion of the material under the mask, the etched film is said to have *undercut* the mask. The amount of undercutting is a measure of an etching parameter known as the *bias*. Bias is the difference between the lateral dimensions of the etched image and the masked image. Thus, the mask used in etching must compensate for whatever bias an etchant is known to produce, in order to create the desired feature on the wafer. Because of the isotropic nature of wet etching, it results in high bias values that are not practical for use in pattern images that have features measuring less than 3 microns. Thus, wafer feature patterns that are smaller than 3 microns are not wet-etched.

Another important consideration in any etching process is the *selectivity* of the etchant. An etchant not only attacks the material being removed, but the mask and the substrate (the surface under the material being etched) as well. The selectivity of an etchant refers to its ability to remove only the material intended for etching, while leaving the mask and substrate materials intact. Selectivity, S_f, is the ratio between the different etch rates of the etchant for different materials. A good etchant has a high selectivity value with respect to both the mask, S_{fm}, and the substrate, S_{fs}, that is, its etching rate for the film being etched must be much higher than its etching rates for both the mask and the substrate.

Despite the resolution limitations of wet etching, it has found widespread use because of its following advantages:

- low cost, the cost per wafer for dry etching is 1 to 2 orders of magnitude higher;
- high reliability;
- high throughput; and
- excellent selectivity in most cases with respect to both mask and substrate materials.

Automated wet etching systems add even more advantages:

- greater ease of use;
- higher reproducibility; and
- better efficiency in the use of etchants.

Wet etching has disadvantages including:

- limited resolution;
- higher safety risks due to the direct chemical exposure of the personnel;
- high cost of etchants in some cases;
- problems related to the resist's loss of adhesion to the substrate;
- problems due to the possible formation of bubbles which inhibit the etching process; and
- problems related to incomplete or non-uniform etching.

Silicon (single-crystal or poly-crystalline), which is hydrophobic, may be isotropic wet-etched using a mixture of nitric acid ($HN0_3$) and hydrofluoric acid (HF) housed in polypropylene temperature controlled baths. The nitric acid consumes the silicon surface to form a layer of silicon dioxide, which in turn is dissolved away by the HF. The over-all reaction is as follows:

$$Si + HNO_3 + 6\,HF \rightarrow H_2SiF_6 + HNO_2 + H_2 + H_2O$$

Silicon dioxide, which is hydroscopic, is isotropic wet-etched using a variety of HF based solutions or vapour. The over-all reaction for this is:

$$SiO_2 + 6\,HF \rightarrow H_2 + SiF_6 + 2\,H_2O$$

H_2SiF_6 is a gas soluble in water. Specifically, water-diluted HF with some buffering agents such as ammonium fluoride (NH_4F) is a commonly used $Si0_2$ etchant formulation. Etchants dissolve Si slower in the <111> plane because for silicon this plane has more bonds per unit area. HF is also used for silicon nitride Si_3N_4 (or H_3P0_4 at 180°C giving a 10nm/minute etch rate) and CVD oxide etching. Anisotropic etching can be achieved with K0H, with silicon and silicon dioxide.

Wet etching of aluminium and aluminium alloy layers may be achieved using slightly heated (35 to 45°C) solutions of phosphoric acid, acetic acid, nitric acid, and water. The nitric acid consumes some of the aluminium material to form an aluminium oxide layer. This oxide layer is then dissolved by the phosphoric acid and water, as more $A\ell_20_3$ is formed simultaneously in maintaining the cycle.
Other materials on the wafer may be wet-etched by using the appropriate etching solutions. Generally nitric acid, including aqua regia ($HC\ell$ plus $HN0_3$) is involved with metal ($A\ell$, Cr, Ni, Au, Ag, etc.) etching.

Table 1.4: Comparison between dry and wet etching

Etching	Wet	Dry
Method	Chemical Solutions	Ion Bombardment or Chemical Reactive
Environment and Equipment	Atmosphere, Bath	Vacuum Chamber
Advantage	• Low cost, easy to implement • High etching rate, high throughput • Good selectivity for most materials • Highly selective	• Highly selective • Capable of defining small feature size (<100 nm)
Disadvantage	• Inadequate for defining feature size • Potential of chemical handling hazards • Wafer contamination issues	• High cost, hard to implement • Low throughput • Poor selectivity • Potential radiation damage
Directionality	Isotropic (except for etching Crystalline Materials)	Anisotropic (etching mainly normal to surface)

1.6.2 Dry Chemical Etching

Dry etching refers to the removal of material, typically a masked pattern of semiconductor material, by exposing the material to a bombardment of ions (usually a plasma of nitrogen, chlorine and boron trichloride) that dislodge portions of the material from the exposed surface. Unlike with many of the wet chemical etchants used in wet etching, the dry etching process typically etches directionally or anisotropically.
Dry etching does not utilize any liquid chemicals or etchants to remove materials from the wafer, generating only volatile by-products in the process. Dry etching may be realised by any of the following:

- through chemical reactions that consume the material, using chemically reactive gases or plasma;
- physical removal of the material, usually by momentum transfer; or
- a combination of both physical removal and chemical reactions.

Drying etching is used due to its ability to better control the etching process and reduce contamination levels.
The dry etching technology can split in three separate classes:

- reactive ion etching (RIE);
- sputter etching; and
- vapour phase etching.

Plasma etching is an example of a purely chemical dry etching technique. While physical sputtering and ion beam milling are examples of purely physical dry etching techniques. Lastly, reactive ion etching is an example of dry etching that employs both physical and chemical processes.
Like wet etching, dry etching also follows the resist mask patterns on the wafer, that is, it only etches away materials that are not covered by mask material, and are therefore exposed to its etching species, while leaving areas covered by the masks almost intact. These masks were previously deposited on the wafer by a lithography fabrication step - see section 1.5.

i. Plasma etching

Plasma etching, which can etch virtually any substrate compatible material, is a purely chemical dry etching technique that consists of the following steps:

- generation of reactive species in a plasma;
- diffusion of these species to the surface of the material being etched;
- adsorption of these species on the surface;
- occurrence of chemical reactions between the species and the material being etched, forming volatile by-products;
- desorption of the by-products from the surface; and
- diffusion of the desorbed by-products into the bulk of the gas.

Two kinds of plasma etching reactor systems are in use — the barrel (cylindrical), and the parallel plate (planar), as shown in figure 1.15. Both reactor types operate on the same principles and vary primarily in configuration only. The typical reactor consists of a vacuum reactor chamber made usually of aluminium, glass, or quartz. A radio-frequency (RF) energy source is used to activate fluorine-based or chlorine-based gases which act as etchants. Wafers are loaded into the chamber, a pump evacuates the chamber (10^{-1} to 5 Torr), and the reagent gas is introduced. The RF energy ionizes the gas and forms the etching plasma, which reacts with the wafers to form volatile products which are pumped away.
The reactant gas is $CF + 0_2$ for silicon, C_2F_6 for $Si0_2$ and $CF_4 + Ar$ for Si_3N_4. Similar reactants are used to etch metals, for example $CC\ell_4$ is used to etch aluminium $A\ell$, and chromium Cr.
The desorption of the reaction by-products from the material surface being plasma etched is as important as the occurrence of the chemical reactions that consume the material. If such desorption does not occur, then etching cannot occur even if the chemical reactions have been completed.

The selectivity of the species used in dry etching that employs chemical reactions is important. Selectivity refers to the ability of the reactive species to etch away only the material intended for removal, while leaving all other materials intact. The species used must not attack the mask material over the material being etched as well as the material beneath it.

The reactive species used in dry chemical etching must be selected to meet the following criteria:

- high selectivity against etching the mask material over the layer being etched;
- high selectivity against etching the material under the layer being etched;
- high etch rate for the material being removed; and
- excellent etching uniformity.

They should also allow a safe, clean, and automation-ready etching process.

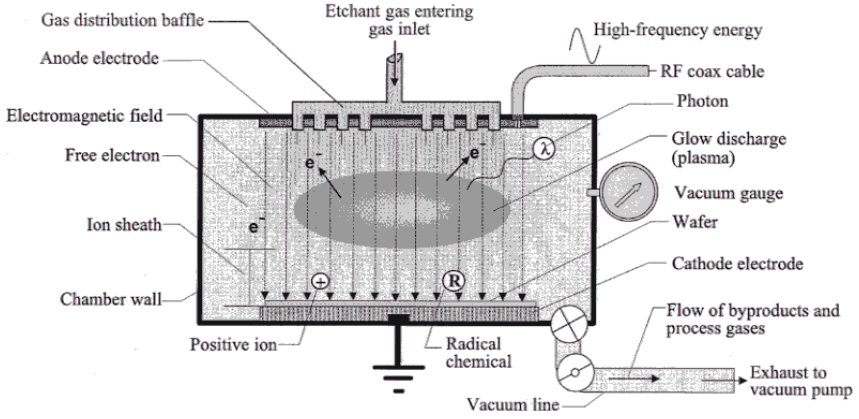


Figure 1.15. *Typical dry plasma etching system.*

Another important consideration in any etching process is its anisotropy, or property of etching in one direction only. A completely anisotropic etching process that removes material in the vertical direction only is desirable, since it will follow the mask patterns on the wafer faithfully, leaving any material covered by mask material basically untouched. Most etching techniques employing purely chemical means to remove the material (whether through wet or dry etching) do not exhibit high anisotropy. This

is because chemical reactions can and do occur in all directions. Thus, chemical reactions can attack in the horizontal direction and consume a portion of the material covered by the mask, termed *undercutting*.
If maximum anisotropy is of utmost concern, then dry etching techniques that employ physical removal of material must be considered. One such technique is physical sputtering, which involves purely physical removal of material by bombarding it with highly energetic but chemically inert species or ions. These energetic ions collide with atoms of the material as they hit the material's surface, dislodging surface atoms in the process.

ii. Reactive ion etching

Reactive ion etching (RIE), which is sometimes referred to as reactive sputter etching, is a combination of chemical and physical etching, and consists of bombarding the exposed surface material to be etched with highly energetic chemically reactive ion species. These species are usually oxidizing and reducing agents produced from process gases that have been ionized and fragmentized by a glow discharge. Such high-energy ion bombardment dislodges atoms from the material (just like purely physical sputtering), in effect achieving material removal by sputtering.

In addition to sputter-removal, the bombarding ions used in RIE are chosen to chemically react with and remove the exposed material being bombarded to produce highly volatile reaction by-products that can be pumped out of the system. This is the reason why RIE is widely used in wafer fabrication - it achieves the required anisotropy (by means of sputter-removal) and the required selectivity (through chemical reactions). Table 1.5 presents some examples of the process gases employed in the reactive ion etching of common wafer related materials.

Table 1.5: Examples of gases used in the RIE of common wafer materials

Material to be Etched	Examples of Gases used in the RIE	mask	selectivity	Etch rate A/min	Etch product
Polysilicon/silicon, Si	$CF+0_2$; CF_4; SF_6; $C\ell_2$; $CC\ell_3F$; $BC\ell_2+C\ell_2$	Resist (Cr, Ni, $A\ell$)	20:1 40:1	500	$SiC\ell_2$, $SiC\ell_4$
Silicon dioxide, $Si0_2$	C_2F_6; CF_4; SF_6; HF, CF_4+H_2; CHF_3+0_2	Resist (Cr, Ni, $A\ell$)	10:1 30:1	200	SiF_4
Silicon nitide, Si_3N_4	$CF_4+Ar/0_2/H_2$; CHF_3	Resist (Cr, Ni, $A\ell$)	10:1 20:1	100	SiF_4
$A\ell$; $A\ell$ doped with Si, Cu, Ti	$CC\ell_4$; $CC\ell_4+C\ell_2$; $BC\ell_3$; $BC\ell_3+C\ell_2$; $C\ell_2$	Resist SI_3N_4	40:1	300	$A\ell_2C\ell_6$, $A\ell C\ell_3$
Tungsten W, V, Ti, Ta, Mo	Fluorinated gases, CF_4				WF_6 ….
Refractory Silicides	Fluorinated plus chlorinated gases (with or without 0_2)				SiF_4
Resist / polymer	0_2	SI_3N_4 (Cr, Ni)	50:1	500	

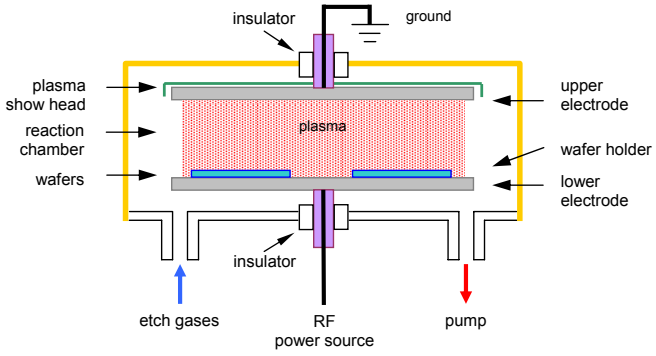


Figure 1.16. *Typical parallel-plate reactive ion etching system.*

In RIE, the substrate is placed inside a reactor in which several chemically reactive gases (CF_4 or $CC\ell_4$) are introduced at low pressure (10^{-4} to 10^{-3} Torr). A plasma is struck, by electrical discharge, in the gas mixture using an RF power source, stripping the gas molecules down to ions. The ions are accelerated towards, and reacts at, the surface of the material being etched, forming another gaseous material, which is removed by the low pressure in-line vacuum pressure. This is the chemical part of reactive ion etching. There is also a physical part which is similar in nature to the sputtering deposition process. If the ions have high enough energy, a few hundred eV, they can knock atoms out of the material to be etched without a chemical reaction. It is a complex task to develop dry etch processes that balance chemical and physical etching, since there are many parameters to adjust. By changing the balance it is possible to influence the anisotropy of the etching, since the chemical part is isotropic and the physical part is highly anisotropic, the combination can form sidewalls that have shapes from rounded to vertical. A schematic of a typical reactive ion etching system is shown in figure 1.16.

A subclass of RIE which continues to grow rapidly in popularity is deep RIE. In this process, etch depths of hundreds of microns can be achieved with almost vertical sidewalls. Two different gas compositions are alternated in the reactor. The first gas composition creates a polymer on the surface of the substrate, and the second gas composition etches the substrate. The polymer is immediately sputtered away by the physical part of the etching, but only on the horizontal surfaces and not the sidewalls. Since the polymer only dissolves very slowly in the chemical part of the etching, it builds up on the sidewalls and protects them from etching. As a result, etching aspect ratios of 50 to 1 can be achieved. The process can easily be used to etch completely through a silicon substrate, and etch rates are 3 to 4 times higher than wet etching.

iii. Sputter etching

Sputter etching (ion milling) is essentially RIE without reactive ions. The systems used are similar in principle to sputtering deposition systems. The difference is that substrate is subjected to the ion bombardment instead of the material target used in sputter deposition, as shown in figure 1.17.
The wafer to be etched is attached to a negative electrode, or target, in a glow-discharge circuit. Positive argon ions bombard the wafer surface, resulting in the dislocation of the surface atoms. Power is provided by an RF energy source.
Targeting the layer to be etched with incident ions that are perpendicular to its surface ensures that only the uncovered material is removed. Unfortunately, such a purely physical process is also non-selective, that is, it also attacks the mask layer covering the material being etched, since the mask is also directly hit by the bombarding species. For this reason, physical sputtering is not popular as a dry etching technique for wafer fabrication.
A good balance between isotropy and selectivity may be achieved by employing both physical sputtering and chemical means in the same dry etching process, such as reactive ion etching.

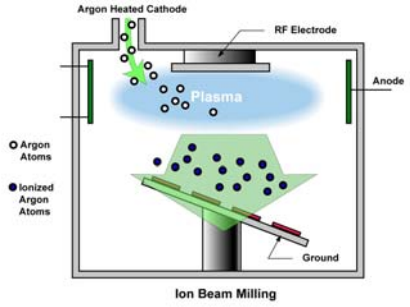


Figure 1.17. *Typical sputter etching system.*

iv. Vapour phase etching

Vapour phase etching is another dry etching method, which can be performed with simpler equipment than what RIE requires. In this process, the wafer to be etched is placed inside a chamber, into which one or more gases are introduced. The material to be etched is dissolved at the surface in a chemical reaction with the gas molecules. The two most common vapour phase etching technologies are silicon dioxide etching using hydrogen fluoride (HF) and silicon etching using xenon diflouride (XeF_2), both of which are isotropic in nature. Care must be taken in the formation of a vapour phase process to not have bi-products form in the chemical reaction that condense on the surface and interfere with the etching process.
Dry etching technology is expensive to run compared to wet etching. If feature resolution in thin film structures or vertical sidewalls for deep etchings in the substrate are required, consider dry etching.

1.7 Lift-off Processing

In semiconductor wafer fabrication, the term *lift-off* refers to the process of creating patterns on the wafer surface through an additive process, as opposed to the more familiar patterning techniques that involve subtractive processes, such as etching. Lift-off is used in patterning metal films for interconnections.
Lift-off consists of forming an inverse image of the pattern desired on the wafer using a stencil layer, which covers certain areas on the wafer and exposes the rest. The layer to be patterned is then deposited over the stencilled wafer. In the exposed areas of the stencil, the layer material is deposited directly on the wafer substrate, while in the covered areas, the material is deposited on the stencil film.
After the layer material has been deposited, the wafer is immersed in a liquid that dissolves the stencil layer. Once the stencil is dissolved by the liquid, the layer material lifts off (hence the term *lift-off*), leaving behind the layer material that was deposited directly onto the wafer substrate itself, which forms the final pattern on the wafer.
The lift-off process as a patterning technique offers the following advantages:

- composite layers consisting of several different materials may be deposited one material at a time and then patterned with a single lift-off;
- residues that are difficult to remove are prevented in the absence of etching of the patterned layer;
- sloped side walls become possible, resulting in good step coverage.

On the other hand, the main disadvantage of the lift-off process is the difficulty of creating the required stencil patterns for successful lift-off.

Materials that are used as stencil film for lift-off include:

- a single photoresist layer;
- two photoresist layers;
- a photoresist-aluminium-photoresist layer;
- polyimide/molybdenum layer;
- polyimide/polysulphone/$Si0_2$ layer; and
- inorganic dielectric-photoresist layer.

The key to successful lift-off is to ensure the existence of a distinct break between the layer material deposited on top of the stencil and the layer material deposited on the wafer substrate. Such a separation allows the dissolving liquid to reach and attack the stencil layer. One technique to create such 'breaks' is cold evaporation over steep steps.

1.8 Resistor Fabrication

Circuits built on semiconductor wafers may require resistive components. There are several types of resistors built on wafers:

- diffused resistors;
- ion-implanted resistors;
- thin-film resistors; and
- polysilicon resistors.

Diffused resistors are fabricated through p-type diffusion into an n-type background, which is usually accomplished simultaneously with base diffusion. The sheet resistance of a base diffusion is usually 100 to 200 Ohms per square (see example 1.1). As such, the use of base diffusion results in good layout proportions for resistors ranging in value from 50 to 10K Ohms. A diffused resistor is isolated from its background by the contact potential of its corresponding p-n junction or by a high reverse voltage if the tub is biased properly. Such practice allows several diffused resistors to be built in a single tub, resulting in savings in chip area.
Ion-implanted resistors exhibit sheet resistances that are as high as 5 kilo-Ohms per square, allowing significant reductions in chip area requirements of high resistance-valued resistors. Ion-implanted resistors are fabricated by first forming two base diffusions and then ion-implanting the resistor between them. Contacts are then formed on the base diffusions. Ion-implanted resistors are suited for low-power digital and linear circuits because of their high sheet resistance.
Thin-film resistors offer high precision and stability. They are fabricated by vacuum evaporation or sputtering of thin films of resistive materials directly on the substrate oxide layer. Materials used for thin-film resistors include nichrome, sichrome, and a variety of refractory silicides. These materials exhibit good adhesion on the oxide as thin films, and are usually built with a film thickness of about 10nm to 100 nm. Thin film resistance can be adjusted precisely by laser trimming.
Polysilicon resistors are fabricated from undoped polysilicon films that are deposited onto the wafer. These are then implanted with the correct type and amount of impurity (n or p type), and then annealed at about 600 to 1000°C. Polysilicon resistors are used when high resistance is needed but wide tolerances are acceptable.

1.9 Isolation Techniques

The individual components and regions that comprise the circuit on a monolithic die need to have electrical isolation from each other in order to function. The most common techniques used for achieving component isolation during wafer fabrication are:

- by employing reverse-biased p-n junctions;
- through what is known as mesa isolation;
- by wafer bonding to an insulating substrate;
- by oxide isolation;
- by trenching; and
- through a combination of any of these processes.

A reverse-biased p-n junction has an extremely low leakage current, so is use as an isolation technique during wafer fabrication. By doping two adjacent regions with opposite types of conductivity and providing them with adequate reverse biasing, they become effectively isolated from each other. Under such a situation, the coupling between the regions is only capacitive, which becomes an issue at high frequencies.
Another technique for achieving component isolation is known as *mesa isolation*. This involves the building of the components on an active film that was grown on an insulating (or semi-insulating) film, and then etching moats around the components. This results in the components becoming individual 'islands', or 'mesas', hence the name 'mesa isolation' given to this isolation technique. Circuits fabricated on silicon on insulators, as well as those made on epitaxial GaAs over semi-insulating GaAs substrate, are examples of applications of mesa isolation.
Wafer bonding to an insulative substrate may be considered as a variant of mesa isolation. This isolation technique takes advantage of the fact that any two flat, smooth, clean, and hydrophilic surfaces can be bonded at ambient temperature without the use of external forces. Wafer bonding can be applied to widely dissimilar materials. Once the moats are etched around the 'mesas', isolation is provided by the insulating substrate.

Oxide isolation techniques consist of a series of material deposition and removal steps that lead to the formation of active single-crystal tubs that are surrounded by an oxide layer. Such oxide layers, once formed, provide near-perfect isolation between the active tubs.

Trenching is a process wherein anisotropic wet etching or reactive ion etching is employed to dig a trench around the active region. The trench is then filled with isolating material. Planarization is performed after filling the trenches, see section 1.11.

1.10 Wafer Cleaning

There are a number of wafer cleaning techniques or steps employed to ensure that a semiconductor wafer is always free of contaminants and foreign materials as it undergoes the wafer fabrication process. Different contaminants have different properties, and therefore have different requirements for removal from the wafer.

Basic Wet Concepts of Cleaning
Remove organic contamination and particles by oxidation:
$NH_40H:H_20_2:H_20$ = 1:1:5 to 1:2:7 at 70 to 90°C.
Remove metal contamination by forming a soluble complex:
$HC\ell:H_20_2:H_20$ =1:1:6 to 1:2:8 at 70 to 90°C.

Removing Metal Contamination
Wet cleaning process is the most effective method for removing metallic contamination:
HF: 0.5%, H_20_2: 10% clean, due to high oxidizing mechanism of H_20_2.
Ca contamination causes a rough surface and a defect density in the oxide:
The threshold value for Ca contamination is 10^9 atoms/cm^2.

Removal Organic Contamination
Photoresist removal by Plasma and Wet cleaning ($H_2S0_4:H_20_2$ = 3:1 to 4:1 at 120 to 130°C).
Depletion of H_20 due to high wafer temperature, 120°C, causes unstable process control. The alternative is to add ozone in water, which can be used as a strong oxidizing agent that decomposes organic impurities. The oxide thickness increases as the immersion time increases and with the concentration of ozone.

Photoresist Stripping
Resist stripping, is the removal of unwanted photoresist layers from the wafer. Its objective is to eliminate the photoresist material from the wafer as quickly as possible, without allowing any surface materials under the resist to be attacked by the chemicals used. Resist stripping is classified into:

- organic stripping;
- inorganic stripping; and
- dry stripping.

Organic stripping employs organic strippers, which are chemicals that break down the structure of the resist layer. Organic strippers are phenol-based, but their short pot-life and difficulties with phenol disposal make low-phenol or phenol-free organic strippers more common. Glycol ethers or Na0H and silicates are used for positive resists.
Wet inorganic strippers, also known as oxidizing-type strippers, are used for inorganic stripping, usually to remove photoresist from non-metallised wafers, as well as post-baked and other hard-to-remove resists. Inorganic strippers are solutions of sulphuric acid H_2S0_4 and an oxidant (such as ammonium persulphate $Na_2S_20_8$, hydrogen peroxide H_20_2, or chromic $Cr0_3$), heated to about 125°C.
Dry stripping or *plasma ashing*, pertains to the removal of photoresist by dry etching using plasma etching equipment. The wafers are placed into a chamber under vacuum, and oxygen is introduced and subjected to radio frequency (13.56MHz) power which creates oxygen radicals. The radicals react with the resist to oxidize it to water, carbon monoxide, and carbon dioxide. The ashing step is usually done to remove the top layer or skin of the resist, then additional wet or dry etching processes can be used to strip away the remaining resist. Its advantages over wet etching with organic or inorganic strippers include better safety, absence of metal ion contamination, decreased pollution issues, and a lower tendency to attach to underlying substrate layers.

Chemical Removal of Film Contaminants
Chemically bonded films of contaminant material can be removed from a wafer surface by chemical cleaning. Chemical cleaning comes in various forms, depending on the nature of the film contaminants that need to be removed and from which type of surface. Chemical cleaning is performed with a series of acid (HF, H_2S0_4, H_20_2, $HC\ell$, and $HN0_3$) and rinse baths involving de-ionised water. As an example, removal of film contaminants from a wafer with nothing but thermally grown oxide may consist of the following steps:

- preliminary cleaning,
- removal of residual organic contaminants and some metals,
- stripping of the hydrous oxide film created by the previous step,
- desorption of atomic and ionic contaminants, and
- drying.

Storage of cleaned wafers should be avoided but, if necessary, must be done using closed glass containers inside a nitrogen dry box.

Sputter Etching of Native Oxide Films
A thin oxide layer grows over silicon or aluminium when exposed to air, forming $Si0_2$ and $A\ell_20_3$, respectively. These native oxide layers need to be removed from specific areas, because they exhibit adverse effects such as higher contact resistance or hampered interfacial reactions. In-situ sputter or plasma etching are the techniques commonly utilized to remove such native oxides from the wafer. *In-situ* means performing the sputter or plasma etching in the same vacuum environment where the overlying layer will be deposited.

Elimination of Particulates
Wafer contamination with insoluble particulates is a common problem. There are two frequently utilized techniques for removing particulates from a wafer:

- ultrasonic scrubbing and
- a combination of mechanical scrubbing and high-pressure spraying.

Ultrasonic scrubbing consists of immersing the wafer in a liquid medium, which is supplied with ultrasonic energy. The sonic agitation causes microscopic bubbles to form and collapse, creating shock waves that loosen and displace particles. Ultrasonic scrubbing requires a filtration system that removes the particles from the bath as they are detached. One drawback of ultrasonic scrubbing is that it can cause mechanical damage to substrate layers.
Mechanical scrubbing employs a brush that rotates and hydroplanes over a solvent applied on the wafer surface. This means that the brush does not actually contact the wafer, but the solvent moved by the rotating brush dislodges particles from the wafer surface. Simultaneously, high-pressure spraying of the wafer surface with a de-ionised water jet helps in clearing the wafer surface of particulate contamination.

1.11 Planarization

Planarization is the process of improving the flatness or planarity of the surface of a semiconductor wafer through various methods known as planarization techniques.
The starting raw wafers for semiconductor device fabrication are ideally flat or planar. However, as the wafer goes through the various device fabrication steps, layers of different materials, shapes, and depths are deposited over the wafer surface through different growth and deposition techniques. Also, portions of materials already deposited on the wafer need to be removed at different processing stages. This series of material growth, deposition, and removal steps decreases the flatness or planarity of the wafer surface.
Device fabrication that involves multi-layers of metallization aggravates the problem of wafer non-planarity. Narrower metallisation has necessitated the need for thicker metallization in order to meet the current requirements of the device. Modern fabrication techniques that increase the number of metal layers on the wafer while decreasing the width of the metal interconnects increases the problem of wafer non-planarity.
A decrease in the flatness of the wafer's surface introduces at least two problems to device fabrication. First, ensuring ample step coverage of fine lines so that no breaks in the continuity of the lines arise becomes more difficult as the wafer becomes less flat. Second, progressive loss of planarity eventually makes the imaging of fine-featured patterns on the wafer problematic.
There are several planarization techniques used in wafer fabrication. There are two categories for planarization techniques, namely, local planarization and global planarization. Local planarization refers to smoothing techniques that increase planarity over small areas. Global planarization consist of techniques that decrease long-range variations in wafer surface topology, especially those that occur over the entire image field of the stepper.

Planarization techniques include:

- oxidation;
- chemical etching;
- taper control by ion implant damage;
- deposition of films of low-melting point glass;
- resputtering of deposited films to smooth them out;
- use of polyimide films;
- use of resins and low-viscosity liquid epoxies;
- use of spin-on glass materials;
- sacrificial etch-back; and
- mechanical-chemical polishing of the wafer.

1.12 Gettering

Gettering is defined as the process of removing device-degrading impurities from the active regions of the wafer. Gettering, which can be performed during crystal growth or in subsequent wafer fabrication steps, is an important factor for enhancing the yield of semiconductor devices. The general mechanism by which gettering removes impurities from device regions can be described by the following steps:

- the impurities to be gettered are released into solid solution from whatever precipitate they are in;
- they undergo diffusion through the silicon; then
- they are trapped by defects such as dislocations or precipitates in an area away from active regions.

There are two general classifications of gettering, namely, extrinsic and intrinsic.

Extrinsic gettering refers to gettering that employs external means to create damage or stress in the silicon lattice in such a way that extended defects needed for trapping impurities are formed. These chemically reactive trapping sites are usually located at the wafer backside, which are removed if the backside is subsequently lapped to produce a thinner substrate.
Several methods have been used to achieve external gettering. For instance, the introduction of mechanical damage by abrasion, grooving, or sandblasting can produce stresses at the backside of a wafer, which when annealed create dislocations that tend to relieve these stresses. These locations serve as gettering sites. The main drawback of this method is its tendency to initiate and propagate wafer backside microcracks that may compromise the mechanical strength of the wafer.
Diffusing phosphorus into the wafer backside is another technique used for external gettering. Phosphorus diffusion into silicon results in phosphorus vacancies or dislocations that serve as trapping sites for impurity atoms, such as gold. Another effect of P diffusion is the creation of Si-P precipitates, which are capable of removing Ni impurities through interactions between Si self-interstitials and Ni atoms, nucleating $NiSi_2$ particles in the process.

The introduction of damage by a laser is another external gettering method. Scanning a laser beam across the wafer surface induces damage that is similar to mechanical damage, but the laser damage is controlled and cleaner. Laser subjects the irradiated areas to thermal shock, forming dislocation nests that serve as gettering sites.
Ion-bombardment to produce wafer backside damage is another method for external gettering, using high-energy ions to induce the necessary stress within the lattices of the wafer backside. Deposition of a polysilicon layer on the wafer backside has also been used for external gettering. Polysilicon layers introduce grain boundaries and lattice disorder that can act as traps for mobile impurities.

Intrinsic gettering refers to gettering that involves impurity trapping sites created by precipitating supersaturated oxygen out of the silicon wafer. The precipitation of supersaturated oxygen creates clusters that continuously grow, progressively introducing stress into the wafer.
Eventually these stresses reach the point where they need to be relieved. Dislocation loops or stacking faults are thus formed to provide the necessary stress relief. These dislocations and faults subsequently serve as trapping sites for impurities.
A basic requirement of intrinsic gettering is starting wafers that have sufficient, but not excessive oxygen levels (15 to 20 ppma).

The advantages of intrinsic gettering over extrinsic gettering are:

- it does not require subjecting the wafer to any treatment except for heating;
- its volume of impurity sink is significantly larger than that of external gettering on the wafer backside;
- its gettering regions are much closer to the device operating regions.

1.13 Lifetime control

Two basic processes have been developed to reduce the lifetime of carriers in power semiconductor devices.

- Thermal diffusion of gold or platinum or
- Bombardment of the silicon with high-energy particles such as electrons, protons and Helium.

The diffusion of gold or platinum occurs more rapidly than the diffusion of group III and V dopants, hence the precious metal is diffused at 800°C to 900°C, just prior to metallization, which is performed at a lower temperature. The higher the precious metal diffusion temperature the higher the solubility and the lower the carrier lifetime. Disadvantages of precious metal diffusion include:

- devices cannot be tested prior to or immediately after impurity diffusion and
- small temperature changes cause a wide variation in device characteristics.

In high-resistivity silicon used to fabricate high-voltage power devices, irradiation bombardment causes defects composed of complexes of vacancies with impurity atoms of oxygen and of two adjacent vacancy sites in the lattice. The advantages of the use of irradiation in order to reduce carrier lifetime in power semiconductor devices are:

- irradiation is performed at room temperature, after fabrication;
- irradiation can be accurately controlled hence a tighter distribution of electrical characteristics results;
- overdose annealing can be performed at only 400°C; and
- it is a clean, non-contaminating process.

Much attention has been focussed on proton irradiation, which has high costs and long processing scan times, but offers the most accurate and precise form of lifetime control. The electrical consequences of lifetime control are an improvement in switching speed at the expense of increased leakage and on-state voltage.

1.14 Silicide formation

Metallization refers to the metal layers that electrically interconnect the various device structures fabricated on the silicon substrate. Interconnection paths that possess low resistivities and the ability to withstand subsequent high temperature processes are critical to semiconductor manufacturing. Thin-film aluminium is the most widely used material for metallization because of its low resistivity and its adhesion compatibility with SiO_2. The resistivity of Aℓ is low enough for IC interconnection purposes, but its low melting temperature of 660°C and the low Aℓ-Si (11.3%:88.7%) eutectic temperature of 577°C, restrict subsequent processes to operating temperatures of less than 500°C. Aluminium alloys (lightly doped Aℓ) such as Aℓ-Cu (or TiN barrier metal) are preferred to pure aluminium for metallization because these inhibit problems like electromigration and junction spiking – see figure 1.18.

Aℓ metal layers are usually deposited through Physical Vapour Deposition (PVD) by sputtering.
Upon exposure to oxygen, aluminium readily forms a native thin oxide on its surface, $A\ell_2 0_3$, even at ambient temperature. The presence of such an oxide layer can increase the contact resistance of the Aℓ layer. It can also inhibit the sputtering of an Aℓ target or etching of an Aℓ thin film, resulting in processing difficulties. Aℓ can readily suffer from corrosion in the presence of a corrosive contaminant and moisture. For instance, if phosphorus-doped silicon dioxide is deposited over Aℓ lines, phosphoric acid results if moisture ingresses through the glass. The acid corrodes the Aℓ connects.

For these reasons, instead of using Aℓ, low-resistivity interconnections are usually fabricated using materials known as *refractory metal silicides* (MSi_x), which can handle much higher processing temperatures than Aℓ.
The formation of refractory metal silicides (such as WSi_2, $TiSi_2$, $MoSi_2$, and $TaSi_2$) at the wafer surface can be accomplished in four ways:

- by deposition of the pure metal onto a Si layer (which can be the single-crystal substrate or poly-crystalline Si);
- simultaneous evaporation of the silicon and the refractory metal from two sources (or co-evaporation);
- sputter-deposition of the silicide, either from a composite target or by co-sputtering; and
- chemical vapour deposition (CVD).

i. The *silicide formation* technique of directly depositing a refractory metal on a silicon surface to form the required silicide layer employs the process of direct metallurgical reaction. After the metal is deposited on the silicon, the wafer is exposed to high temperatures that promote the chemical reactions between the metal and the silicon needed to form the silicide.
In such a metallurgical reaction, metal-rich silicides generally form first, and continue to grow until all the metal is consumed. When the metal has been consumed, silicides of lower metal content start appearing, which can continue to grow by consuming the metal-rich silicides. To illustrate this with titanium Ti as the metal, TiSi is the first silicide to form on Si, typically appearing at a temperature above 500°C and peaking at 700°C. $TiSi_2$ only starts to appear at 600°C and peaks at 800°C. Beyond 800°C, TiSi would be fully converted into $TiSi_2$, at which point the system attains stability.
Silicide formation by direct metallurgical reaction consumes silicon from the substrate onto which the metal is placed. Thus, it is important that enough silicon is available when this technique is employed to form the silicide layers.

ii. *Co-evaporation*, another technique for silicide formation, consists of the simultaneous deposition of the metal and the silicon under high vacuum conditions. The metal and silicon are vaporized through one of several possible heating techniques: with an electron beam, by rf induction, with a laser, or by resistive heating. However, e-beam heating is the preferred technique because the refractory metals (Ti, Ta, Mo, W) have high melting points (1670 to 2996°C) while silicon has a low vapour pressure. With the use of two e-beam guns whose power supplies are individually controlled, the proper metal-to-Si ratio can be achieved.
Critical aspects of an evaporation process that ensure the deposition of films with repeatable properties include:

- the evaporation base pressure (must be < 1 micro-torr);
- evaporation rates;
- purity of the elements; and
- residual gases present in the evaporation chamber.

iii. The third technique for silicide formation is *sputter deposition*. Sputtering is a deposition process wherein atoms or molecules are ejected from a target material by high-energy particle bombardment so that the ejected atoms or molecules can condense on a substrate as a thin film. As in co-evaporation, the correct sputtering rates of the metal and Si must be determined and applied to ensure proper deposition of the silicide film. The step coverage of co-sputtered films are superior to those of evaporated films.
Sputter-deposition of silicides has various forms. For instance, sputtering from two targets using multi-pass sputtering systems achieves the appropriate mixture of metal and Si in a layered structure. Sintering then completes the chemical reaction between the metal and Si to form the silicide. Sputtering from a composite target (MSi_x) can also be performed, allowing better compositional control, but vulnerability to contamination issues are associated with composite targets.

iv. *Chemical vapour deposition* CVD of silicides, the fourth technique for silicide formation, involves chemical reactions between vapours to form the silicide film, and offers advantages over the other techniques, namely, better step coverage, higher purity of the deposited films, and higher wafer throughput. However, the availability of gas reactants whose chemical reactions will produce the desired silicide is necessary for CVD in silicide formation.

Silicides in Table 1.6 are highly conductive alloys of silicon and metals; contact materials in silicon device manufacturing; combine advantageous features of metal contacts (significantly lower resistivity than poly-Si) and poly-Si contacts (no electromigration), and have superior resistance to temperature than metals like Aℓ. NiSi may be preferred because of it low sintering temperature and shallow penetration into the silicon.

Table 1.6: Low resistivity silicides

silicide	composition	resistivity	sintered at	nm of Si per nm of metal
		μohm-cm	°C	
cobalt silicide	$CoSi_2$	16-20	900	3.65
titanium silicide	$TiSi_2$	13-16	900	2.25
tungsten silicide	WSi_2	60-80	1000	
tantalum silicide	$TaSi_2$	35-40	1000	
platinum silicide	PtSi	25-35	600-700	
nickel silicide	NiSi	14-20	400-700	1.85

1.15 Ohmic contact

An ohmic contact is a resistive connection which is voltage independent (and technically has a Schottky barrier height of $\Phi_b \leq 0$). Aluminium is commonly evaporated in a vacuum at near room temperature, onto the wafer surface to form a metallised electrical contact. Deposit rates of ½ μm/minute are typical. Thermal annealing reduces contact resistivity.
If the silicon is n-type, a p-n Schottky junction is formed ($\Phi_b > 0$), which is undesirable as an ohmic contact. This junction forming aspect is discussed at the end of section 3.1.4. Ohmic metal contact to both p-type and n-type semiconductors with a large bandgap, like silicon carbide or gallium nitride, is technically difficult – as introduced in section 1.20.

Metallization: Metal Deposition

After devices have been fabricated in the silicon substrate, a metallization process is the fabrication step in which proper interconnection of circuit elements is made.
Metal layers are deposited by a vacuum deposition technique on the wafer to form conductive pathways. The most common metals include aluminium, nickel, chromium, gold, germanium, copper, silver, titanium, tungsten, platinum, and tantalum. Selected metal alloys may also be used.
Aluminium is the most common metal used to interconnect ICs, both to make ohmic contact to the devices and to connect these to the bonding pads on the chip's edge. Aluminium adheres well to both silicon and silicon dioxide, can be easily vacuum deposited - since it has a low boiling point, and has high conductivity. In addition to pure aluminium, alloys of aluminium are used to form IC interconnections for different performance-related reasons. For example, small amounts of copper are added to reduce the potential for electromigration effects (in which current applied to the device induces mass transport of the metal, accumulation and depletion, depending on current direction, as shown in figure 1.18a). Small amounts of silicon also are added to aluminium metallization to reduce the formation of metal 'spikes' that occur over contact holes. Copper has a higher conductivity (ρ_{Cu} = 1.67 μΩcm, ρ_{Al}=1.65μΩcm), is more malleable, and better electromigration resistance than Aℓ, but copper tends to oxidise, corrode, is not amenable to dry etching, and adheres poorly to, and contaminates, $Si0_2$. Anywhere aluminium comes in contact with silicon, some silicon is absorbed leaving voids. Random pits result which after annealing fill with aluminium forming spikes which penetrate through to the surface shorting shallow junctions, as shown in figure 1.18b.

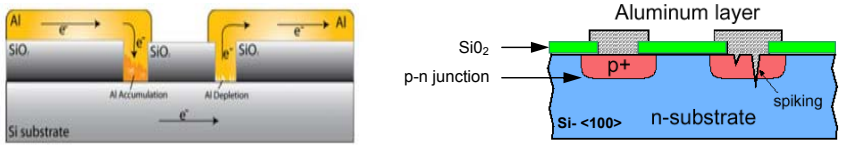


Figure 1.18. *Aluminium: (a) current direction dependant migration and (b) metallised junction spiking.*

i. *Photolithographic* techniques similar to those used during device fabrication are used to deposit conductive patterns during metallization. The metal is deposited then covered with a patterned photoresist, and subsequently etched. Alternatively, the resist is applied first, followed by deposition of the metal. The wafer is then placed in a solvent that causes swelling of the resist. As the resist swells it lifts the overlaid metal away from the wafer surface.

ii. *Etching* deposited metals requires a more aggressive chemical attack than etching $Si0_2$ during device fabrication. Consequently, the resists need to be tougher. The process of *silyation* is frequently performed to harden the photoresist before it is subjected to etching. During silyation, silicon atoms are introduced into the surface of the organic resist. This process can be accomplished with either wet or dry procedures. Most commonly, a wet bath using either hexamethyldisilazane or silazone in xylene is used.

iii. Metal layers are *vacuum-deposited* onto wafers by one of the following methods:

- Filament Evaporation;
- Flash Evaporation;
- Electron-beam Evaporation;
- Sputtering; and
- Induction Evaporation.

Filament Evaporation

Filament evaporation, also called resistive evaporation, is the simplest method, is accomplished by gradually heating a filament of the metal to be evaporated. This metal may be in one of several different forms: pellets, wire, crystal, etc. Gold, platinum and aluminium are metals typically used. The PMOS process uses aluminium wire.
The metal is placed in a basket. Electrodes are connected to either side of the basket or bell jar and a high current passed through it, causing the basket to heat by thermal resistance. As the power, and therefore heat, is increased, the metallic filament partially melts and wets the filament and as the current through the filament is increased further, the metal eventually vaporizes. In this way, atoms of aluminium break free from the filament and condense on the cooler surface of the semiconductor wafers, forming the desired metal layer on the wafers. While filament evaporation is the simplest of all metallization approaches, problems of contamination during evaporation preclude its widespread use in IC fabrication.
The procedure can be summarised as follows:

- Metal sources – pellets, are placed on filaments (tungsten, molybdenum, quartz, graphite, etc.);
- Metals are heated via a resistive filament under vacuum conditions to their melting point;
- Metal pellets give off a vapour, the atoms of which are kinetic energy dependant on temperature;
- Metal atoms travel in a straight line from the source to a sample;
- Deposition rates of order of 1 nm/s are standard; and
- Contamination may result from the filament being at least the same temperature as the source.

Flash Evaporation

Flash or partial evaporation uses the principle of thermal-resistance heating to evaporate metals.
The evaporation process requires high temperature, and low pressure; and can be divided into three steps.

- The solid aluminium metal is changed into a gaseous vapour.
- The gaseous aluminium is transported to the substrate.
- The gaseous aluminium is condensed onto the substrate.

The sources are usually either powder or thin wires. In the latter case, wire is continuously fed from a spool until it contacts the heated ceramic bar. Upon contact, the metal evaporates and is subsequently deposited on the substrate.

The evaporation process does not produce a uniform layer of aluminium across the substrate. The deposition rate changes when moving radially from the centre of the substrate.
Like filament evaporation, flash evaporation offers radiation-free coatings. This technique does offer some benefits beyond filament evaporation: contamination-free coatings, speed or good throughput of wafers, and the ability to coat materials or layers that are composite in nature.

Electron-beam Evaporation

Electron-beam evaporation, frequently called e-beam, functions by focusing an intense beam of electrons into a crucible, or pocket, in the evaporator that contains the aluminium to be deposited. As the beam is directed into the source area, the aluminium is heated to its melting point, and eventually, evaporation temperature. The benefits of this technique are speed and low contamination, since only the electron beam touches the aluminium source material. The process can be summarised as follows.

- Thermal emission of electrons from a filament source, usually tungsten, is used to heat samples to high temperatures.
- Typically, electron beams are used when the required temperatures are too high for thermal evaporation.
- Magnetic field and rastering are used to steer the beam 270° into metal source. This is done to allow shielding of the tungsten filament and to prevent contamination.
- Electrons striking metals can produce X-rays which sometimes causes damage to material layers on a wafer. An annealing stage takes care of these damage problems.

Sputtering

Aluminium sputtering is used commonly in IC metallization processes because the adhesion of the deposited metals is excellent. RF sputtering is done by ionizing inert gas particles, argon, in an electric field - producing a gas plasma, in a low-pressure or partial-vacuum atmosphere. Then the ions are directed toward the source or target - comprised of the metal to be subsequently deposited, where the energy of these gas particles physically dislodges, or sputters off, atoms of the source material. The dislodged atoms are then deposited in a thin film on the silicon substrate facing the target, figure 1.19.
The procedure can be summarised as follows.

- A parallel plate system generates high-energy ions that accelerate to bombard the source material – the material to be deposited.
- If the ion energy is high enough (typically 4 times the bond energy of source) atoms will be kicked lose, ejected - sputtered. Typical bond energies are 5eV.
- the sputtered atoms reach the substrate and the gas providing ions must be inert, that is, must not chemically react with sample substrate.
- the sputtered atoms condense and form a thin film over the substrate. Low pressures are incompatible with sputtering, thus the sample must be located close to target source.
- Insulating materials must use an RF energy source.

Sputtering is a versatile process since almost any material can be deposited by this technique, using both direct current and radio-frequency voltages. Sputtering works well for materials with extremely high melting points such as carbon, silicon, and alloys.

Induction Evaporation

Induction evaporation uses radio-frequency radiation to evaporate the metal in a crucible. The metal is then deposited as with other methods.

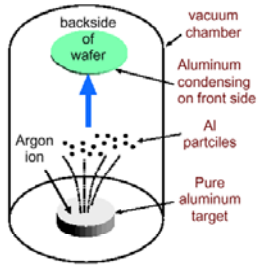


Figure 1.19. *Typical sputtering chamber.*

Table 1.7: Summary of Pros and Cons to metal layer deposition and evaporation methods

Method	Advantages	Disadvantages
E-Beam Evaporation	• high temperature materials • good for lift-off • highest purity	• some CMOS processes sensitive to radiation • alloys difficult • poor step coverage
Filament Evaporation	• simple to implement • good for lift-off	• limited source material (no high temperatures) • alloys difficult • poor step coverage
Sputter Deposition	• better step coverage of alloys • high temperature materials • less radiation damage	• possible grainy films • porous films • plasma damage and contamination
Flash Evaporation	• contamination free • composite layers	•

iv. Metal *reactive ion etching* (RIE) is commonly used to etch metal layers. This process uses a combination of physical sputtering and chemically reactive species for etching at low pressures. RIE uses ion bombardment to achieve directional etching and also a chemically reactive gas (carbon tetrafluoride, carbon tetrachloride, boron trichloride, and others) to maintain good etched layer selectivity. A wafer is placed into a chamber and given a negative electrical charge. The chamber is heated and taken to a low pressure, and then filled with a positively charged plasma of the reactive gas. The opposing electrical charges cause the rapidly moving plasma molecules to align themselves and strike the wafer surface vertically, thereby reacting with and volatizing the exposed metal layer. After etching, the remaining photoresist is stripped in a similar manner as that considered in section 1.10.

v. Alloying and Annealing. After the metallised interconnections have been deposited and etched, a final step of alloying and annealing may be performed. To perform alloying, the metallised substrate is placed in a low-temperature diffusion furnace. Usually aluminium is placed in the furnace to form a low-resistance contact between the aluminium metal and silicon substrate. Finally, either during the alloy step or following it, the wafers are exposed to a gas mixture containing hydrogen in a diffusion furnace at 400 to 500ºC. This annealing step is designed to optimize and stabilize the characteristics of the device by combining hydrogen with uncommitted atoms at or near the silicon-silicon dioxide interface.

1.16 Glassivation

Glassivation is the deposition of the final passivating layer on top of the die to protect it from mechanical damage and corrosion. This final layer is often composed of an amorphous insulating material, or glass.
Silicon nitride (Si_3N_4) is a common glassivating material because of its suitability for this purpose, as highlight by the features given in Table 1.3. Highly resistant to diffusion, it is almost impenetrable to moisture and ionic contaminants such as sodium, Na. It can also be deposited with a low residual compressive stress, making it less prone to delamination or cracking. Its interfacing with the underlying metal layers is conformal. Finally, it can be prepared with a low pinhole density. Since silicon nitride has a high dielectric constant, it is not popular as an interlayer dielectric for the simple reason that it results in a high inter-metal capacitance.
Silicon nitride can be deposited using plasma-enhanced chemical vapour deposition (PECVD) or low-pressure chemical vapour deposition (LPCVD), using silane and ammonia gases NH_3, although the former is the technique of choice for glassivation purposes because it allows a lower processing temperature. Sometimes a layer known as "p-glass" is deposited with a layer of $Si0_2$ doped with phosphorous. Phosphine gas is used as a source of phosphorous for this type of deposition.

1.17 Back side metallisation and die separation

A final processing step called back-lapping is sometimes performed. The backside of the wafer may be lapped or ground down using a wet abrasive solution under pressure. Backside metallization with a metal such as gold may be deposited on the back of the wafer with sputtering. This makes attachment of the separated die to the package easier in the final assembly.
After sorting and testing, the individual dies are physically separated. Diamond scribing, laser scribing, and diamond wheel sawing are used for die separation. Diamond scribing involves scoring a line across the wafer surface with a diamond tip. The wafer is then bent along the line, causing a fracture and separation. Laser scribing is similar except that a laser is used to score the fracture line. Diamond sawing involves wet-cutting the substrates with a high-speed circular diamond saw. Sawing may be used to either partially cut and scribe the surface, or can be used to completely cut through the wafer.

1.18 Wire bonding

Wire bonding is the process of electrically connecting the silicon to the package electrical pins or legs. Power devices use aluminium rather than gold, Au, although copper is being increasingly used.
After separation into individual dies, the functional devices are attached to a lead frame assembly. Usually the chips are attached with an epoxy material. Once attached to the lead frame, electrical connections must be provided between the die and assembly leads. This is accomplished by attaching aluminium or gold leads via thermal compression or ultrasonic welding.

The three wire bonding methodologies, as shown in figure 1.20, are:

a Wire bonding
b Flip-chip bonding
c Tape-automated bonding

a. Wire bonding
As summarised in Table 1.8, Au or Aℓ wires are wire bonded between pads and the substrate using:

- Ultrasonic,
- Thermo-sonic, or
- Thermo-compression bonding.

Ultrasonic bonding
Due to problems with thermo-compression bonding:

- Oxidation of Aℓ makes it difficult to form a good ball.
- Epoxies cannot withstand high temperatures.

Ultrasonic is a lower temperature alternative, which relies on pressure and rapid mechanical vibration to form bonds. The approach is:

i. The wire fed from a spool through a hole in the bonding tool
ii. Wire lowered into position as ultrasonic vibration at 20-60 kHz causes the metal to deform and flow.
iii. Tool raised after the bond to the package is formed.
iv. Clamp pulls and breaks wire.

Thermo-sonic bonding

- Combination of thermo-compression and ultrasonic
- Temperature maintained at approximately 150°C
- Ultrasonic vibration and pressure used to cause metal to flow to form weld
- Capable of producing 5 to10 bonds/s

Thermo-compression bonding

- Fine wire (15 to 75 mm diameter) fed from a spool through a heated capillary.
- H_2 torch or electric spark melts the wire end, forming a ball.
- Ball is positioned over the chip bonding pad, capillary is lowered, and ball deforms into a 'nail head'.
- Capillary raised and wire fed from spool and positioned over substrate; bond to package is a wedge produced by deforming the wire with the edge of the capillary.
- Capillary is raised and wire is broken near the edge of the bond.

Table 1.8: Three wire bonding processes

Wire bonding	Pressure	Temperature	Ultrasonic energy	Wire
Thermo-compression	High	300-500°C	No	Au,
Ultrasonic	Low	25°C	Yes	Au, Aℓ
Thermo-sonic	Low	100-150°C	Yes	Au

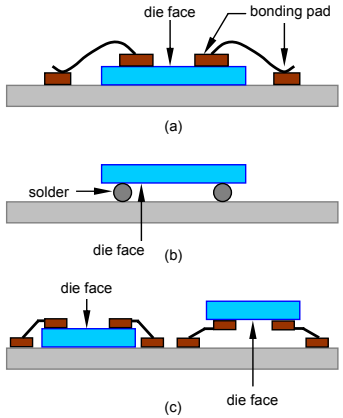


Figure 1.20. *Bonding methodologies: (a) wire; (b) basic flip-chip; and (c) tape-automated.*

b. Flip-chip bonding

Direct interconnection where die is mounted upside-down onto module or PCB, as shown in figure 1.20b. Connections made via solder bumps located over the surface of die. I/O density limited only by minimum distance between adjacent bond pads. A number of different flip-chip bonding techniques are shown in figure 1.21.

The bonding process, shown in figure 1.21c for stud bump bonding, is:

- Die are placed face down on the module substrate so that I/O pads on the chip are aligned with those on the substrate
- Solder reflow process is used to simultaneously form all the required connections,
- Drawback: bump fabrication process itself is fairly complex and capital intensive.
- Solderless flip-chip technology is another alternative; involves stencil printing of organic polymer onto a die.

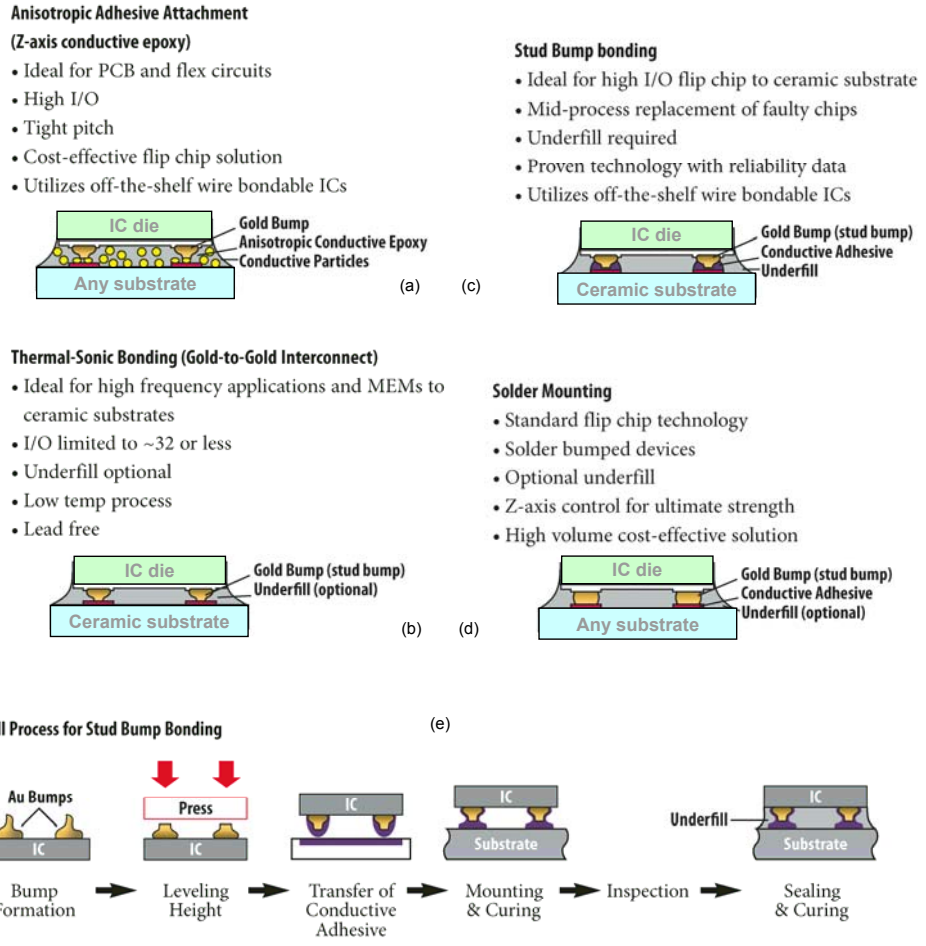


Figure 1.21. *Flip-chip bonding methodologies:*
(a) anisotropic adhesive attachment (Z-axis conductive epoxy);
(b) thermal-sonic bonding (gold-to-gold interconnect); (c) stud bump bonding;
(d) solder mounting; and (e) full process for stud bump bonding.

c. Tape-automated bonding

Tape Automated Bonding is an interconnect technology between the substrate and the die, using a prefabricated carrier with copper leads adapted to the die pads instead of single wires.
This prefabricated carrier, usually a tape consists of a perforated polyamide film, like a camera film, and of the same dimensions, which has a transport perforation and stamped openings for the die and the connection leads. On this film is then glued a copper foil, and its copper is structured by photolithography like a flexible circuit.
The advantage of this process is the creation of freestanding fingers, as seen in figure 1.20c, in the tape openings, which are then soldered or welded to bumps previously created on die pads (Inner lead bonding). The mounted die can be burned-in, tested, and afterwards punched out from the tape. No mechanical protection is needed, the bumps sealing hermetically the die, and the leads have a mechanical strength of about ten times the strength of a bonded wire.

Advantage:

- all bonds formed simultaneously, improving throughput.

Disadvantages:

- Requires multilayer solder bumps with complex metallurgy.
- A particular tape can only be used for a chip and package that matches its interconnect pattern.

1.19 Types of silicon

Silicon is the most common material used for semiconductors. After oxygen, silicon is the second-most abundant element in the earth's crust. It is not poisonous, and it is environment friendly, its waste does not represent any problems. However, to be useful as a semiconductor material, silicon must be refined to a chemical purity of better than 99.9999%. Pure silicon is not a natural state but is refined from silicon dioxides $Si0_2$ such as quartzite gravel, which is the purest silica, or crushed quartz, silicates.

1.19.1 Purifying silicon

Silicon dioxide of either quartzite gravel or crushed quartz is placed in an electric arc furnace. A carbon arc is then applied to release the oxygen, at temperatures over 1800°C, leaving the products carbon dioxide and molten silicon. The reduction to silicon is via the reaction:

$$SiO_2 + SiC \rightarrow Si + SiO + CO$$
$$SiO_2 + 2C \rightarrow Si + 2CO \qquad (1.29)$$

This process yields a metallurgical-grade, medium-grey metallic looking, 99% pure silicon.
Silicon with a one percent impurity is not useful in the semiconductor industry. Impurities of the order 10^{14} will make major changes in the electrical behaviour of silicon. Since there are about $5x10^{22}$ atoms/cm^3 in a silicon crystal, a purity of better than 1 part in 10^8 or 99.999999% pure material is needed. Next the silicon is crushed and reacted with HCℓ gas, in the presence of copper-containing catalyst, to make trichlorosilane, a high vapour pressure liquid that boils at 31.8°C as in:

$$Si + 3HCl_{(gas)} \xrightarrow{Cu} SiHCl_3 + H_2 \qquad (1.30)$$

Many of the impurities in the silicon (aluminium, iron, phosphorus, chromium, manganese, titanium, vanadium, and carbon) also react with the HCℓ, forming various chlorides, which are highly reactive. Each of these chlorides have different boiling points, so by fractional distillation it is possible to separate out the $SiHCl_3$ from most of the impurities. The pure trichlorosilane is then reacted with hydrogen gas at an elevated temperature of about 1100°C to form pure electronic grade silicon (1 pp10^{-9} of impurities).

$$SiHCl_3 + H_2 \rightarrow 2Si + 3HCl \qquad (1.31)$$

Although the resultant silicon is relatively pure, it is in a polycrystalline form that is not suitable for semiconductor device manufacture. This so-called electronic grade polysilicon, EGS, requires further processing. The presented example is one way of producing pure silicon. There are other production procedures with different chemical reactions used, yet the end-product is the same - pure silicon.
The next step in the process is to grow single crystal silicon via one of three methods, as considered in section 1.19.3.

- the Czochralski process
- the float-zone process or
- the ribbon silicon process.

In *single-crystal silicon*, the molecular structure, which is the arrangement of atoms in the material, is uniform and defect free, because the entire structure is grown from the same crystal. This uniformity is ideal for transferring electrons efficiently through the material. To make an effective semiconductor, however, silicon is doped with other elements to make it n-type or p-type.

Multi-crystalline silicon, in contrast, consists of several smaller crystals or grains, which introduce boundaries. These boundaries impede the flow of electrons and encourage them to recombine with holes, thereby reducing the efficiency of the silicon. However, multi-crystalline silicon is much less expensive to produce than single-crystalline silicon.

1.19.2 Crystallinity

The crystallinity of a material indicates how perfectly ordered the atoms are in the crystal structure. Silicon, as well as other semiconductor materials, can come in various crystalline forms:

- single-crystalline,
- multi-crystalline,
- polycrystalline or
- amorphous.

In a single-crystal material, the atoms making up the structure of the crystal are repeated in a very regular, orderly manner from layer to layer. In contrast, a multi-crystalline material is composed of numerous smaller crystals, with the orderly arrangement disrupted when moving from one crystal to another. Multi and poly crystalline silicon are particularly important in areas like photovoltaic cells and light-emitting and laser diodes, when cost is an important factor. One classification scheme for silicon uses approximate crystal size and includes the methods typically used to grow or deposit such material, as shown in Table 1.9.

Table 1.9: Types of crystalline silicon and formation processing methods

Type of Silicon	abbreviation	Crystal Size Range	Deposition Method
single-crystal silicon	sc-Si	> 10cm	Czochralski, float-zone
multicrystalline silicon	mc-Si	1mm-10cm	Cast, sheet, ribbon
polycrystalline silicon	pc-Si	< 1mm-1mm	Chemical-vapour deposition
microcrystalline silicon	μc-Si	< 1mm	Plasma deposition

1.19.3 Single crystal silicon

Several different processes can be used to grow an ingot or boule of single or mono-crystal silicon. The most established and dependable processes are the *Czochralski method* and the *float-zone technique*. The *ribbon-growth* technique is used for lower cost and quality silicon crystal growth.

1.19.3i *Czochralski process*

The most commonly used process for creating the boule is called the *Czochralski* method, as illustrated in figure 1.22a. Electronic grade polysilicon silicon is heated in a quartz crucible to 1400°C in an argon atmosphere, using RF or resistance heating. A starter seed of single-crystal silicon on a puller contacts the top surface of molten polycrystalline silicon at 1415°C to 1420°C. As the seed crystal is slowly withdrawn - pulled and rotated, if the temperature gradient of the melt is adjusted so that the melting/freezing temperature is just at the seed-melt interface, atoms of the molten silicon solidify in the pattern of the seed and extend its single-crystal structure, forming a cylindrical boule of near perfect, pure silicon.

- The ingot pull is unusually pure, because impurities either burn or tend to be drawn into the liquid silicon. An argon atmosphere precludes any oxygen impurity. The rod and crucible are rotated in opposite directions to minimise the effects of convection in the melt. The pull-rate (1μm to 1mm/s), the rotation-rate (10 to 40 turns per minute), and the temperature gradient are carefully optimised for a particular wafer diameter (up to 30mm) and lattice structure orientation growth direction (direction <111>, along the diagonal of the sides of the cube crystal structure, for bipolar devices). Lengths of boule of several metres are attainable.
- A small amount of boron (or phosphorous) is usually added during the Czochralski process to pre-dope the substrate silicon.

1.19.3ii *Float-zone process*

The *float-zone process* produces purer crystals than the Czochralski method, because the pull is not contaminated by the crucible used in growing Czochralski crystals. In the float-zone process, as illustrated in figure 1.22b, a polycrystalline silicon rod is set atop a seed mono-crystal and then

effectively lowered through an electromagnetic induction coil. The coil's magnetic field induces an electric field in the rod, heating and melting the interface between the rod and the seed. Single-crystal silicon forms at the interface, growing upward as the rod is slowly raised. In this floating zone technique, the molten silicon is unsupported, maintaining itself through surface tension.
Rods of mono-silicon measure up to 10cm in diameter and 1m in length.

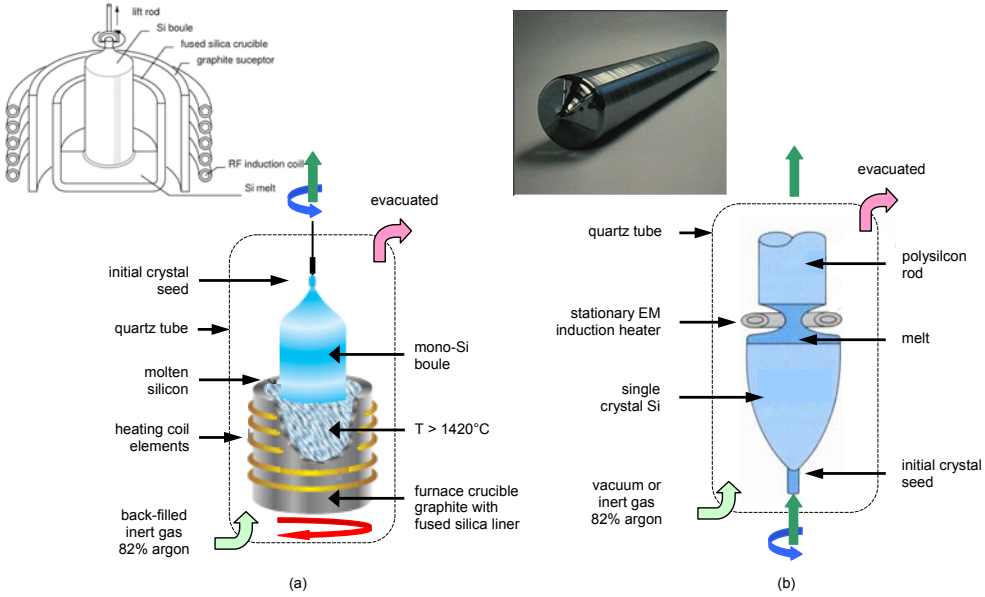


Figure 1.22. *Techniques for producing a single crystal silicon boule: (a) Czochralski method and (b) the float-zone method.*

Wafer preparation

The next step is the same for both single-crystal formation methods. The boule ends are cropped using a water-lubricated, single-blade diamond saw. The ingot is then ground to a uniform diameter in a lathe, and each end is bevelled with a sand belt to reduce the possibility of shattering the ingot. X-ray diffraction can then be used to determine the crystal structure orientation, which is marked by grinding the length of the cylindrical side of the boule.
The cylindrical single-crystal ingot is sawed, using a multi-blade, inner-diameter saw in conjunction with a wet lubricant, into thin wafers for further processing. The sawing wastes 20% to 50% of the silicon as sawdust, known as *kerf*.
The sliced wafers are mechanically lapped under pressure using a counter-rotating machine to achieve flatness and parallelism on both wafer sides. Most lapping operations use slurries of either aluminium oxide or silicon carbide. The edges of the individual wafers are also rounded by wet automatic grinders.
After lapping, the wafers are etched with a solution containing nitric, acetic, and hydrofluoric acids (HF, CH_3C00H, and $HN0_3$). This etching process removes external surface damage and reduces the thickness of the wafer.
Next, the wafers are polished using an aqueous mixture of colloidal silica and sodium hydroxide. The wafers are mounted onto a metal carrier plate that is attached by vacuum to the polishing machine. The chemical polishing process usually involves two or three grinding and polishing steps with progressively finer slurry, which decreases wafer thickness and results in a mirror-like lustre finish. Sometimes carrier pads must be stripped from the metal carrier plates. The pads are usually stripped with solvents such as methylene chloride, methyl ethyl ketone, or a glycol ether mixture.
Finally, the wafers are cleaned to remove any particles or residue remaining on the exterior surface of the polished wafer. Various cleaning steps and solutions containing ammonia, hydrogen peroxide, hydrofluoric acid, hydrochloric acid (NH_3, H_20_2, HF, and HCℓ), and deionised water may be used.
The finished wafers are inspected and packaged for shipping, since most semiconductor manufacturers purchase wafers from specialist wafer producers.

1.19.3iii *Ribbon silicon*

Although single-crystal silicon technology is well developed, the Czochralski and float-zone processes are complex and expensive, as are the ingot-casting processes discussed under multi-crystalline silicon. Another crystal-producing process is ribbon silicon growth, where the single crystals cost less than from other processes, because they form the silicon directly into thin, usable wafers of single-crystal silicon. By forming thin crystalline sheets directly, sawing and slicing steps of cylindrical boules are avoided.
One ribbon growth technique, termed edge-defined film-fed growth, starts with two crystal seeds that grow and capture a sheet of material between them as they are pulled from a source of molten silicon. A frame entrains a thin sheet of material when drawn from a melt. This technique does not waste much material, but the quality of the material is not as high as Czochralski process and float zone produced silicon. The resultant silicon quality is inferior for large-area, high-voltage, power semiconductor switching devices.

1.19.4 Multi-crystalline Silicon

Multi-crystalline (or poly-crystalline) silicon describes when the active portion of the silicon is made up of several relatively large crystals, called grains, up to a square centimetre or so in area.
Having several large crystals in a cell introduces a problem. Charge carriers can move around relatively freely within one crystal, but at the interface between two crystals, called the grain boundary, the atomic order is disrupted. Free electrons and holes are much more likely to recombine at grain boundaries than within a single crystal.

There are several ways to minimize the problems caused by grain boundaries:

- adjusting growth conditions through treatments such as annealing (heating followed by a slow cooling rate stage) the semiconductor material so that grains are columnar and as large as possible. The impurities are also better distributed;
- designing cells so that the charge carriers are generated within or close to the built-in electric field; and
- filling broken bonds at grain edges with elements such as hydrogen or oxygen, which is called passivating the grain boundaries.

Multi-crystalline silicon based devices are generally less efficient than those made of single-crystal silicon, but they can be less expensive to produce. Multi-crystalline silicon is produced in a variety of ways.

- The most common commercial methods involve a casting process in which molten silicon is directly cast into a mould and allowed to slowly solidify into an ingot. The starting material can be a refined lower-grade silicon, rather that the higher-grade semiconductor grade required for single-crystal material. The mould is usually square, producing an ingot that can be cut and sliced into square cells, minimising wasted silicon.
- The procedure of extracting pure multi or poly-crystalline silicon from tri-chlorine-silane can be (among others) performed in special furnaces. Furnaces are heated by electric current, which flows through (in most cases) silicon electrodes. The 2m long electrodes measure 8mm in diameter. The current flowing through electrodes can reach up to 6000A. The furnace walls are additionally cooled preventing the formation of any unwanted reactions due to gas side products. The procedure results in pure poly-crystalline silicon used as a raw material for solar cell production. Poly-crystalline silicon can be extracted from silicon by heating it up to 1500°C and then cooling it down to 1412°C, which is just above solidification of the material. The cooling is accompanied by origination of an ingot of fibrous-structured poly-crystalline silicon of dimensions 40x40x30 cm. The structure of poly-crystalline silicon in part of the material is settled, yet it is not adjusted to the structure of the other part.

1.19.5 Amorphous Silicon

Amorphous silicon is produced in high frequency furnaces in a partial vacuum atmosphere. In the presence of a high frequency electrical field, gases like silane, B_2H_6 or PH_3 are blown through the furnaces, supplying the silicon deposit with boron and phosphorus.
Amorphous solids, like common glass, are materials whose atoms are not arranged in any particular order. They do not form crystalline structures, and they contain large numbers of structural and bonding defects. Economic advantages are that it can be produced at lower temperatures and can be deposited on low-cost substrates such as plastic, glass, and metal. These characteristics make amorphous silicon the leading thin-film material.

Since amorphous silicon does not have the structural uniformity of single or multi crystalline silicon, small structural deviations in the material result in defects such as *dangling bonds*, where atoms lack a neighbour to which they can bond. These defects provide sites for electrons to recombine with holes, rather than contributing to the electrical circuit. Ordinarily, this kind of material would be unacceptable for electronic devices, because defects limit the flow of current. However, amorphous silicon can be deposited so that it contains a small amount of hydrogen, 5% to 10%, in a process called hydrogenation. The result is that the hydrogen atoms combine chemically with many of the dangling bonds, as shown in figure 1.23, essentially neutralising or removing them and permitting electrons to move through the material.

Staebler-Wronski Effect

Instability currently retards amorphous silicon exploitation in some semiconductor applications. In the case of photo-voltaic cells, the amorphous cells experience an electrical output decreases over a period of time when first exposed to sunlight. The electrical output stabilizes with a net output loss of 20%. The reason is related to the amorphous hydrogenated nature of the material, including tiny microvoids or atomic-level gaps in the amorphous silicon structure several angstroms in diameter (1 angstrom $=10^{-10}$ m). Other causes include oxygen or carbon impurities that are in the cells and ordinary stresses in the system that break silicon-silicon bonds in the region of the imperfections.

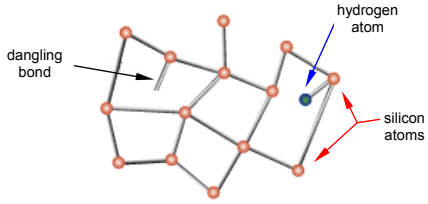


Figure 1.23. *Amorphous silicon showing the dangling bonds and hydrogen sites.*

Devices suffering from light induced degradation can recover their effectiveness if they are annealed at 150°C for a few minutes. Annealing is also effective at the normal operating temperatures of silicon, about 50° to 80°C. This is called self-annealing.

Summary of substrate structural features

Monocrystalline: An inorganic or organic compound functions in a device in a single crystal form (mono-crystal). Only small molecules/oligomers (finite monomers) can be used.

- Advantages: high reproducibility of properties, very high conductivity and carrier mobility.
- Drawbacks: complexity of fabrication of devices, low mechanical strength.

Polycrystalline: An inorganic or organic compound functions in a device in a form of a film/layer/bulk of many microscopic crystals. Both small molecules and low molecular weight polymers can be used.

- Advantages: good reproducibility of properties, high conductivity and mobility.
- Drawbacks: poor luminescent/optoelectronic properties.

Microcrystalline: composed of micrometre-size well-shaped crystals/lamellas (thin plates or layers).
Nanocrystalline: composed of nanometre-size crystals, often featured in shape (porous, hollow, etc).

Small-molecule amorphous solids:

- Advantages: simple fabrication of devices, good reproducibility of properties, good luminescent/ optoelectronic properties.
- Drawbacks: low thermal stability (glass transition).

Amorphous polymers:

- Advantages: simple fabrication of devices, high mechanical and thermal stability, good luminescent/ optoelectronic properties.
- Drawbacks: low reproducibility of properties, low conductivity and carrier mobility.

1.20 Silicon Carbide

Wide bandgap semiconductors (GaN (III-V), SiC, diamond, etc.) have better high voltage and temperature characteristics than silicon devices. However, because silicon carbide, SiC, sublimes at high temperature, ≈1800ºC, processing is more difficult than for silicon (which melts at a lower temperature of 1415°C). The similar chemistry properties of silicon and silicon carbide (both in group IV) means that many of the existing processes for silicon can be applied to silicon carbide, but with some refinement and higher processing temperatures. The exception is thermal diffusion which is not effective if a good SiC surface morphology is to be retained.

The SiC crystal boules are grown by seeded sublimation using the physical vapour transport (PVT) method. Alternatively, chemical vapour deposition (CVD) can be used, where SiH_4, C_3H_8, and H_2 are typically injected into the chamber. This process is mainly used for producing SiC epitaxial growth. A hot walled CVD reactor can deposit 100µm at a rate of 1 to 5 µm/hour at 1200ºC to 1500ºC. Crystal defects (micropipes, stack faults, etc.) occur at a rate of less than 1 per cm^2. Proprietary defect healing technology can significantly decrease the defect rate. The main single crystal polytypes for power switching device fabrication are 4H-SiC and 6H-SiC (this lattice structure terminology is based on the Ramsdell notation).

Nitrogen for n-type and aluminium or boron for p-type can be used in epitaxial growth and ion implantation. Substrates usually have an n or p epitaxial drift layer. Typical n-type epitaxy (50µm) can be thicker than a p-type layer (10µm), and the n-type epitaxy has a thin 1µm n-type buffer or fieldstop.

Ion implantation is shallow, typically less than 1µm, and requires high temperature and 30 to 300keV. Subsequent annealing in argon is at 1650ºC in the presence of a silicon over pressure. The lower the temperature, the longer the annealing time. Contact metallization can use sintered nickel on highly n-doped SiC, which is annealed at 1150ºC for a few minutes (with deuterium) (then overlaid with thick gold). A nickel, aluminium, and titanium Schottky metal sintered combination is suitable for p^+ region metallization, which is overlaid with gold. Aluminium metallisation is used at the Schottky diode anode.

$Si0_2$ is an electrical-insulator that can be grown on both Si and SiC. Oxide growth for SiC is slower than that on silicon and involves nitridation of nitric oxide, N_20, at 1300ºC. Because of the physical and chemical stability of silicon carbide, acid wet etching is ineffective and dry reactive ion etching tends to be used for etching processes.

1.21 Si and SiC physical and electrical properties compared

The processing of silicon is a mature, cost efficient technology, with 300mm wafers and submicron resolution common within the microelectronics industry. So-called wide bandgap semiconductors like silicon carbide (processed on 100mm wafers) offer promising high voltage and temperature power switching device possibilities as material quality and process yields improve. Figure 1.23 shows and allows comparison of the key physical and electrical properties of the main semiconductor materials applicable to power switching device fabrication.

The higher

- the energy bandgap, E_g, the higher the possible operating temperature before intrinsic conduction mechanisms produce adverse effects;
- the avalanche breakdown electric field, ξ_b, the higher the possible rated voltage;
- the thermal conductivity, λ, the more readily heat dissipated can be removed; and
- the saturation electron drift velocity, v_{sat}, and the electron mobility, μ_n, the faster possible switching speeds.

Although the attributes of wide bandgap materials are evident, processing is more difficult than with Si and some of the parameters vary significantly with a wide operating (and processing) temperature range. SiC performance characteristic figures are slightly better than those for GaN, except, importantly, GaN has better carrier mobility. GaN growth is complicated by the fact that nitrogen tends to revert to the gaseous state, and therefore only thin layers are usually grown on sapphire or SiC substrates. Lattice-substrate boundary misfit occurs because of the significant difference in molecule sizes and packing. This limitation is more accentuated with GaN on silicon. There is a 17% misfit in molecule package and a 56% mismatch in thermal expansion ($\alpha_{GaN} = 5.59\times10^{-6}$ and $\alpha_{Si} = 3.59\times10^{-7}$ @ 300K). To prevent cracking during processing cooling, an intermediate transition layer like AℓN, is introduced. Wide band gap based, low-voltage (<100V) GaN, lateral RF transistors are viable. But boundary imperfections may prove problematic with high-voltage, power devices where the principle current flow is usually vertically through the structure, hence through the imperfect mechanical and electrical lattice boundary. In the case of GaN, ohmic contacts involve annealing Ti/Aℓ/Ni/Au, while Schottky contacts can be Ni/Au. SiN_x passivation is deposited by plasma enhanced CVD.

Some of the physical and electrical parameters and their values in figure 1.24 will be explained and used in subsequent chapters. Other useful substrate data is given in Table 1.10.

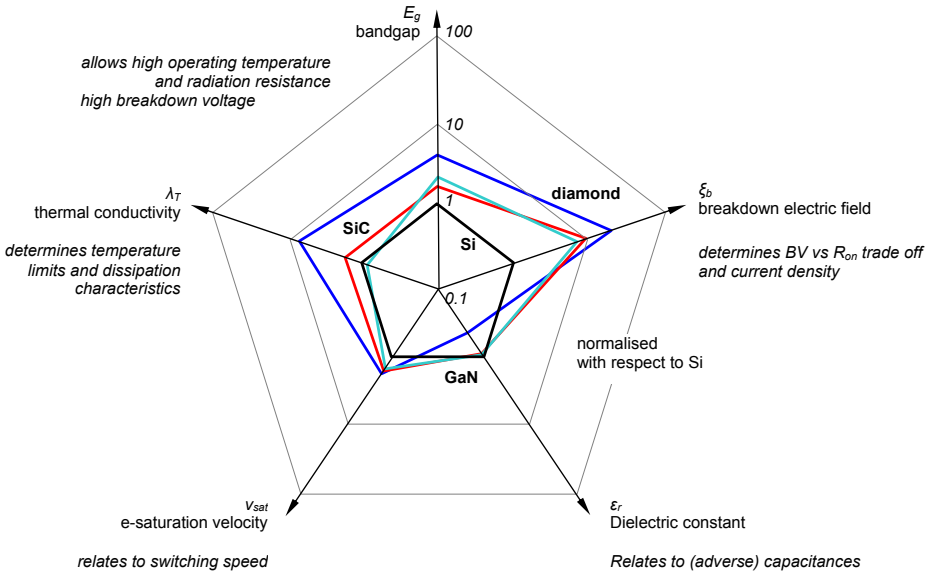


Figure 1.24. *Key electrical and thermal characteristics of group IV monocrystalline silicon, diamond, gallium nitride, and 3 polytypes of silicon carbide, at room temperature.*

material	Bandgap energy	Dielectric constant	Electron mobility	Breakdown electric field	Saturated electron drift velocity	Thermal conductivity	Figure of merit *w.r.t* Si	coefficient of linear thermal expansion
	E_g	ε_r	μ_n / μ_p	ξ_b	v_{sat}	λ_T	**FoM**	α
	eV	pu	cm^2/Vs	MV/cm	10^7 cm/s	W/mK	$\lambda_T \times(\xi_b \times v_{sat})^2$	x 10^{-6} K^{-1}
Si	1.12	11.9	1400 / 450	0.3	1.0	130	1	2.6
GaN	3.44	9.5	900 / 10	3.0	2.5	110	407	5.6
3C-SiC	2.36	9.7	800 / 20	1.3	2.7	700	2381	2.8
4H-SiC	3.25	9.9	700 / 100	3.2	2.7	700	3241	5.2
6H-SiC	2.86	9.7	400 / 75	2.4	2.0	700	1307	5.2
diamond	5.45	5.7	2200 / 1800	5.7	2.7	1000	54000	0.8

Table 1.10: Other useful substrate material data

parameter			Si	SiC	GaN	Diamond
maximum operating temperature	T_{max}	°C	300	1240		1100
melting temperature	T_{melt}	°C	1415	sublime >1800	2500	phase change
density	ρ	g / cm^3	2.33	3.17-3.21	6.15	3.52
electrical resistivity	ρ_e	Ω m	10^{-3}		10^{-3}	

Reading list

Streetman, B. G. and Banerjie, S. K., *Solid State Electronic Devices*, Prentice-Hall International, 6th Edition, 2005.

Van Zeghbroeck, B., *Principles of Semiconductor Devices*, http://ece-www.colorado.edu/~bart/book/

Zetterling, C. M., *Process technology for Silicon Carbide devices*, IEE, 2002.

http://www.siliconfareast.com/

http://www.semiconductorglossary.com/

2

The pn Junction

The diode is the simplest bipolar semiconductor device. It comprises p-type and n-type semiconductor materials brought together, usually after diffusion, to form a (step or abrupt) junction as shown in figures 2.1a and 2.2a.

A *depletion layer*, or alternatively a *space charge layer*, scl, is built up at the junction as a result of diffusion caused by the large carrier concentration gradients. The holes diffuse from the p-side into the n-side while electrons diffuse from the n-side to the p-side, as shown in figure 2.1b. The n-side, losing electrons, is charged positively because of the net donor charge left behind, while the p-side conversely becomes negatively charged. An electric potential barrier, ξ, builds up, creating a drift current which opposes the diffusion flow, both of which balance at thermo-dynamic equilibrium as shown in figure 2.1c. The electric field is a maximum at the junction and zero at the scl edges. There are no free carriers in the scl.

The zero external bias, built-in, junction potential or scl potential is given by

$$\Phi = \frac{kT_j}{q}\,\ell n\frac{N_A N_D}{n_i^2} \qquad \text{(V)} \qquad (2.1)$$

where q is the electron charge, 1.6 x 10^{-19} C
k is Boltzmann's constant, 1.38 x 10^{-23} J/K
T_j is the junction temperature, K.
Thus $\Phi = kT_j/q = 0.0259\,\text{eV}$ at room temperature, 300 K.

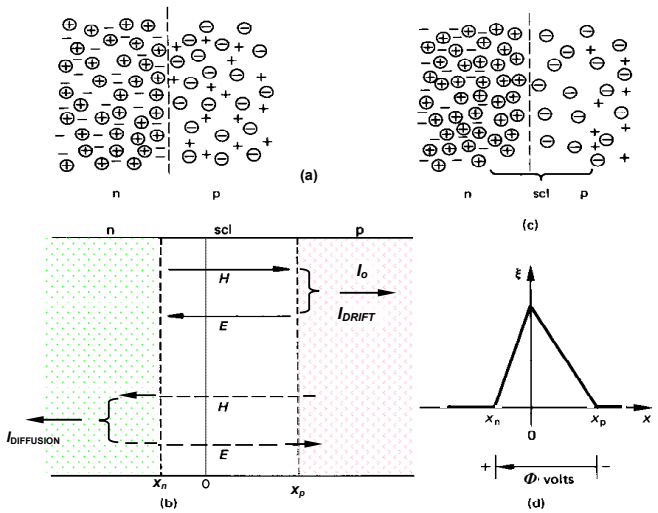


Figure 2.1. *The step junction: (a) the junction if carriers did not diffuse: ⊕ionised donors, ⊖ionised acceptors, + holes and - electrons; (b) electron and hole movements: ---- diffusion flow, —— drift flow; (c) ionised impurities and free carriers equilibrium distribution; and (d) scl electric field and voltage.*

One important feature of the pn junction is that current (holes) flows freely in the p to n direction when forward-biased, that is, the p-region is biased positive with respect to the n-region. Only a small leakage current flows in the reverse voltage bias case. This asymmetry makes the pn junction diode useful as a rectifier, exhibiting static voltage-current characteristics as illustrated in figure 2.2.

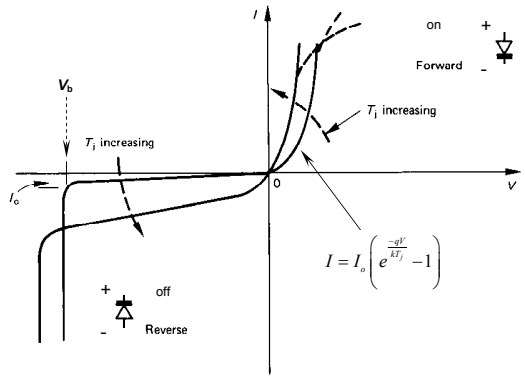


Figure 2.2. *Typical I-V static characteristics of a silicon pn junction diode, and the effects of junction temperature, T_j.*

Example 2.1: ***Built-in potential of an abrupt junction***

A silicon abrupt p-n junction has a p-type region of 2x10^{16} cm^{-3} acceptors and an n-type region containing 2x10^{16} cm^{-3} acceptors in addition to 10^{17} cm^{-3} donors.

Calculate

i. the thermal equilibrium density of electrons and holes in the p-type region, and both densities in the n-type region
ii. the built-in potential of the p-n junction at room temperature
iii. the built-in potential of the p-n junction at 400K, assuming the intrinsic concentration increases 300 fold over that at 300K

Solution

i. The thermal equilibrium densities, using $n_i^2 = p \times n$ from chapter 1, are:

in the p-type region

$$p = N_A = 2\times10^{16}\ \text{cm}^{-3}$$

$$n = n_i^2/p = \left(1.5\times10^{10}\right)^2 \Big/ 2\times10^{16}\ \text{cm}^3 = 1.125\times10^{3}\ \text{cm}^{-3}$$

in the n-type region

$$n = N_D - N_A = 8\times10^{16}\ \text{cm}^{-3}$$

$$p = n_i^2/n = \left(1.5\times10^{10}\right)^2 \Big/ 8\times10^{16}\ \text{cm}^3 = 2.813\times10^{3}\ \text{cm}^{-3}$$

ii. The built-in potential at room temperature from equation (2.1) is

$$\Phi = \frac{kT_j}{q}\,\ell n\frac{p_n n_p}{n_i^2} = 0.0259\text{V}\times\ell n\left(\frac{2\times10^{16}\times8\times10^{16}}{\left(1.5\times10^{10}\right)^2}\right) = 0.766\text{V}$$

iii. The intrinsic carrier density is temperature dependant, and increases to 300 x 1.5x10^{10} = 4.5x10^{12} at 400K.

The built-in potential at 400K is

$$\Phi = \frac{kT_j}{q}\,\ell n\frac{p_n n_p}{n_i^2} = 0.0259\text{V}\times\frac{400\text{K}}{300\text{K}}\times\ell n\left(\frac{2\times10^{16}\times8\times10^{16}}{\left(4.5\times10^{12}\right)^2}\right) = 0.630\text{V}$$

♣

2.1 The pn junction under forward bias (steady-state)

If the p-region is externally positively-biased with respect to the n-region as shown in figure 2.3b, the scl narrows and current flows freely. The emf positive potential supplies holes to the p-region, while the negative emf potential provides electrons to the n-region. The carriers both combine, but are continuously replenished from the emf source. A large emf source current flows through the diode, which is termed *forward-biased*.

2.2 The pn junction under reverse bias (steady-state)

If a bias voltage is applied across the p and n regions as shown in figure 2.3c, with the p-terminal negative with respect to the n-terminal, then the scl widens. This is because electrons in the n-region are attracted to the positive external emf source while holes in the p-region are attracted to the negative emf potential. As the scl widens, the peak electric field ξ_m at the junction increases as shown in figure 2.3d. The only current that flows is the small *leakage current* which is due to carriers generated in the scl or minority carriers which diffuse to the junction and are collected. The junction condition is termed *reverse-biased*.

Increasing applied reverse bias eventually leads to junction reverse voltage breakdown, V_b, as shown in figure 2.2 (third quadrant), and the diode current is controlled (limited) by the external circuit. Junction breakdown is due to one of three phenomena, depending on the doping levels of the regions and, most importantly, on the concentration of the lower doped side of the junction.

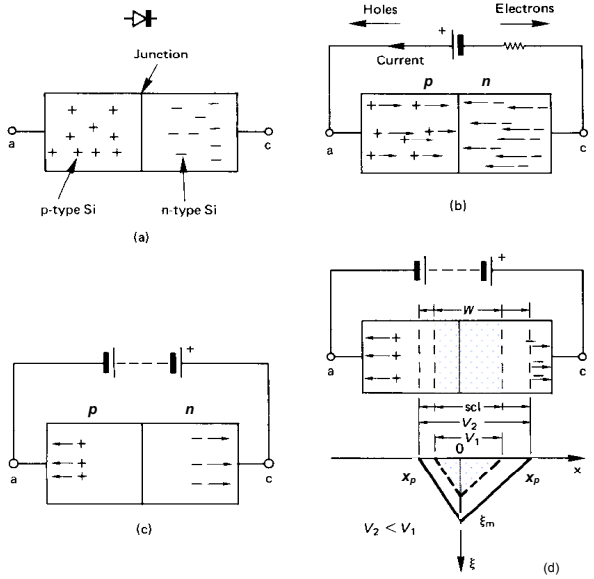


Figure 2.3. *Diagrammatic representation of a pn junction diode showing minority carrier flow: (a) without external applied voltage; (b) with forward applied voltage; (c) with reverse applied voltage; and (d) electric field and scl change with increased reverse applied voltage.*

2.2.1 Punch-through voltage

The reverse voltage extends the scl to at least one of the ohmic contacts and the device presents a short circuit to that voltage in excess of the *punch-through* voltage, V_{PT}. Punch-through tends to occur at low temperatures with devices which employ a low concentration region (usually the n-side), as is usual with high-voltage devices. The punch-through voltage for silicon can be approximated by

$$V_{PT} = 7.67\times10^{-16} N_c W_c^2 \qquad \text{(V)} \qquad (2.2)$$

where N_c is the concentration in /cc of the lighter doped region and
W_c is the width of that region in μm.

2.2.2 Avalanche breakdown

Avalanche breakdown or *multiplication breakdown*, is the most common mode of breakdown and occurs when the peak electric field, ξ_m, in the scl at the junction exceeds a certain level which is dependent on the doping level of the lighter doped region. Minority carriers associated with the leakage current are accelerated to kinetic energies high enough for them to ionise silicon atoms on collision, thereby creating a new hole-electron pair. These are accelerated in opposite directions, because of the high electric field strength, colliding and ionising repeatedly - hence the term avalanche, impact ionisation or carrier multiplication. If the lighter doped silicon region has a concentration of

$$10^{13} < N_c < 5\times10^{14} \qquad (\text{/cc})$$

then the avalanche voltage may be approximated by

$$V_b = 5.34\times10^{13}\, N_c^{-3/4} \qquad \text{(V)} \qquad (2.3)$$

The peak electric field at the junction will be

$$\xi_b = 3.91\times10^{5}\, N_c^{1/8} \qquad \text{(V/m)} \qquad (2.4)$$

and the width of the scl, mainly in the lighter doped region, at breakdown is given by

$$W_{scl} = 2\,V_b / \xi_b = 2.73\times10^{8}\, N_c^{-7/8} \quad \left(\approx 1\times10^{-14}\, V_b\right) \qquad \text{(m)} \qquad (2.5)$$

2.2.3 Zener breakdown

Field or Zener breakdown occurs with heavily doped junction regions and at usually less than 5V reverse bias. It occurs when the scl is too narrow for avalanche yet the electric field grows very large and electrons tunnel directly from the valence band on the p-side to the conduction band on the n-side. This reverse current is called the *Zener effect*.

These three modes of reverse voltage breakdown are not necessarily destructive provided the current is uniformly distributed and limited by the external circuit. If the current density in a particular area is too high, a local hot spot may occur, leading to device thermal destruction.

2.3 Thermal effects

The pn junction current, I, shown in figure 2.2, is related to the scl voltage, v, according to

$$I(v) = I_o\left[e^{-qv/kT_j} - 1\right] \qquad \text{(A)} \qquad (2.6)$$

where I_o is the reverse (saturation) leakage current in amps.

The forward conduction voltage decreases with increased junction temperature, T_j. That is, the on-state voltage has a negative temperature coefficient. In practical silicon pn diodes, at low currents, the temperature coefficient is typically -2.4 mV/K, becoming less negative with increased current. At higher currents, the coefficient becomes positive because of the reduced carrier mobility at higher temperatures, which causes non-scl regions to increase in resistance. The effects of the change in temperature coefficient at higher currents, in practical devices, are shown dotted in figure 2.2. Neglecting the exponential silicon bad gap temperature dependence, the temperature effects at high current, on the diffusion constant component of the leakage current I_o in equation (2.6), called the saturation current, is given by

$$I_o(T) = I_o(25°\text{C})\times\left(\frac{T}{300}\right)^{1.8} \qquad (2.7)$$

Silicon carbide diodes have a higher temperature coefficient, typically +8mV/K.

The avalanche voltage increases with temperature, as does the reverse leakage current. The effects of temperature on the reverse bias characteristics are shown in figure 2.2. In the case of silicon carbide, increased temperature decreases the avalanche voltage and increases the leakage current.

The silicon temperature coefficient for avalanche is positive since the mean distance between collisions is reduced due to the increased thermal energy, which increases the vibrational amplitude. Higher electric fields are necessary for the carriers to gain sufficient kinetic energy for ionisation.

Equation (2.6) also indicates that the reverse bias current increases with increased junction temperature. This positive temperature coefficient does not generally result in thermal instability with silicon devices, provided sufficient heat sinking is employed on smaller devices.

Example 2.2: ***Diode forward bias characteristics***

A pn junction diode has a reverse saturation current of 100nA at a junction operating temperature of 28°C. What is the forward current when the forward bias voltage is 0.5V and the dynamic ac resistance at that current?

Solution

The diode current is given by equation (2.6), namely

$$I(v) = I_o\left[e^{-qv/kT_j} - 1\right] \quad \text{(A)}$$

where $V_T = q/kT_j = 0.0259\text{V}$ at 300K, 28°C.

$$I(0.5\text{V}) = 100\times10^{-9}\times\left[e^{v/0.0259} - 1\right] = 26.1\text{A}$$

Differentiating the diode *V-I* equation with respect to voltage gives

$$\frac{di}{dv} = \frac{I_o}{V_T}e^{\frac{V_F}{V_T}} = \frac{1}{R_{ac}} = \frac{100\times10^{-9}\text{A}}{0.0258\text{V}}\times e^{\frac{0.5\text{V}}{0.0258\text{V}}}$$

$$\Rightarrow R_{ac} = 1.0\text{m}\Omega$$

♣

2.4 Models for the bipolar junction diode

Semiconductor device electrical models are used extensively for power electronic circuit simulation. A basic piecewise-linear model is applicable to simple manual calculations, where the terminal *I-V* characteristics are empirically modelled based on ideal circuit elements. A more complex and accurate model is required for computer transient circuit analysis simulation. Such accurate models are based on the semiconductor physics of the device. Many power switching semiconductor device manufacturers provide values for the model parameters suitable for circuit simulation in the packages PSpice and SABER.

2.4.1 Piecewise-linear junction diode model

The pn junction diode is a unilateral device that, to a good approximation, conducts current in only one direction. Figure 2.4a shows a *piecewise-linear* (pwl) model of the diode that is suitable for static modelling in power electronic circuits. It includes a perfect diode, an on-state voltage source E_o, and a series resistor of resistance value R_o to account for the slope in the actual forward conduction characteristic. The forward *I-V* characteristic at a given temperature is given by

$$V_F(I_F) = E_0 + I_F R_0 \quad \text{for } V_F > E_0 \quad \text{(V)} \qquad (2.8)$$

The model in figure 2.4a does not incorporate the static reverse characteristics of leakage and avalanche. These are shown in figure 2.4b, where V_b from equation (2.3) models the avalanche limit and $R_l\ (=V_b/I_o)$ gives linear leakage current properties for a given junction temperature. The three diode components in figure 2.4b are assumed ideal.

The model given by equation (2.8) is adequate for calculation of static balancing requirements of parallel and series connected diodes and thyristors, as considered in section 10.1 and the associated problems, 10.4, 10.5, and 10.9 to 10.12.

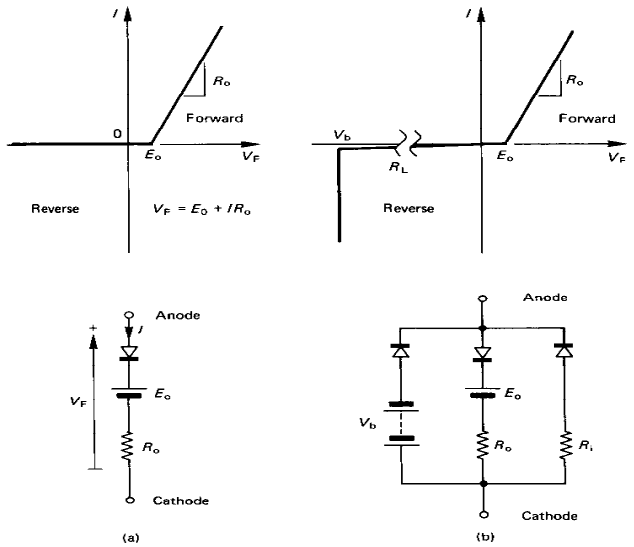


Figure 2.4. *Piecewise-linear approximations of junction diode characteristics: (a) ideal diode with an offset voltage and resistance to account for slope in the forward characteristic and (b) model including reverse bias characteristics.*

Example 2.3: ***Using the pwl junction diode model***

An approximation to the forward *I-V* characteristic of the diode shown in figure 2.4a, is given by $V_F(I_F) = 1.0 + 0.01\,I_F$. For a constant current of 45A for ⅔ of a cycle, calculate the diode

i. on-state voltage;
ii. mean power loss; and
iii. rms current.

Solution

i. The on-state voltage at 45A is given by

$$V_F(45\text{A}) = 1.0 + I_F\ 0.01 = 1.0\text{V} + 45\text{A}\times0.01\Omega = 1.45\text{V}$$

ii. If the on-state duty cycle is δ = ⅔, the average power loss is

$$\bar{P} = \delta\times V_F\times I_F = \tfrac{2}{3}\times1.45\text{V}\times45\text{A} = 43.2\text{W}$$

iii. The diode rms current is given by

$$I_{rms} = \sqrt{\delta}\times I_{dc} = \sqrt{\tfrac{2}{3}}\times45\text{A} = 36.7\text{A}$$

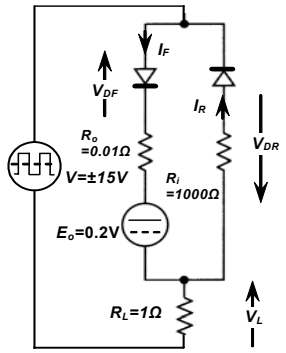
♣

Example 2.4: ***Static linear diode model***

A Schottky diode is used to half-wave rectify a square wave ±15V source in series with a 1Ω load resistor. If the diode model shown in figure 2.4b is modelled with R_o = 0.01 Ω, E_o = 0.2V, R_l = 1000Ω, and V_b = 30V, determine:

i. the diode model forward and reverse bias operating point equations for the series circuit
ii. the load current and diode voltage
iii. the rectifier losses (neglecting any recovery effects) and the load power dissipation
iv. estimate the power dissipated in the load if the source is ac with the same fundamental component as the square wave
v. what is the non-fundamental power dissipated with the square wave source?

Solution



i. When the diode is *forward biased*

$$i_F = \frac{1}{R_o}\left(v_{DF} - E_o\right) \quad \text{for} \quad v_{DF} \geq 0.2\text{V}$$

Kirchhoff's voltage law for the series circuit gives

$$V_s = i_F R_L + v_{DF}$$

Eliminating the diode voltage v_{DF} gives the series circuit current

$$i_F = \frac{V_s - E_o}{R_o + R_L} \quad \text{for} \quad V_s \geq E_o$$

$$i_F = 0 \quad \text{for} \quad 0 < V_s < E_o$$

The diode forward voltage is therefore given by

$$v_{DF} = \frac{V_s R_o + E_o R_L}{R_o + R_L} = E_o + i_F R_o \quad \text{for} \quad i > 0$$

When the diode is *reversed biased*, below the reverse breakdown voltage V_b = 30V

$$i_R = \frac{1}{R_l} v_{DR} \quad \text{for} \quad v_{DR} < V_b$$

$$V_s = i_R R_L + v_{DR}$$

Eliminating the diode voltage v_{DR} gives the series circuit leakage current

$$i_R = \frac{V_s}{R_l + R_L}$$

The diode reverse voltage is thus given by

$$v_{DR} = \frac{V_s R_l}{R_l + R_L} = i_R R_l$$

ii. The circuit voltages and current are, when the diode is forward biased,

$$i_F = \frac{15\text{V} - 0.2\text{V}}{0.01\Omega + 1\Omega} = 14.65\text{A}$$

$$V_{D_F} = 0.2\text{V} + 14.65\text{A} \times 0.01\Omega = 0.35\text{V}$$

If $R_L >> R_o$, the diode forward current equation can be simplified using R_o = 0.
When the diode is reverse biased

$$i_R = \frac{15\text{V}}{1000\Omega + 1\Omega} = 15.0\text{mA}$$

$$V_{D_R} = 15\text{mA} \times 1000\Omega = 15.0\text{V}$$

If $R_l >> R_L$ the diode reverse current and voltage equations can be simplified using R_L = 0.

iii. The rectifier losses are, when forward biased,

$$P_{d_F} = v_{D_F} \times i_F$$

$$= 0.35\text{V} \times 14.65\text{A} = 5.127\text{W}$$

and when reverse biased

$$P_{d_R} = v_{D_R} \times i_R$$

$$= 15\text{V} \times 15\text{mA} = 0.225\text{W}$$

Total diode losses for a square wave are therefore ½×(5.127W + 0.225W) = 2.68W.
The power from the square wave supply is

$$\tfrac{1}{2} \times \left(15\text{V} \times 14.65\text{A} + 15\text{V} \times 15\text{mA}\right) = 110.0\,\text{W}$$

with 110W – 2.68W = 107.32W dissipated in the 1Ω load resistor.

iv. The magnitude of the fundamental of a square wave is 4/π times the square wave magnitude, that is, 15V×4/π = 19.1V peak.

The forward biased diode does not conduct until the supply voltage exceeds 0.2V. This is a small percentage of the sine wave magnitude (≈1%), hence can be neglected in the loss estimate. The forward current flow is approximately

$$i_F = \frac{19.1\text{V} - 0.2\text{V}}{1\Omega + 0.01\Omega} \times \sin \omega t = 18.7 \times \sin \omega t$$

The rms of a sine is 1/√2 its magnitude and 1/√2 again for a half-wave rectified sine. That is

$$i_{Frms} = \frac{18.7\text{A}}{\sqrt{2}\,\sqrt{2}} = 9.35\text{A rms}$$

The reverse leakage current is given by

$$i_R = \frac{19.1\text{V}}{1000\Omega + 1\Omega} \times \sin \omega t = 0.019 \times \sin \omega t$$

which gives an rms current of

$$i_{Rrms} = \frac{19\text{mA}}{\sqrt{2}\sqrt{2}} = 9.5\text{mA rms}$$

The power dissipated in the 1Ω load resistor is

$$P_L = \left(i_{F_{rms}}^2 + i_{R_{rms}}^2\right) \times R_L$$

$$= \left(9.35^2 + 0.0095^2\right) \times 1\Omega = 87.42\text{W} + 90\mu\text{W} = 87.42\text{W}$$

Clearly, if $R_L << R_l$, the reverse leakage current related power component is negligible.

v. With the square wave, from part iii., 107.32W are dissipated in the load but from part iv., only 87.42W are dissipated for a sine wave with the same fundamental magnitude. The 19.9W difference is power produced by the harmonics of the fundamental (3rd, 5th, …). For a resistive heating load this power produces useful heating, but in a motor the harmonic power would produce unwanted torque pulsations and motor heating.

♣

2.4.2 Semiconductor physics based junction diode model

The charge-carrier diode model shown in figure 2.5 is necessary for transient (time domain) circuit analysis involving diodes. The pn junction diode is assumed to have an abrupt or step junction. The model components are voltage dependant current sources, I and I_b, voltage dependant capacitance C_t and C_j, and series access resistance R_s.

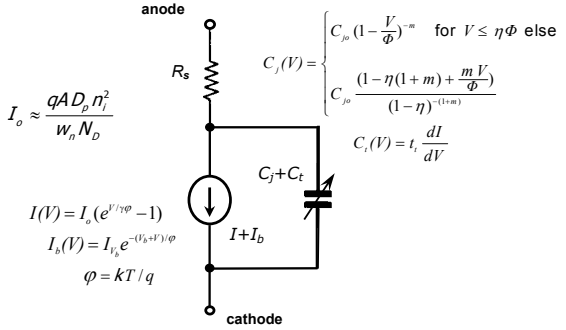


Figure 2.5. *PSpice transient analysis circuit model of the pn junction diode.*

The ideal diode current I is given by equation (2.6). The diode current I_b models reverse voltage breakdown, where the breakdown voltage V_b is assumed due to avalanche and is given by equation (2.3). The voltage dependant transit capacitance, C_t, which is dominant under forward bias, is related to the minority carrier lifetime t_t. The voltage dependant scl (depletion layer) capacitance C_j, which is dominant under reverse bias, involves the zero bias junction potential voltage Φ, given by equation (2.1) and the zero bias junction capacitance C_{jo}. In the case of the silicon carbide Schottky diode, $C_j >> C_t$. The scl capacitance, $C_j(V)$ can be evaluated from the pn diode structure and doping profile, as follows.

2.4.2i - Determination of zero bias junction capacitance, C_{jo}

Poisson's equation, in conjunction with Gauss's law, for the one dimensional step junction shown in figure 2.6, give

$$\frac{d^2V}{dx^2} = -\frac{d\xi}{dx} = \frac{qN_D}{\varepsilon_s} = -\frac{qN_A}{\varepsilon_s} \qquad (2.9)$$

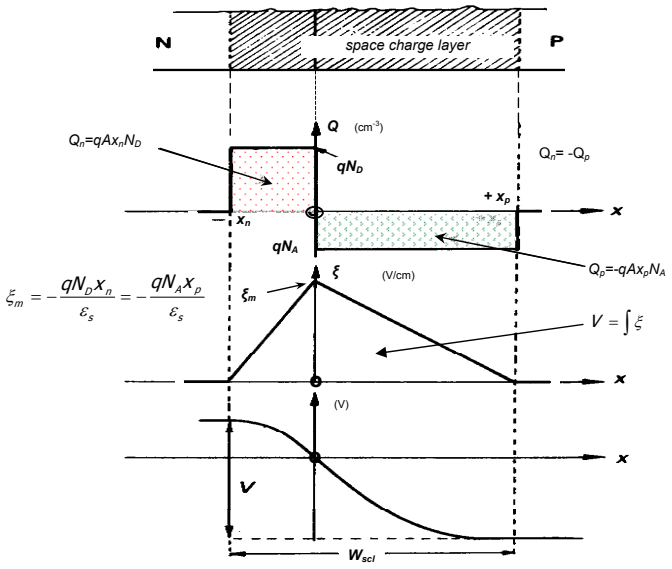


Figure 2.6. *The charge Q, electric field ξ, and voltage potential V, in the space charge layer of a step pn junction.*

The dielectric permittivity $\varepsilon_s = \varepsilon_r \varepsilon_o$ comprises the free space permittivity ε_o = 8.854x10^{-12} F/m and the relative permittivity ε_r = 11.8 for silicon and 9.7 for SiC.

$$\frac{d\xi(x)}{dx} = \begin{cases} \frac{q}{\varepsilon_s} N_D & \text{for} \quad -x_n < x < 0 \\ -\frac{q}{\varepsilon_s} N_A & \text{for} \quad 0 < x < x_p \end{cases} \tag{2.10}$$

Integrating both parts of equation (2.10) over the shown bounds, gives $\xi(x)$:

$$\frac{dV(x)}{dx} = \xi(x) = \begin{cases} \frac{q}{\varepsilon_s} N_D x + \xi_m & \text{for} \quad -x_n < x < 0 \\ -\frac{q}{\varepsilon_s} N_A x + \xi_m & \text{for} \quad 0 < x < x_p \end{cases} \tag{2.11}$$

where the maximium field intensity (at $x = 0$) is $\xi_m = \frac{q}{\varepsilon_s} N_D x_n = \frac{q}{\varepsilon_s} N_A x_p$

The piece-wise parabolic voltage potential across the scl shown in figure 2.6, is given by integration of the electric field, that is

$$V = \int_{-x_n}^{0} \left(\frac{q}{\varepsilon_s} N_D x + \xi_m\right) dx + \int_{0}^{x_p} \left(-\frac{q}{\varepsilon_s} N_A x + \xi_m\right) dx = \tfrac{1}{2} \xi_m W_o \tag{2.12}$$

Since the charges each side of the metallurgical junction must balance, equation (2.12) can be rearranged to give the scl width.

$$W_o = \sqrt{\left[\frac{2\varepsilon_s V}{q}\left(\frac{1}{N_A} + \frac{1}{N_D}\right)\right]} \tag{2.13}$$

From equation (2.1), a zero bias voltage Φ exists without the presence of any external voltage. Therefore, to incorporate non-equilibrium conditions, the electrostatic barrier potential becomes Φ-V, where V is the externally applied reverse bias voltage. Consequently the scl width expression becomes:

$$W_{scl} = \sqrt{\left[\frac{2\varepsilon_s (\Phi - V)}{q}\left(\frac{1}{N_A} + \frac{1}{N_D}\right)\right]} \tag{2.14}$$

The scl width voltage dependence can be expressed in terms of the zero bias scl width, W_o

$$W_{scl}(V) = \sqrt{\left[\frac{2\varepsilon_s \Phi}{q}\left(\frac{1}{N_A} + \frac{1}{N_D}\right)\right]}\sqrt{1 - \frac{V}{\Phi}} = W_0 \sqrt{1 - \frac{V}{\Phi}} \tag{2.15}$$

$$x_{n0} = \frac{W_0}{1 + N_D/N_A} \qquad x_{p0} = \frac{W_0}{1 + N_A/N_D}$$
$$x_n(V) = x_{n0}\sqrt{1 - \frac{V}{\Phi}} \qquad x_p(V) = x_{p0}\sqrt{1 - \frac{V}{\Phi}} \tag{2.16}$$

The magnitude of the voltage dependant charge on each side of the junction is

$$|Q(V)| = qA\frac{N_D N_A}{N_D + N_A} W = A\left[2q\varepsilon_s \Phi \frac{N_D N_A}{N_D + N_A}\right]^{\frac{1}{2}} \sqrt{1 - \frac{V}{\Phi}} = Q_0 \sqrt{1 - \frac{V}{\Phi}} \tag{2.17}$$

The junction capacitance is given by differentiation of equation (2.17) with respect to V

$$C_j = \left|\frac{dQ}{dV}\right| = \varepsilon_s A\left[\frac{q}{2\varepsilon_s(\Phi - V)} \frac{N_D N_A}{N_D + N_A}\right]^{\frac{1}{2}} = \frac{\varepsilon_s A}{W} \tag{2.18}$$

Equation (2.18) can be rearranged to give the PSpice capacitance form, in terms of the zero bias junction capacitance C_{jo}.

$$C_j(V) = \frac{C_{jo}}{\left(1 - \frac{V}{\Phi}\right)^{\frac{1}{2}}} \qquad \text{where } C_{jo} = \varepsilon_s A\left[\frac{q}{2\varepsilon_s \Phi} \frac{N_D N_A}{N_D + N_A}\right]^{\frac{1}{2}} = \frac{\varepsilon_s A}{W_o} \tag{2.19}$$

The electric field at the metallurgical junction, from equation (2.12) is given by

$$\xi_j(V) = \xi_0 \sqrt{1 - \frac{V}{\Phi}} \quad \text{where} \quad \xi_0 = 2\Phi / W_0 \tag{2.20}$$

2.4.2ii - One-sided pn diode equations

When $N_A >> N_D$, which is the usual case in high voltage pn diodes, equations (2.12) to (2.20) are approximated by the following one-sided diode equations.

$$W_0 = \sqrt{\left[\frac{2\varepsilon_s \Phi}{q N_D}\right]} \approx x_{no} \quad \text{and} \quad x_{po} \approx 0$$

$$Q_0 = A\sqrt{2q\varepsilon_s \Phi N_D} \tag{2.21}$$

$$C_{jo} = \varepsilon_s A\sqrt{\frac{q N_D}{2\varepsilon_s \Phi}} = \frac{\varepsilon_s A}{W_o}$$

These equations show that the scl penetrates mostly into the n-side, (hence the name one-sided), which supports most of the voltage, as shown in the last diagram in figure 2.6.

Example 2.5: ***Space charge layer parameter values***

A 10μm thick p-type $2\text{x}10^{16}$ /cc silicon epitaxial layer is grown on an n^- - type $1\text{x}10^{14}$ /cc silicon substrate, of area 1 cm^2, to form an abrupt pn junction.

Calculate the following PSpice parameter values, at room temperature:

i. zero bias junction potential, Φ;
ii. zero bias scl width, maximum electric field, charge, and junction capacitance, W_0, ξ_0, Q_o, C_{jo}; and
iii. avalanche breakdown voltage, V_b.

If the substrate is 150μm thick, for a 1000V reverse bias, calculate:

iv. scl width and penetration depth each side of the junction, W, x_n, x_p;
v. charge each side of the junction, maximum electric field, and the capacitance, Q, ξ_j, C_j.

Solution

i. From equation (2.1), the zero bias built-in voltage is

$$\Phi = \frac{kT_j}{q}\,\ell n\frac{N_A N_D}{n_i^2} = 0.0259\,\ell n\frac{2\times10^{16}\times\ 1\times10^{14}}{2.25\times10^{20}}$$
$$= 0.0259\times\ell n(8.89\times10^{9}) = 0.534\text{V}$$

ii. From equations (2.15), (2.20), (2.17), and (2.19)

$$W_0 = \sqrt{\left[\frac{2\Phi\,\varepsilon_s}{q}\times\left(\frac{1}{N_A}+\frac{1}{N_D}\right)\right]}$$

$$W_0 = \sqrt{\left[\frac{2\times0.53\,\text{V}\times11.8\times8.85\times10^{-12}}{1.6\times10^{-19}}\times\left\{\frac{1}{2\times10^{22}}+\frac{1}{1\times10^{20}}\right\}\right]} \quad = \quad 2.65\,\mu\text{m}$$

$$\xi_0 = 2\ \Phi/W = 2\times0.534/2.65\mu\text{m}$$
$$\xi_0 = 0.40\ \text{MV/m}$$

$$Q_o = A\left[2\Phi q\varepsilon_s\frac{N_D N_A}{N_D+N_A}\right]^{\frac{1}{2}}$$

$$Q_o = 1\times10^{-4}\sqrt{2\times0.53\text{V}\times1.6\times10^{-19}\times11.8\times8.85\times10^{-12}\ \times\frac{2\times10^{42}}{2.01\times10^{22}}} \quad = \quad 4.21\ \text{nC}$$

$$C_{jo} = A\left[\frac{q\varepsilon_s}{2\Phi}\,\frac{N_D N_A}{N_D+N_A}\right]^{\frac{1}{2}}$$

$$C_{jo} = 1\times10^{-4}\times\left[\frac{1.6\times10^{-19}\times11.8\times8.85\times10^{-12}}{2\times0.53\text{V}}\times\frac{2\times10^{42}}{2.01\times10^{22}}\right]^{\frac{1}{2}}$$
$$= 3.95\times10^{-9}\,\text{F} \quad = \quad 3.95\ \text{nF}$$

iii. From equation (2.2), the estimated punch-through voltage is

$$V_{PT} = 7.67\times10^{-16}\,N_c\,W_c^2 = 7.67\times10^{-16}\times10^{14}\times(150)^2$$
$$= 1727\text{V}$$

That is, punch through occurs when the reverse bias is greater than the operating voltage, 1000V. If the diode is to breakdown due to avalanche then the avalanche breakdown voltage given by V_b (equation (2.3)) must be less than V_{PT}, 1727V.

$$V_b \ = \ 5.34\times10^{13}\ N_c^{-3/4} = 5.34\times10^{13}\times(1\times10^{14})^{-3/4}$$
$$= 1689\text{V}$$

iv. From equation (2.15) the scl width at -1000V reverse bias is

$$W_{scl} = W_0\sqrt{1-\frac{V}{\Phi}} \quad = \quad 2.65\mu\sqrt{1+\frac{1000\text{V}}{0.533\text{V}}}$$
$$= 114.6\mu\text{m}$$

From equation (2.16) the scl penetration into each side of the junction at -1000V is

$$x_n = \frac{W_{scl}}{1+N_D/N_A} = \frac{114.6}{1+0.005} \qquad x_p = \frac{W_{scl}}{1+N_A/N_D} = \frac{114.6}{1+200}$$

$$x_n = 114.0\mu\text{m} \qquad x_p = 0.57\mu\text{m}$$

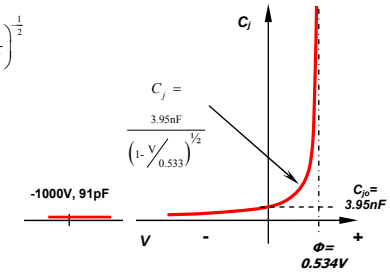
Note that when $N_A >> N_D$, $x_n \approx W_{scl}$, thus the lower the relative concentration of N_D, the deeper the scl penetration and the higher the portion of V supported in N_D. The junction scl can under these circumstances be analysed based on simplified equations – called one-sided junction equations.

v. The charge magnitude each side of the junction, shown in figure 2.6, is given by equation (2.17). The electric field at the junction is given by equation (2.20), while the junction capacitance at -1000V is given by equation (2.19):

$$Q_j = \ Q_0\sqrt{1-\frac{V}{\Phi}} \quad = \quad 4.2\text{nC}\times\sqrt{1+\frac{1000\text{V}}{0.533\text{V}}}$$
$$= \ 182{\cdot}4\,\text{nC}$$

$$\xi_j = \ \xi_0\sqrt{1-\frac{V}{\Phi}} \quad = \quad 0.40\text{M}\times\sqrt{1+\frac{1000\text{V}}{0.533\text{V}}}$$
$$\xi_j = \ 17.5\,\text{MV/m} \qquad \left(<\xi_b = 25\text{MV/m}\right)$$

$$C_j = \ C_{jo}\left(1-\frac{V}{\Phi}\right)^{\frac{1}{2}} = \ 3.95\text{n}\times\left(1+\frac{1000\text{V}}{0.533\text{V}}\right)^{\frac{1}{2}}$$
$$C_j = \ 91\text{pF}$$



Reading list

See chapter 1 reading list.

Fraser, D. A., *The Physics of Semiconductor Devices*, OUP, 1977.

Wolf, H. F., *Semiconductors*, John Wiley-Interscience, New York, 1977.

Van Zeghbroeck, B., *Principles of Semiconductor Devices*, //ece-www.colorado.edu/~bart/book, 2004.

Yang, E. S., *Fundamentals of Semiconductor Devices*, McGraw-Hill, 1978.

Problems

2.1. A silicon diode is to have a breakdown voltage of 1000 V. If breakdown is due to the avalanche mechanism calculate
   i. the concentration of the n^- region
   ii. the width of the n^- region
   iii. the maximum electric field
   iv. the expected punch-through voltage based on parts i. and ii.
   [2×10^{14} /cc, 83μm, 2.4×10^5 V/cm, 1057 V]

2.2. What is the punch-through voltage for a silicon step junction with an n^- doping level of 5×10^{13} /cc and a width of 20μm? Calculate the doping level and scl width for a similarly voltage rated silicon avalanche diode assuming equations (2.3) and (2.5) are valid.
   [15.3 V, 5.27×10^{16} /cc, 0.63μm]

2.3. An abrupt silicon pn junction consists of a p-type region containing 10^{16} cm^{-3} acceptors and an n-type region containing 5×10^{14} cm^{-3} donors. Calculate the built-in potential of this p-n junction.
   When the applied voltage equals 0, 0.5V, and -100V calculate:
   i. the total width of the scl region.
   ii. maximum electric field in the scl region.
   iii. the potential across the scl region in the n-type semiconductor.

2.4. Consider an abrupt pn diode with $N_A = 10^{18}$ cm^{-3} and $N_D = 10^{16}$ cm^{-3}. Calculate the junction capacitance at zero bias if the diode area is 10^{-4} cm^2. Repeat the problem while treating the diode as a one-sided diode and calculate the relative error.

2.5. Repeat example 2.2 using the single-sided diode equations in equation (2.21), where $N_A >> N_D$. Calculate the percentage error in using the assumptions.

2.6. A silicon pn diode with $N_A = 10^{18}$ cm^{-3} has a capacitance of 10^{-7} F/cm^2 at an applied reverse voltage of 1V. Calculate the donor density N_D.

2.7. A silicon pn diode has a maximum electric field magnitude of 10^7 V/cm and a scl width of 200μm. The acceptor concentration is 100 times the donor density. Calculate each doping density.

2.8. Repeat example 2.2 for the equivalent 4H silicon carbide junction diode having the same electrical operating conditions. Use the silicon carbide data given below.

See problems 10.4, 10.5, and 10.9 to 10.12.

Useful *SI* data for silicon and silicon carbide:

$q = -1.6 \times 10^{-19}$ C $\xi_o = 8.85 \times 10^{-12}$ F/m $\xi_{r\,Si} = 11.8$ $\xi_{r\,SiC} = 9.7$ $kT/q = 0.0259$ eV at 300K

$n_{i\,Si} = 1.5 \times 10^{16}$ m^{-3} $n_{i\,SiC} = 2.5 \times 10^{-3}$ m^{-3}

Blank

3

Power Switching Devices and their Static Electrical Characteristics

There is a vast proliferation of power switching semiconductor devices, each offering various features, attributes, and limitations. The principal device families of concern in the power switching semiconductor range are the diode, transistor, and thyristor. Each family category has numerous different members. The basic characteristics of the three families and a range of their members will be presented.

3.1 Power diodes

The homojunction p-n diode is the simplest semiconductor device, comprising one pn junction. In attempts to improve both static and dynamic diode electrical properties for different application conditions, numerous diode types have evolved.

3.1.1 The pn fast-recovery diode

The doping concentration on each side of the junction influences the avalanche breakdown voltage, the contact potential, and the series resistance of the diode. The junction diode normally has the p-side highly doped compared with the n-side, and the lightly doped n-region determines many of the properties of the device. The n-region gives the device its high-voltage breakdown and under reverse bias, the scl penetrates deeply into the n-side. The lower the n-type concentration and the wider the n-side, the higher will be the reverse voltage rating and also, the higher the forward resistance. These n-region requirements can lead to thermal I^2R problems in silicon. Larger junction areas help reduce the thermal instability problem.
It is usual to terminate the lightly doped n-region with a heavily doped n^+ layer to simplify ohmic contact and to reduce the access resistance to the scl. For better n-region width control, n-type silicon is epitaxially grown on an n^+ substrate. The p^+ anode is diffused or implanted into the epitaxial region, forming an epitaxial diode.

In devices specifically designed for high reverse bias applications, care must be taken to avoid premature breakdown across the edge of the die or where the junction surfaces. Premature edge breakdown is reduced by *bevelling* the edge as shown in figure 3.1a, or by diffusing a *guard ring* as shown in figure 3.1b, which isolates the junction from the edge of the wafer. The scl electric field is lower at the bevelled edge than it is in the main body of the device. In the case of a lightly doped p-type guard ring, the scl is wider in the p-ring, because of its lower concentration, than in the p^+ region. The maximum electric field is therefore lower at the pn-ring junction for a given reverse bias voltage. Negatively charged glass film techniques are also employed to widen the scl near the surface, as shown in figures 3.1c and 3.1d. Multiple guard rings are sometimes employed for very high breakdown voltage devices. Similar techniques are extendable to devices other than diodes, such as thyristors. Field

control bevelling on more complex junction structures is achieved with double-negative or double-positive bevelling as shown in parts e and f of figure 3.1. The bevelling is accomplished by grinding, followed by etching of the bevel surface to restore the silicon crystalline mechanical and structure quality. The processed area is passivated with a thin layer of polyimide, which is covered in silicon rubber. Negative bevels tend to be more stable electrically with ageing.
The foregoing discussion is directly applicable to the rectifier diode, but other considerations are also important if fast switching properties are required. The turn-on and reverse recovery time of a junction are minimised by reducing the amount of stored charge in the neutral regions and by minimising carrier lifetimes. *Lifetime killing* is achieved by adding gold or platinum, which is an efficient recombination centre. Electron and proton irradiation are preferred non-invasive lifetime control methods. Irradiation gives the lowest forward recovery voltage and the lowest reverse leakage current. The improved switching times must be traded off against increased leakage current and on-state voltage. Switching times are also improved by minimising the length (thickness) of the n-region.

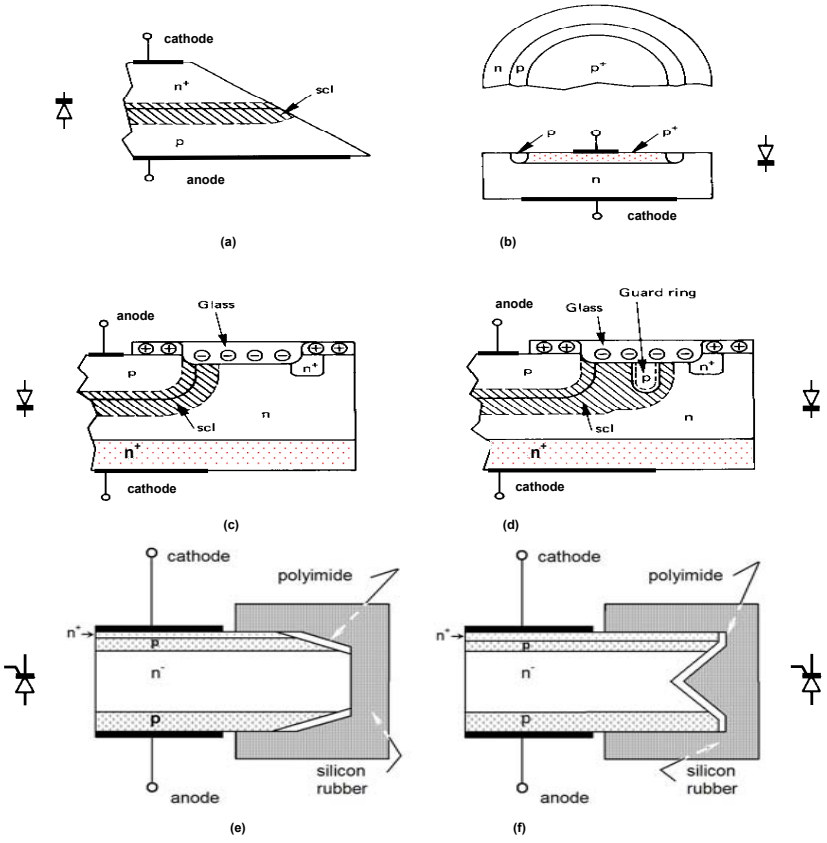


Figure 3.1. *To prevent edge breakdown under junction reverse bias: (a) reduction of the space charge region near the bevel; (b) p-type guard ring; (c) glass guard ring; (d) glass plus p-type guard ring; (e) double negative bevel; and (f) double positive bevel angle.*

3.1.2 The p-i-n diode

The transient performance of diodes tends to deteriorate as the thickness of the silicon wafer is increased in attaining higher reverse voltage ratings. Gold lifetime killing only aggravates the adverse effects incurred with increased thickness. The p-i-n diode allows a much thinner wafer than its conventional pn counterpart, thus facilitating improved switching properties.

The p-*i*-n diode is a pn junction with a doping profile tailored so that an intrinsic layer, the i-region, is sandwiched between the p-layer and the n-layer, as shown in figure 3.2. In practice, the idealised *i-region* is approximated by a high resistivity n-layer referred to as a *v-layer.* Because of the low doping in the *v*-layer, the scl will penetrate deeply and most of the reverse bias potential will be supported across this region.

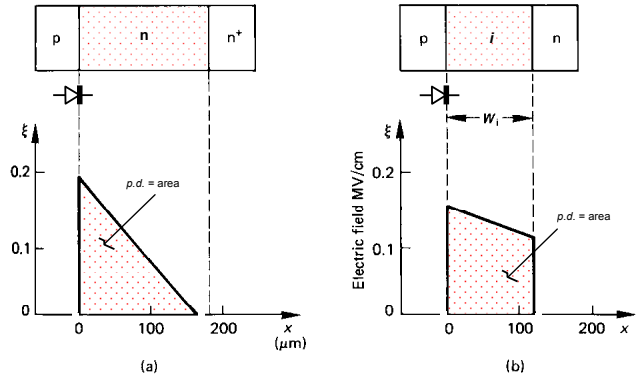


Figure 3.2. *Cross-section and electric field distribution of: (a) a pn diode and (b) a p-i-n diode.*

The power p-*i*-n diode can be fabricated by using either the epitaxial process or the diffusion of p and n-regions into a high-resistivity semiconductor substrate. The i-region width W_i, specifies the reverse voltage breakdown of the p-*i*-n diode, which is the area under the electric field in figure 3.2b, viz.,

$$V_b \approx \xi_b W \approx 25 W_i \quad \text{(in μm)} \qquad \text{(V)} \qquad (3.1)$$

The thickness W_i, along with the distribution of any gold within it, determines the nature of the reverse and forward-conducting characteristics. These characteristics are more effective and efficient in fast p-i-n diodes than in the traditional pn structures.

3.1.3 The power Zener diode

Zener diodes are pn diodes used extensively as voltage reference sources and voltage clamps. The diode reverse breakdown voltage is used as the reference or clamping voltage level.

The leakage current in a good pn diode remains small up to the reverse breakdown point where the characteristic has a sharp bend. Such an electrical characteristic is called *hard*. Premature breakdown at weak spots in the junction area or periphery cause high leakage currents before final breakdown, and such diodes are said to have *soft* breakdown characteristics.

Zener diodes are especially made to operate in the breakdown range. Above a few volts, the breakdown mechanism is avalanche multiplication rather than Zener and the breakdown reference voltage V_Z is obtained by proper selection of the pn junction doping levels. Once in breakdown V_Z remains almost constant provided the manufacturer's power rating, $P = V_Z I$, is not exceeded. Where the breakdown mechanism is due to the Zener effect, the temperature coefficient is negative, about -0.1 per cent/K, changing to positive, +0.1 per cent/K, after about 4.5V when the avalanche multiplication mechanism predominates.

Zener diodes require a hard breakdown characteristic not involving any local hot spots. They are available in a voltage range from a few volts to about 280V and with power dissipations ranging from 250mW to 75W, with heat sinking. Transient suppressing Zener diodes can absorb up to 50kW, provided energy limits and number of cycles are not exceeded, as shown in figure 10.21.

Practically, Zener diodes are difficult to make, less than ideal in application, and should be avoided if possible. The basic *I-V* characteristics, and electrical circuit symbol for the different types of diodes, are shown in figure 3.3.

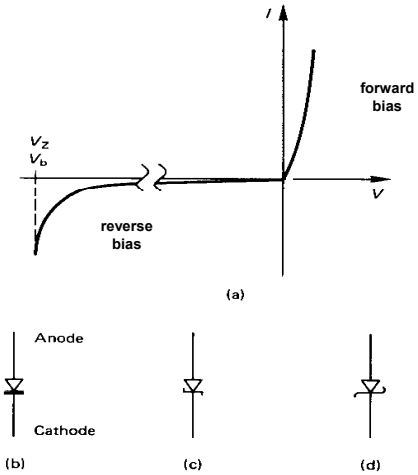


Figure 3.3. *Diodes: (a) static I-V characteristic; (b) symbol for a rectifier diode; (c) voltage reference or Zener diode; and (d) Schottky barrier diode.*

3.1.4 The Schottky barrier diode

The *Schottky diode* is a metal-semiconductor diode device which offers low on-state voltages, but in silicon is presently restricted to applications imposing a reverse bias of less than 400V. At lower voltages, less than 40V, devices of up to 300A are available and the maximum junction operating temperature is 175°C, which is higher than for conventional silicon pn junction devices.

The Schottky diode is formed by a metal (such as chromium, platinum, tungsten or molybdenum) in homogeneous contact with a substrate piece of n-type silicon, as shown in figure 3.4a. The contact is characterised by a potential barrier $\Phi_b > 0$ (termed *Schottky barrier height*) which determines the forward and reverse properties of the Schottky diode.

In forward conduction, electrons are emitted from the negative potential n-type silicon to the positive potential metal, passing over the barrier potential. Unlike the bipolar pn diode, only electrons are carriers, hence the Schottky barrier diode is a unipolar device. The forward on-state voltage drop is dominated by and proportional to the barrier potential Φ_b, while unfortunately the reverse leakage current is approximately inversely related. Thus a Schottky diode with a very low forward voltage drop will have very high reverse leakage current relative to the pn diode counterpart, as shown in figure 3.5.

Chromium provides the lowest forward voltage drop but is limited to an operating temperature of 125°C and has a high leakage current. Platinum allows operating temperatures to 175°C with a leakage current several orders of magnitude lower than chromium. The trade-off is a higher forward voltage.

A guard ring is used to improve device robustness, but its function is to act like a Zener diode and thus protect the Schottky barrier under excessive reverse bias. An optimally designed epitaxial layer, as shown in figure 3.4b, is also employed which reduces the field at the less than perfect metal-semiconductor interface and allows the whole interface to go safely into reverse bias breakdown.

There are a number of important differences between Schottky barrier and pn junction diodes.

- In a pn diode, the reverse bias leakage current is the result of minority carriers diffusing into the scl and being swept across it. This current level is highly temperature-sensitive. In the Schottky-barrier case, reverse current is the result of majority carriers that overcome the barrier. A much higher leakage value results at room temperature, but is not temperature-dependent.
- The forward current is mostly injected from the n-type semiconductor into the metal and very little excess minority charge is able to accumulate in the semiconductor. Since minimal minority carrier recombination occurs, the Schottky barrier diode is able to switch rapidly from forward conduction to reverse voltage blocking.
- Since under forward bias, barrier injection comes only from the semiconductor, and there is little recombination in the scl; thus the device can be represented by the ideal diode equation (2.6).
- The majority electrons injected over the barrier into the metal have much higher energy than the other metal electrons which are in thermal equilibrium. Those injected electrons are therefore called *hot,* and the diode in some applications is referred to as a *hot electron diode*.

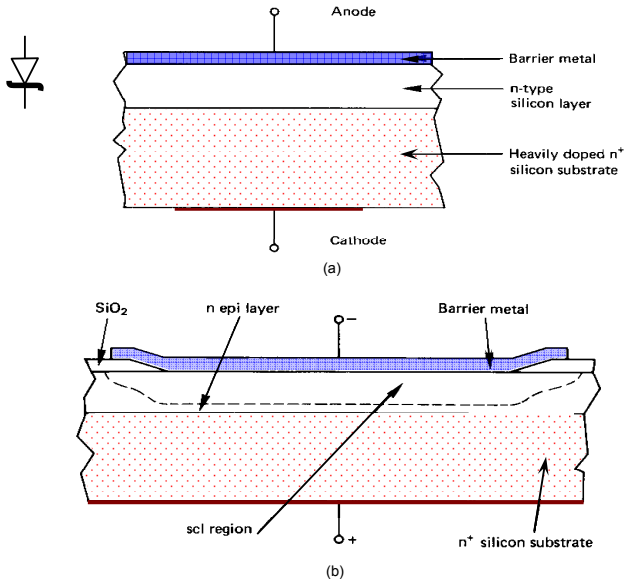


Figure 3.4. *The Schottky barrier diode: (a) the basic structure and (b) the space charge layer region extending into the epi-substrate region under reverse bias.*

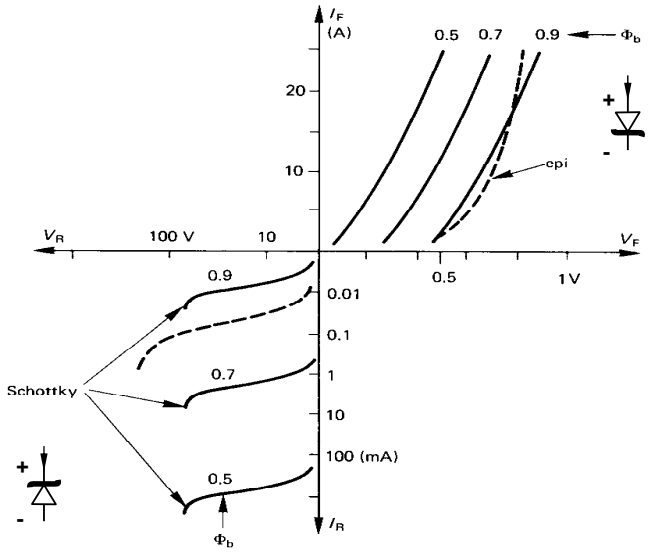


Figure 3.5. *Schottky and epi diode I-V characteristics with different Schottky barrier potentials.*

An important point arising from this brief consideration of the Schottky barrier diode is the importance of the connection of an n-type semiconductor region to aluminium metallization that occurs in unipolar and bipolar semiconductor devices. A practical method of forming aluminium ohmic contacts on n-type materials where $\Phi_b > 0$, is by doping the semiconductor very heavily (>10^{19} /cm^3), above the degeneracy level. Thus, in the contact region, if a barrier exists, the scl width is small enough (<3nm) to appear transparent, allowing electron carriers to tunnel through the barrier in both directions. On the other hand, aluminium makes a good ohmic contact on p-type silicon since the required p^+ surface layer is formed during the heat treatment of the contact after the aluminium is deposited. An ohmic contact acts as a virtual sink for minority carriers, because it has an enormous supply of majority carriers.

3.1.5 The silicon carbide Schottky barrier diode

Silicon carbide Schottky diodes are attractive for high voltages because the field breakdown of silicon carbide is eight times that of silicon. Additionally, the wide band gap allows higher operating temperatures. Both nickel and titanium can be used as Schottky metals. Boron atoms (a dose of 1x10^{15} /cm^2 at 30keV) are implanted to form the edge termination that spreads any field crowding at the edge of the metal contact, as shown in figure 3.6. The lower barrier height of titanium produces a lower forward voltage device, but with a higher reverse leakage current, than when nickel is used as the barrier metal.

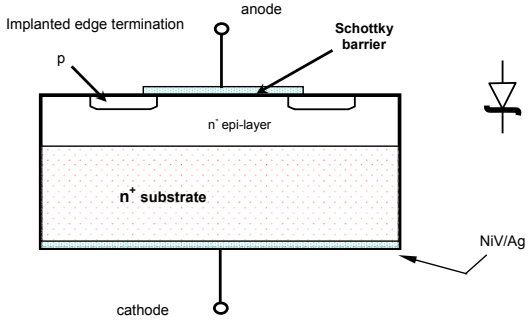


Figure 3.6. *The silicon carbide Schottky barrier diode structure.*

3.2 Power switching transistors

Two types of transistor are extensively used in power switching circuits, namely the power metal oxide semiconductor field effect transistor (MOSFET) and the insulated gate bipolar transistor (IGBT). The IGBT has a bipolar junction transistor (BJT) output stage and a MOSFET input stage, in an integrated *Darlington* pair configuration. Many of the IGBT power handling properties are associated with the limitations of the BJT. Thus some attention to the BJT's electrical characteristics is necessary, even though it is virtually obsolete as a discrete power-switching device. SiC BJT s may offer a short reprieve.

- The **BJT** consists of a pnp or npn single-crystal silicon structure. It operates by the injection and collection of minority carriers, both electrons and holes, and is therefore termed a *bipolar transistor*.
- The **MOSFET** depends on the voltage control of a depletion width and is a majority carrier device. It is therefore a *unipolar transistor*.
- The **IGBT** has the desirable voltage input drive characteristics of the MOSFET but the power switching disadvantages of the minority carrier mechanisms of the BJT.

3.2.1 The bipolar npn power switching junction transistor (BJT)

As a discrete electrical device, the high-voltage, power-switching bipolar junction transistor, BJT, is virtually obsolete. The BJT has one unique redeeming electrical characteristic, viz.; it can be designed to conduct hundreds of amperes with an extremely low on-state voltage of less than 100mV, when saturated. Although superseded, its basic electrical operating characteristics are fundamental to the operation of most other power switching devices. Specifically, the MOSFET has a parasitic npn BJT, as shown in figure 3.14, that can cause false turn-on other than for the fact that understanding of BJT

characteristics allows circumvention of the problem. The fundamental operation of thyristors (SCR, GTO, and GCT) relies totally on BJT characteristics and electrical mechanisms. The IGBT has two parasitic BJTs, as shown in figure 3.16, that form an undesirable pnp-npn SCR structure. Understanding of BJT gain mechanisms allows virtual deactivation of the parasitic SCR.

The first bipolar transistors were mainly pnp, fabricated by alloying techniques and employed germanium semiconductor materials. Most transistors are now npn, made of silicon, and utilise selective diffusion and oxide masking.

A typical high-voltage *triple-diffused* transistor doping profile is shown in figure 3.7a. The n-collector region is the initial high-resistivity silicon material and the collector n^+ diffusion is performed first, usually into both sides. One n^+ diffusion is lapped off and the p-base and n^+ emitter diffusions are sequentially performed.

A *planar epitaxial* structure is often used for transistors with voltage ratings of less than 1000V. The basic structure and processing steps are shown in figure 3.7b. The n-type collector region is an epitaxial layer grown on an n-substrate. The base and emitter are sequentially diffused into the epitaxy. Ion implantation is also used. This approach allows greater control on the depth of the n-type collector region, which is particularly important in specifying device switching and high-voltage properties. Also, the parasitic series collector resistance of the substrate is minimised without compromising the pellet's mechanical strength as a result of a possible reduction in wafer thickness.

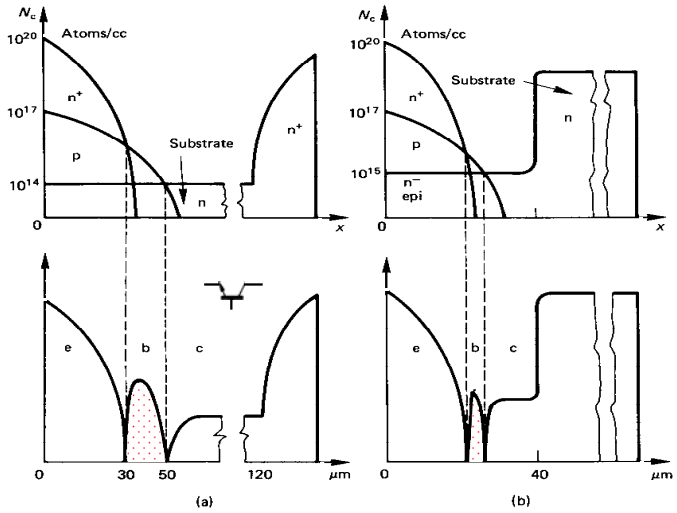


Figure 3.7. *Impurity profile in two types of npn transistors: (a) a planar triple-diffused npn transistor obtained by three consecutive diffusions into a uniformly doped n-type substrate and (b) planar epitaxial npn transistor, obtained after two diffusions into n-type epitaxial layer which is first grown on a low-resistivity n-type silicon substrate.*

3.2.1i - BJT gain

Figure 3.8 shows an npn bipolar junction transistor connected in the *common emitter* configuration. In this configuration, injection of electrons from the lower n^+p junction into the centre p-region supplies minority carrier electrons to participate in the reverse current through the upper np junction.

The n^+ region which serves as the source of injected electrons is called the *emitter* and forms the emitter junction with the *p-base,* while the n-region into which electrons are swept by the reverse bias np junction is called the *collector* and, with the p-base, forms the collector junction.

To have an efficient npn transistor almost all the electrons injected by the emitter into the base should be collected. Thus the p-base region should be narrow and the electron minority carrier lifetime should be long to ensure that the average electron injected at the emitter will diffuse to the collector scl without recombining in the base. The average lifetime of electrons in the p-base increases as the p-base concentration decreases, that is as the hole concentration decreases. The fraction of electrons which reach the collector is called the *base transport factor,* b_t. Electrons lost to recombination in the p-base

must be re-supplied through the base contact. It is also required that the emitter junction carrier flow should be composed almost entirely of electrons injected into the base, rather than holes crossing from the base region to the emitter. Any such holes must be provided by the base current, which is minimised by doping the base region lightly compared with the emitter such that an n^+p emitter results. Such a junction is said to have a high *injection efficiency,* γ_i. A low lattice defect density also increases the injection efficiency.

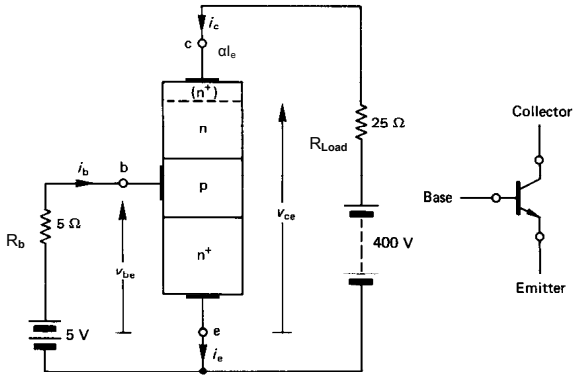


Figure 3.8. *Common emitter junction bias conditions for an npn transistor and the npn bipolar junction transistor circuit symbol.*

The relationship between collector and emitter current is

$$\frac{i_c}{i_e} = b_t \gamma_i = \alpha \tag{3.2}$$

The factor α is called the *current transfer ratio.* Since base current is necessary, α is less than 1, but close to 1, if the BJT

- has a good base transport factor, $b_t \approx 1$
  (narrow base width and with long minority carrier lifetimes) and
- a high emitter injection efficiency, $\gamma_i \approx 1$
  (high emitter doping relative to the base concentration).

In the common emitter configuration shown in figure 3.8, the ratio between the base current i_b and the collector current i_c, is of practical importance. Since the base current is the difference between the emitter and the collector current

$$\frac{i_c}{i_b} = \frac{i_c}{i_e - i_c} = \frac{i_c / i_b}{1 - i_c / i_b} \tag{3.3}$$

$$= \frac{\alpha}{1-\alpha} = \beta \tag{3.4}$$

The factor β, relating the collector current to the base current, is defined as the base-to-collector *current amplification factor.* If α is near unity, β is large, implying the base current is small compared with the collector current.

3.2.1ii - BJT operating states

In power switching applications, a transistor is controlled in two states which can be referred to as the *off-state* or *cut-off state* and the conduction *on-state.* Ideally the transistor should appear as a short circuit when on and an open circuit when in the off-state. Furthermore, the transition time between these two states is ideally zero. In reality, transistors only approximate these requirements.

The typical BJT collector output characteristics are shown in figure 3.9 which illustrates the various BJT operating regions. The *saturated on-state* shown in figure 3.9 occurs when both the collector and emitter junctions are forward biased. Consequently, the collector emitter voltage $V_{ce(sat)}$ is less than the base to emitter saturation voltage $V_{be(sat)}$.

The voltage breakdown phenomenon is of particular importance to the high-voltage, power-switching BJT, and is due to the characteristics of the device structure and geometry.

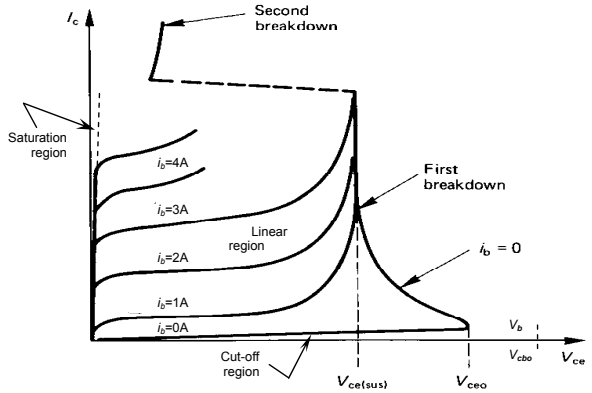


Figure 3.9. *Output characteristics of a common emitter connected transistor showing its operating regions and the voltage breakdown range.*

3.2.1iii - BJT maximum voltage - first and second breakdown

The collector junction supports the off-state voltage and in so doing develops a wide scl. This scl increases in width with increased reverse bias, penetrating into the base. It is unusual that a correctly designed high-voltage power switching BJT would break down as a result of punch-through of the collector scl through the base to the emitter scl. Because of the profile of the diffused base, collector junction voltage breakdown is usually due to the avalanche multiplication mechanism, created by the high electric field at the collector junction. In the common emitter configuration shown in figure 3.9, the transistor usually breaks down gradually, but before the collector junction avalanches at V_b. This occurs because the avalanche-generated holes in the collector scl are swept by the high field into the base. The emitter injects electrons in order to maintain base neutrality. This emitter junction current in turn causes more collector current, creating more avalanche pairs and causing a regenerative action. This voltage dependant avalanche effect is modelled by

$$M = \frac{1}{1-(v_{ce}/V_b)^m} \qquad (3.5)$$

Thus the gain mechanisms of the transistor cause collector to emitter breakdown - *first breakdown,* at voltage V_{ceo}, to occur before collector to base avalanche breakdown, at voltage V_{cbo}, which from $\alpha M = 1$ are related according to

$$V_{ceo} = V_{cbo}(1-\alpha)^{1/m} \approx V_{cbo}/\beta^{1/m}$$
$$= V_b/\beta^{1/m} \quad \text{(V)} \qquad (3.6)$$

where the avalanche breakdown voltage V_b is given by equation (2.3);
$m \approx 6$ for a silicon p^+n collector junction; and
$m \approx 4$ for a silicon n^+p collector junction.

Contradictory device properties are that the higher the forward gain, the lower the breakdown voltage. A much higher collector emitter breakdown voltage level can be attained if the base emitter junction is reverse biased in the off-state.

First breakdown need not be catastrophic provided junction temperature limits are not exceeded. If local hot spots occur because of non-uniform current density distribution as a result of crystal faults, doping fluctuation, etc., *second breakdown* occurs. Silicon crystal melting and irreparable damage result, the collector voltage falls, and the current increases rapidly as shown in figure 3.9.

3.2.2 *The metal oxide semiconductor field effect transistor (MOSFET)*

The basic low-power lateral structure of the enhancement mode, metal oxide semiconductor, field effect transistor (MOSFET) is illustrated in figure 3.10a. The n^+ *source* and *drain* regions are diffused or implanted into the relatively lightly doped p-type substrate, and a thin silicon dioxide layer insulates the aluminium *gate* from the silicon surface. No lateral current flows from the drain to source without a conducting n-channel between them, since the drain-to-source path comprises two opposing series pn junctions.

When a positive gate voltage is applied with respect to the source as shown in figure 3.10b, positive charges are created on the metal gate. In response, negative charges are induced in the underlying silicon, by the formation of a depletion region and a thin surface region containing mobile electrons. Effectively the positive gate potential *inverts* the p-channel, forming an electron-enhanced low-resistance *n-channel,* allowing current to flow freely in either direction between the drain and source. The inversion channel is essentially devoid of the thermal properties associated with the typical BJT.

An important parameter in mos transistors is the *threshold voltage* V_{Th}, which is the minimum positive gate voltage to induce the n-conducting channel. With zero gate voltage the structure is normally off. The device is considered to operate in the *enhancement mode* since the application of a positive gate voltage in excess of V_{Th} induces an n-conducting channel. The typical output characteristics of the MOSFET are shown in figure 3.10c.

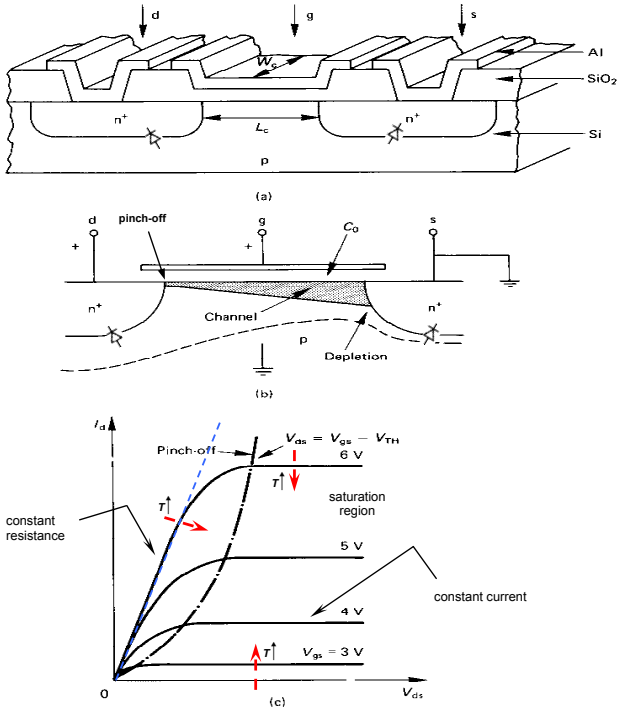


Figure 3.10. *Enhancement-type n-channel mos transistor:*
(a) device cross-section; (b) induced n-channel near pinch-off; and (c) drain I-V characteristics as a function of gate voltage, showing the pinch-off locus and effects of increased temperature.

3.2.2i - MOSFET structure and characteristics

The conventional horizontal structure in figure 3.10a has severe limitations associated with increasing die area that make it uneconomical for consideration as a viable high-current structure. A planar vertical n-channel dmos structure like those shown in figure 3.11 is used to overcome the inherent poor area utilisation of the basic mos structure. The enclosing peripheral p floating field guard is not shown.

The dmos structure is a vertical current flow device. An n^- epitaxial layer is grown on an n^+ substrate. A series of p body regions are next diffused into the epitaxial layer. Then n^+ source regions are diffused within the p body regions and a polycrystalline silicon gate is embedded in the silicon dioxide insulating layer. Source and gate metallization are deposited on the top surface of the die and the drain contact made to the bottom surface. Cell density is inversely related to voltage rating and varies from 200,000 to 1,000,000 cells per cm^2.

With a positive gate voltage, the device turns on and majority carriers flow laterally from the source to the drain region below the gate and vertically to the drain contact. Current can also flow freely in the reverse direction in the channel since the channel is bipolar conducting once enhanced.

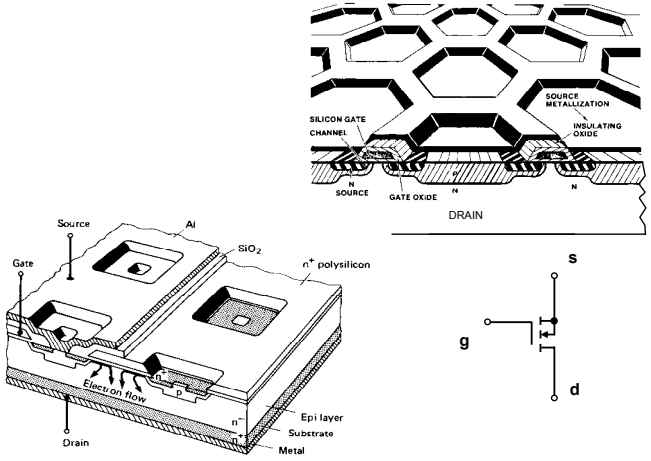


Figure 3.11. *Two designs for the n-channel MOSFET and its circuit symbol (courtesy of Infineon and International Rectifier).*

The obtainable drain-to-source breakdown voltage is not limited by the gate geometry. The scl associated with voltage blocking penetrates mostly in the n-type epitaxial layer. Thickness and doping concentration of this layer are thus decisive in specifying the blocking capability of the power MOSFET.
The basic drain current versus drain to source voltage static operating characteristics (and their temperature dependence) of the power MOSFET are illustrated in figure 3.10c. For a given gate voltage, there are two main operating regions on the drain current-voltage characteristic.

- The first is a *constant resistance* region, where an increase in drain to source voltage results in a proportional increase in drain current. (In practice, the effective resistance increases at higher drain currents.)
- At a certain drain current level, for a given gate voltage, a channel pinch-off effect occurs and the operating characteristic moves into a *constant current* region.

3.2.2ii - MOSFET drain current

When the power MOSFET is used as a switch, it is controlled in the on-condition, such that it is forced to operate in the resistive region. This ensures that the voltage drop across the device is low so that the drain current is essentially defined by the load and the device power dissipation is minimal. Thus for switching applications, the on-resistance $R_{ds(on)}$ is an important characteristic because it will specify the on-state power loss for a given drain current. The lower $R_{ds(on)}$ is, the higher the current-handling capabilities of the device; thus $R_{ds(on)}$ is one important figure of merit of a power MOSFET.
A quadratic MOSFET model allows the inversion layer charge between the source and the drain to vary. For power MOSFETs that have short channels, the drain current I_d is related to the channel dimensions and the gate voltage V_{gs} according to

at low current, above pinch-off

$$I_d = \tfrac{1}{2}\mu \frac{W_c}{L_c} C_a (V_{gs} - V_{Th})^2 \quad \text{(A)} \tag{3.7}$$

if $V_{ds} \geq V_{gs} - V_{Th} > 0$ for n-channel MOSFETs, as shown to the left of the pinch-off locus in figure 3.10c.

at high current after electron velocity saturation, the quadratic model is invalid and

$$I_d = \tfrac{1}{2} v_{sat} W_c C_a (V_{gs} - V_{Th}) \quad \text{(A)} \tag{3.8}$$

where C_a is the capacitance per unit area of the gate oxide (ε / t_{ox})
W_c is the width of the channel
v_{sat} is the saturation velocity of electrons in silicon, (9.0×10^6 cm/s)
L_c is the effective channel length
μ is the conducting channel carrier mobility, (300 cm^2/ V-s).
In the ohmic (linear) region, where $V_{gs} > V_{th}$ and $V_{gs} - V_{th} > V_{ds} > 0$, the drain current is given by

$$I_d = \mu \frac{W_c}{L_c} C_a \left(V_{gs} - V_{TH}\right) V_{ds} - \tfrac{1}{2} V_{ds}^2 \tag{3.9}$$

and when the gate voltage is below the threshold level, $V_{gs} < V_{Th}$,

$$I_d = 0 \tag{3.10}$$

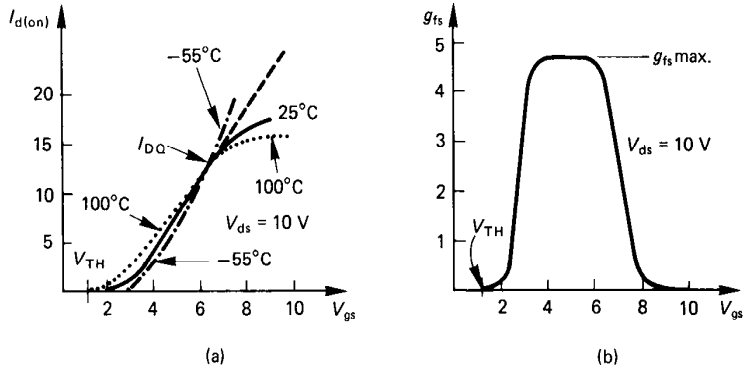


Figure 3.12. *MOSFET gate voltage characteristics:*
(a) transfer characteristics of gate voltage versus drain current and
(b) transconductance characteristics of gate voltage versus transconductance, g_{fs}.

Figure 3.12a shows that drain current exhibits both a positive and negative temperature coefficient with the drain current I_{DQ} being the boundary condition. If the drain current is greater than I_{DQ} there is a possibility of destruction by over-current at low temperatures, while if the drain current is less than I_{DQ}, over-current can produce thermal runaway and destruction. Operation with a gate voltage corresponding to I_{DQ} avoids the need for any gate drive temperature compensation.
At high gate voltages, the on-resistance of the resistive region and the drain current in the constant current region, become somewhat independent of the gate voltage. This phenomenon is best illustrated in the I_d vs V_{ds} characteristic by the curve cramping at high gate voltages in figure 3.10c.

3.2.2iii - MOSFET transconductance and output conductance

Inspection of the static drain source characteristics of figure 3.10c reveals that as the gate voltage increases from zero, initially the drain current does not increase significantly. Only when a certain threshold gate voltage, V_{Th}, has been reached, does the drain current start to increase noticeably. This is more clearly illustrated in figure 3.12b which shows the characteristics of drain current I_d and small signal transconductance g_{fs} versus gate voltage, at a fixed drain voltage. It will be seen from these characteristics that no conduction occurs until V_{gs} reaches the threshold level, V_{Th}, after which the I_d versus V_{gs} characteristic becomes linear, the slope being the transconductance g_{fs}.
The amplification factor, forward *transconductance, g_{fs},* is defined as

$$g_{fs} \triangleq \left| \frac{\partial I_d}{\partial V_{gs}} \right|_{V_{ds}=\text{constant}}$$

Differentiating equations (3.7) and (3.8), for $V_{ds} \geq V_{gs} - V_{Th}$, with respect to gate voltage, gives

at low current

$$g_{fs} = \mu \frac{W_c}{L_c} C_a (V_{gs} - V_{Th}) = \sqrt{2 \frac{W_c}{L_c} \mu_n C_a I_{Dn}} \quad \text{(mho)} \tag{3.11}$$

at high current

$$g_{fs} = \tfrac{1}{2} v_{sat} W_c C_a \quad \text{(mho)} \tag{3.12}$$

At high electric fields, that is high currents, the carrier velocity v_{sat} saturates.
In the ohmic region, $V_{gs} > 0$ and $V_{gs} - V_{th} > V_{ds}$, the forward transconductance is

$$g_{fs} = \mu \frac{W_c}{L_c} C_a V_{ds} \qquad (3.13)$$

The *output conductance*, g_d, is defined as

$$g_d \triangleq \left| \frac{\partial I_d}{\partial V_{ds}} \right|_{V_{gs}=\text{constant}}$$

The output conductance quantifies the drain current variation with gate voltage variation for a constant gate voltage.

Differentiating equations (3.7) and (3.8), with respect to drain voltage, gives zero, g_d = 0, for each case in the saturation region. In the ohmic region the output conductance is

$$g_d = \mu \frac{W_c}{L_c} C_a \left(V_{gs} - V_{th} - V_{ds} \right) \qquad (3.14)$$

A typical minimum threshold voltage is about 2V and exhibits temperature dependence of approximately -10mV per K (α = 0.5 per cent/K), as shown in figure 3.13. At high gate voltages, the drain current becomes constant as the transconductance falls to zero, implying the upper limit of forward drain current. The temperature variation of transconductance is small, typically -0.2 per cent/K, which results in extremely stable switching characteristics. The typical temperature coefficient for the gain of a bipolar junction transistor, the MOSFET equivalent to g_{fs}, is +0.8 per cent/K. The temperature dependence of the MOSFET forward conductance is approximated by

$$g_{fs}(T) \approx g_{fs}(25°\text{C}) \times \left(\frac{T}{300} \right)^{-2.3} \qquad \text{(mho)} \qquad (3.15)$$

since temperature effects are dominated by mobility variation with temperature.

Inherent in the MOSFET structure are voltage-dependent capacitances and on-state resistance.

3.2.2iv - MOSFET on-state resistance

In the fully on-state the drain-source conduction characteristics of the MOSFET can be considered as purely resistive. The *on-resistance* $R_{ds(on)}$ is the sum of the epitaxial region resistance, the channel resistance, which is modulated by the gate source voltage, and the lead and connection resistance. One reason for the wide proliferation of special gate geometries is to produce extremely short, reproducible channels, in order to reduce $R_{ds(on)}$. In high-voltage devices, the on-resistance is dominated by the resistance of the epitaxial drain region when the device is fully enhanced. For high-voltage n-channel devices, the on-state resistance is approximated by

$$R_{ds(on)} = 6.0 \times 10^{-7} \times V_b^{2.5} / A \qquad (\Omega) \qquad (3.16)$$

where V_b is the breakdown voltage in volts
A is the die area in mm^2.

A p-channel device with the same V_b as an n-channel device has an $R_{ds(on)}$ two to three times larger as given by

$$R_{ds(on)} = 1.6 \times 10^{-6} \times V_b^{2.5} / A \qquad (\Omega) \qquad (3.17)$$

The factor $1/g_{fs}$ of $R_{ds(on)}$ is added to give the total $R_{ds(on)}$. On-state drain-source loss can therefore be based on $I_d^2 R_{ds(on)}$. On-resistance $R_{ds(on)}$ increases with temperature and approximately doubles over the range 25°C to 200°C, having a positive temperature coefficient of approximately +0.7 per cent/K above 25ºC, as shown in figure 3.13. The temperature dependence of the on-state resistance is approximated by

$$R_{ds(on)}(T) = R_{ds(on)}(25°\text{C}) \times \left(\frac{T}{300} \right)^{2.3} \qquad (\Omega) \qquad (3.18)$$

where the temperature T is in degrees Kelvin. This relationship (as does forward conductance in equation (3.15)) closely follows the mobility charge dependence with temperature.

Since $R_{ds(on)}$ increases with temperature, current is automatically diverted away from a hot spot. Thus unlike the bipolar junction transistor, second breakdown cannot occur within the MOSFET. The breakdown voltage V_b has a positive temperature coefficient of typically 0.1 per cent/K as shown by $V_{(BR)DSS}$ in figure 3.13.

3.2.2v - MOSFET p-channel device

P-channel MOSFETs are very similar to n-channel devices except that the n and p regions are interchanged. In p-channel devices the on-resistance, for a given die area, will be approximately twice that of a comparable n-channel device. The reason for this is that in the n-channel device the majority carriers are electrons but in the p-channel device, the majority carriers are holes which have lower mobility. If the area of a p-channel device is increased to produce an equal $R_{ds(on)}$, then the various capacitances of the p-channel device will be larger, and the device costs will be greater.

In the linear region, the drain current is

$$-I_{dp} = \frac{W_c \mu_p C_a}{L_c} \left[(V_{gs} + V_{Thp}) V_{ds} - \tfrac{1}{2} V_{ds}^2 \right]$$

For saturation

$$-I_{dp} = \tfrac{1}{2} \frac{W_c \mu_p C_a}{L_c} \left[(V_{gs} + V_{thp})^2 \right]$$

The transconductance in the saturation region is

$$g_{fsp} = \left. \frac{\partial i_{Dp}}{\partial v_{gs}} \right|_Q = \left(\frac{W_c}{L_c} \right) \mu_p C_a (V_{gs} + V_{Thp}) = \sqrt{2 \left(\frac{W_c}{L_c} \right) \mu_p C_a (-I_{dp})}$$

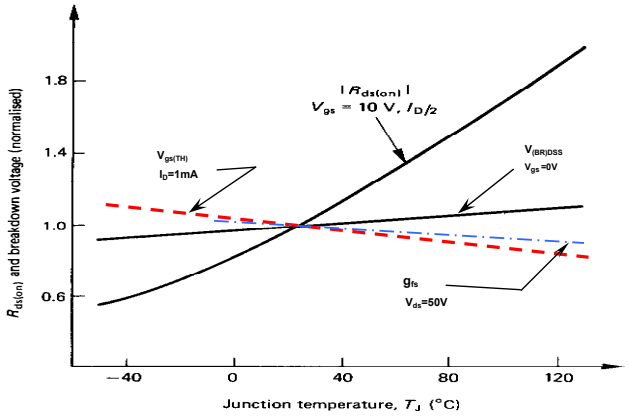


Figure 3.13. *Normalised drain-source on-resistance, transconductance, gate threshold voltage, and breakdown voltage versus junction temperature.*

Example 3.1: Properties of an n-channel MOSFET cell

A silicon n-channel MOSFET cell has a threshold voltage of V_{Th} = 2V, W_c = 10μm, L_c = 1μm, and an oxide thickness of t_{ox} = 50nm. The device is biased with V_{gs} = 10V and V_{ds} = 15V.

i. Assuming a quadratic model and a surface carrier mobility of 300 cm^2/V-s, calculate the drain current, cell dissipation, forward transconductance, and output conductance.

ii. Assuming carrier velocity saturation (5x10^6 cm/s), calculate the drain current, cell dissipation, forward transconductance, and output conductance.

Solution

i. The MOSFET is biased in saturation since $V_{ds} > V_{gs} - V_{Th}$. Therefore, from equation (3.7) the drain current equals:

$$I_d = \tfrac{1}{2} \mu C_a \frac{W_c}{L_c} (V_{gs} - V_{Th})^2 \quad \text{where} \quad C_a = \varepsilon / t_{ox}$$

$$= \tfrac{1}{2} \times 300 \times 10^{-4} \times \frac{3.85 \times 8.85 \times 10^{-12}}{50 \times 10^{-9}} \times \frac{10\mu\text{m}}{1\mu\text{m}} \times (10\text{V} - 2\text{V})^2 = 6.5\ \text{mA}$$

The dc power dissipation is 6.5mAx15V=97.5mW.
From equation (3.11), the transconductance equals:

$$g_{fs} = \mu C_a \frac{W_c}{L_c}(V_{gs} - V_{Th})$$

$$= 300\times10^{-4}\times\frac{3.85\times8.85\times10^{-14}}{50\times10^{-9}}\times\frac{10}{1}\times(10\text{V} - 2\text{V}) = 1.64 \text{ mho}$$

The output conductance g_d is zero.

ii. When the electron velocity saturates, the drain current is given by equation 3.8

$$I_d = ½ v_{sat} W_c C_a (V_{gs} - V_{Th})$$

$$= ½ \times 5\times10^{4} \times 10^{-5} \times \frac{3.85\times8.85\times10^{-12}}{50\times10^{-9}} \times (10\text{V} - 2\text{V}) = 136 \text{ mA}$$

The dc power dissipation is 136mA x 15V=2W, a dc operating condition well in excess of the cell capabilities.

The transconductance is given by equation 3.10

$$g_{fs} = ½ v_{sat} W_c C_a$$

$$= ½ \times 5\times10^{4} \times 10^{-5} \times \frac{3.85\times8.85\times10^{-12}}{50\times10^{-9}} = 16.1 \text{ mho}$$

The output conductance g_d is zero.

♣

3.2.2vi - MOSFET parasitic BJT

Figure 3.14 shows the MOSFET equivalent circuit based on its structure and features. The parasitic npn bipolar junction transistor shown in figure 3.14b is key to device operation and limitations.

Capacitance exists within the structure from the gate to the source, C_{gs}, the gate to the drain, C_{gd}, and from the drain to the source, C_{ds}. The capacitance C_{gs} varies little with voltage; however C_{ds} and C_{gd} vary significantly with voltage. Obviously these capacitances influence the switching intervals, an aspect considered in chapter 4.4.2.

The emitter of the parasitic npn transistor is the source of the MOSFET, the base is the p-type body and the collector is the drain region. In the construction of the MOSFET, the emitter and base of the npn transistor are purposely shorted out by the source metallization to disable the parasitic device by reducing its injection efficiency. However, this short circuit cannot be perfect and R_{be} models the lateral p-body resistance, while C_{ob} is essentially C_{ds}. The npn transistor has a collector-emitter breakdown voltage, between V_{cbo} and V_{ceo}. If an external *dv/dt* is applied between the drain and source as shown in figure 3.14b, enough displacement current could flow through C_{ob} to generate a voltage drop across R_{be} sufficient to turn on the parasitic bipolar device, causing MOSFET failure in second breakdown.

When the drain to source voltage is negative, current can flow from the source to drain through R_{be} and the base to collector junction of the parasitic npn transistor within the structure, the dashed line shown in figure 3.14b. This is termed the *body diode*, inherent in the MOSFET structure.

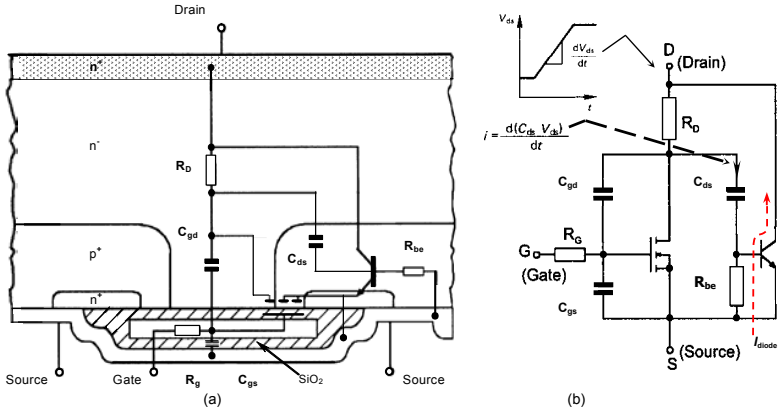


Figure 3.14. *MOSFET – n-channel enhancement mode: (a) structure and (b) equivalent circuit diagram with parasitic npn bipolar transistor forming an inverse diode.*

3.2.2vii - MOSFET on-state resistance reduction

Most power switching devices have a *vertical structure*, where the gate and source of the MOSFET (or emitter in the case of the IGBT) are on one surface of the substrate, while the drain (or collector) is on the other substrate surface. The principal current flows vertically through the substrate but the conductive channel is lateral due to the *planar gate structure*, as shown in figure 3.11. The structure resistance components between the drain and source are:

- the drift region;
- the JFET region;
- the accumulation region; and
- the channel region.

The drift region contribution dominates whilst the contribution from the ohmic contacts and n^+ substrate are not significant, in high voltage devices. The channel voltage drop is proportional to channel length and inversely related to width. The channel should therefore be short, but its length is related to voltage rating since it must support the off-state scl.

Whilst retaining the necessary voltage breakdown length properties, two basic approaches have been pursued to achieve a more vertical gate (channel) structure, viz., the trench gate and vertical super-junction, as shown in parts b and c of figure 3.15. Both techniques involve increased fabrication complexity and extra costs.

1 - Trench gate

A channel is formed on the vertical sidewalls of a trench etched into the die surface as shown in figure 3.14b. The JFET resistive region is eliminated, which not only reduces the total resistance but allows smaller cell size thereby increasing channel density and decreasing the short-circuit capacity. The trench corners must be rounded to avoid high electric field stress points. By extending the gate into the drift region, the gate to drain capacitance increases, hence increasing gate charge requirements.

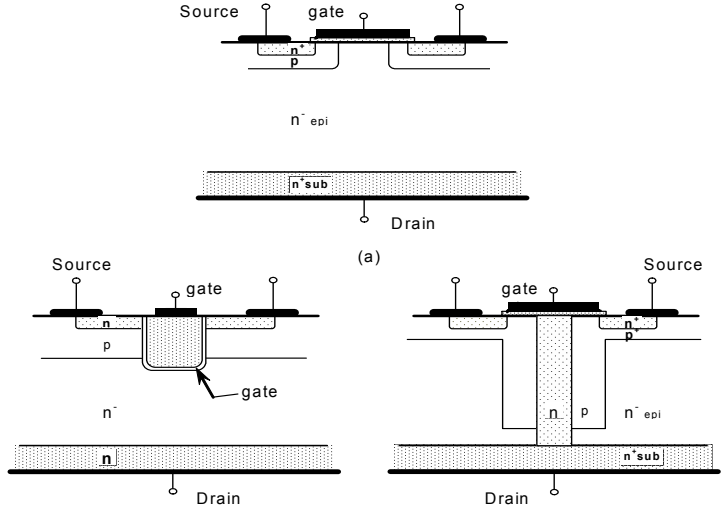


Figure 3.15. *Three MOSFET channel structures: (a) conventional planar gate; (b) trench gate; and (c) vertical superjunction.*

2 - Vertical super-junction

The structure has vertical p-conducting regions in the voltage sustaining n^- drift area, that are extend to the p-wells below the gate, as shown in figure 3.15c. In the off-state, the electric field is not only in the vertical direction but also in the horizontal plane. This means the n-drift region width can be decreased, the on-state resistance is decreased, and the gate charge is reduced for a given surface area. Up to sixteen mask steps are needed which involves repeated cycles of n-type epi-layer growth, masked boron implantation, and finally diffusion. The resultant specific resistance is near linearly related to breakdown voltage, as opposed to $R_{ds(on)} \times Area \propto V_{br}^{2.5}$, equation (3.16). Typically $R_{ds(on)}$ is five times lower than for the conventional MOSFET, which only uses up to six mask steps.

Whilst the trench gate concept can be readily applied to other field effect devices without voltage rating limits, the vertical super-junction is confined to the MOSFET, and then at voltage ratings below about 1000V.

3.2.3 *The insulated gate bipolar transistor (IGBT)*

The high off-state and low on-state voltage characteristics of the bipolar junction transistor are combined with the high input impedance properties of the MOSFET to form the insulated gate bipolar transistor, IGBT, as shown in figure 3.16. The basic structure is that of a MOSFET but with a p^+ implanted into the drain region. This p^+ collector provides reverse blocking capabilities of typically 40V, which can be enhanced if p-wells through the substrate are used to isolate the die periphery.

3.2.3i - IGBT at turn-on

When the IGBT is in the forward blocking mode, and if the positive gate bias (threshold voltage) is applied, which is enough to invert the surface of p-base region under the gate, then an n-type channel forms and current begins to flow. Simultaneously the anode-cathode voltage must be above 0.7V, the potential barrier, so that it can forward bias the p^+ substrate / n^- drift junction, J1. The electron current, which flows from the n^+ emitter via the channel to the n^- drift region, is the base drive current of the vertical pnp transistor. It induces the injection of hole-current from the p^+ region to the n- base region. The conductivity modulation improves because of this high-level injection of minority carriers holes. This increases the conductivity of the drift region, significantly reducing the drift region resistance, which is why the IGBTs can be used in high voltage applications. Two currents flow into the emitter electrode. One is the MOS electron-current flowing through the channel, and the other is the bipolar hole-current flowing through the p^+ body / n^- drift junction, J2.

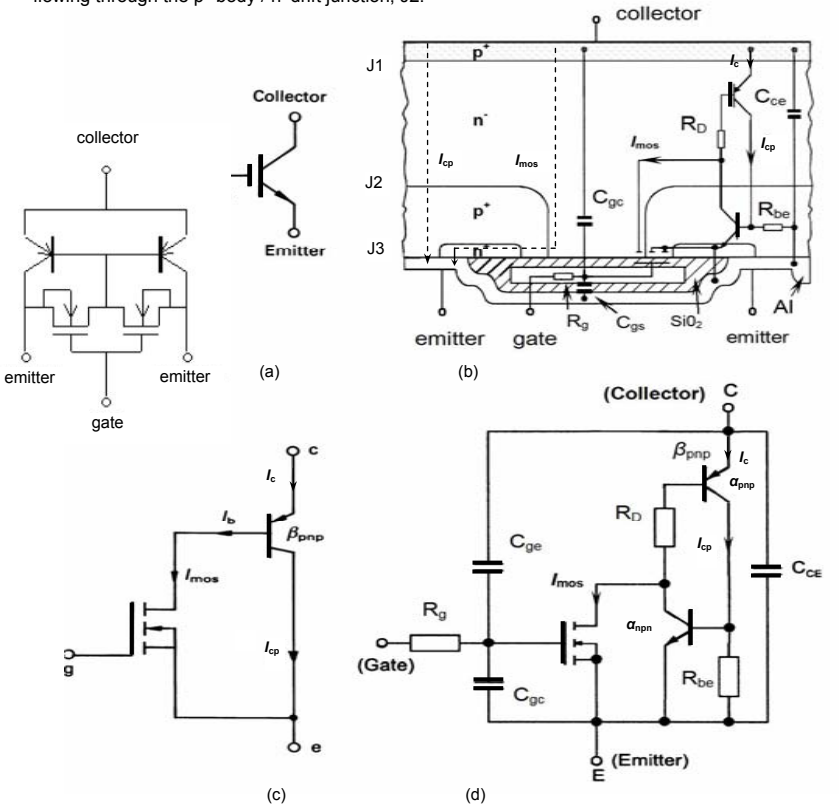


Figure 3.16. *Insulated gate bipolar transistor (IGBT): (a) circuit symbol; (b) physical structure showing current paths: (c) normal operation equivalent circuit; and (d) high current latching equivalent circuit.*

3.2.3ii - IGBT in the on-state

The p^+ substrate conductively modulates the n^- region with minority carriers, which whilst conducting the main collector current, produces a low on-state voltage at the expense of a 0.6 to 0.8V offset in the output voltage characteristics due to the collector pn junction. From figure 3.16c, the IGBT collector current is approximated by

$$I_c = I_{mos}(1+\beta_{pnp}) \tag{3.19}$$

3.2.3iii - IGBT at turn-off

The gate must be shorted to the emitter or a negative bias must be applied to the gate. When the gate voltage falls below the threshold voltage, the inversion layer cannot be maintained, and the supply of electrons into the n^- drift region is blocked, whence, the turn-off process begins. However, the turn-off cannot be quickly completed due to the high concentration minority carrier injected into the n^- drift region during forward conduction. Initially, the collector current rapidly decreases due to the termination of the electron current through the channel (MOSFET turn-off), and then the collector current gradually reduces, as the minority carrier density decays due to recombination, in the externally inaccessible n^- drift region. This storage charge produces a tail current.

The operational mechanisms are those of any minority carrier device and result in slower switching times than the majority carrier MOSFET. On-state voltage and switching characteristics can be significantly improved by using the trench gate technique used on the MOSFET, as considered in section 3.2.2 and shown in figure 3.15b. A less stable structure improvement involves using wider trenches, judiciously spaced, so that accumulated holes under the trench, enhance emitter injection of electrons. This injection enhancement reduces the on-state voltage without degrading the switching performance.

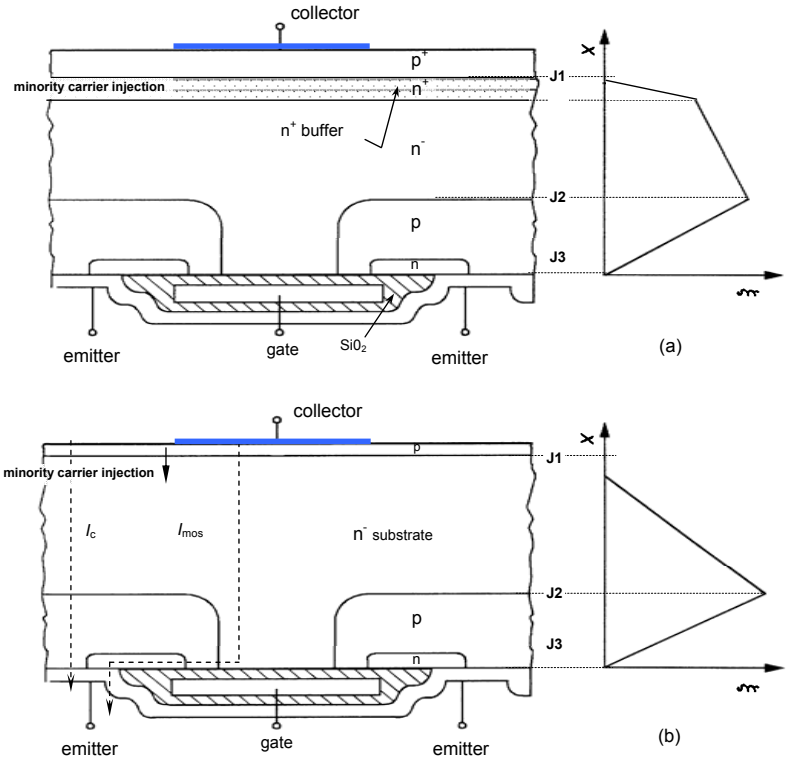


Figure 3.17. *Insulated gate bipolar transistor structures and electric field profile: (a) fieldstop PT-IGBT and (b) conventional NPT-IGBT.*

Further performance enhancement is gained by using the punch through, PT-IGBT, structure shown in figure 3.17a, which incorporates an n^+ buffer region. The conventional non-punch through NPT-IGBT structure is shown in figure 3.17b. Both collector structures can have the same emitter structure, whether a lateral gate as shown, or the MOSFET trench gate in figure 3.15b.

Figure 3.17 shows the electric field in the off-state, where the PT-IGBT develops a field as in the p*i*n diode in figure 3.2b, which allows a thinner wafer. The NPT-IGBT requires a thicker wafer (about 200µm for a 1200V device) which results in a larger substrate resistance and a slower switching device.

- **The PT-IGBT** has n^+ and p^+ layers formed by epitaxial growth on an n^- substrate. The electric field plot in figure 3.17a shows that the off-state voltage scl consumes the n^- substrate and is rapidly reduced to zero in the n^+ buffer.
- **The NPT-IGBT** has a lightly doped n^- substrate with the p-regions (p wells and p collector) formed by ion implantation. The electric field distribution in figure 3.17b shows that the n^- drift region has to be wide enough to support all the off-state voltage, without punch through to the p collector implant.

3.2.3iv - IGBT latch-up

The equivalent circuit in figure 3.16d shows non-ideal components associated with the ideal MOSFET. The parasitic npn bipolar junction transistor (the n^+ emitter/ p^+ well/ n^- drift region are the npn BJT e-b-c) and the pnp transistor (p^+ collector/ n^- drift/ p^+ well are the pnp BJT e-b-c) couple together to form an SCR thyristor structure, as considered in section 3.3. Latching of this parasitic SCR can occur:

- in the on-state if the current density exceeds a critical level, which adversely decreases with increased temperature or
- during the turn-off voltage rise when the hole current increases in sensitive regions of the structure due to the charge movement associated with the scl widening.

1 - IGBT on-state SCR static latch-up is related to the temperature dependant transistor gains which are related to the BJT base transport factor b_t and emitter injection efficiency γ_i, defined for the BJT in equation (3.2)

$$\alpha_{pnp} + \alpha_{npn} = b_{t_{pnp}}\gamma_{i\,pnp} + b_{t_{pnp}}\gamma_{i\,pnp} = 1 \qquad (3.20)$$

Since the conductivity of the drift region under the gate electrode is increased by the introduction of electron current through the channel, most of the holes injected into the drift region are injected at the p-body region under the channel and flow to the source metal along the bottom of n^+ source. This produces a lateral voltage drops across the shunting resistance (R_{be} in figure 3.16b) of the p-body layer. If this voltage drop becomes greater than the potential barrier of the n^+ source / p body layer junction, J3, electrons are injected from the n^+ source to the p-body layer, and the parasitic npn transistor (n^+ source, p body and n^- drift) is turned-on. If the sum of the two (npn and pnp) parasitic transistors' current gains reach unity in equation (3.20), latch-up occurs.
To avoid loss of control and possible IGBT failure, the factors in equation (3.20), which is valid for on-state latch-up, are judiciously adjusted in the device design.

Common to both device types is the gate structure, hence the base-emitter junction of the npn parasitic BJT have the same properties. In each structure, the shorting resistor R_{be} decreases the injection efficiency of the npn BJT emitter. This resistance is minimized by highly doping the p^+ wells directly below the n-emitters and by shortening the length of the n-emitter. The gain α_{npn} in equation (3.20) is decreased since the injection efficiency $\gamma_{i\,npn}$ is lowered.

Reduction of the pnp BJT gain of the PT-IGBT and NPT-IGBT is achieved with different techniques.

- ***For the NPT-IGBT***, the emitter injection efficiency of holes from the p^+ zone into the n^- drift region is high because of the large difference in doping concentrations at the junction. Adversely this yields a high injection efficiency γ_{ipnp}. The base transport factor $b_{t\,pnp}$ is already low because of the large width of the n^- drift region, and is further reduced by lifetime killing of minority carriers in the n^- drift region by using gold doping or electron beam radiation.
- ***For the PT-IGBT***, the p^+ emitting junction at the collector is a well-controlled shallow implant thus reducing the injection efficiency $\gamma_{i\,pnp}$. Charge carrier lifetime killing in the n^- drift region to reduce the base transport factor $b_{t\,pnp}$, is therefore not necessary.

2 - IGBT turn-off SCR dynamic latch-up can occur while the collector voltage is rising, before the collector current decreases. When the IGBT is switched off, the depletion layer of the n^- drift / p-body junction, J2 in figure 3.17, is abruptly extended, and the IGBT latches up due to the displacement current. This limits the safe operating area. Equation (3.20) is modified by equation (3.5) to account for voltage avalanche multiplication effects.

$$M_{npn}\alpha_{npn} + M_{pnp}\alpha_{pnp} = 1$$

where $M = \dfrac{1}{1-(v_{ce}/V_b)^m}$ (3.21)

This dynamic latch-up mode is adversely affected by increased temperature and current magnitude during the voltage rise time at turn-off.
Since $v_{ce} \ll V_b$, $M \to 1$, and the multiplication effect is not significant in the on-state static latch-up analysis. IGBTs are designed and rated so that the latch-up current is 10 to 15 times the rated current.

High temperature characteristics (latching current density)
With a rise in temperature, the current gains of the npn and pnp transistors increase. This decreases the latching current. The effect is aggravated by an increase in the resistance of the p base region due to a decrease in hole mobility.

3.2.4 Reverse blocking NPT IGBT

The conventional IGBT inherently has reverse voltage blocking capabilities, albeit low. Normally, the collector boron ion p+ implant forms a transparent abrupt junction, optimised for on-state voltage and turn-off speed.
When negative voltage is impressed at the collector in figure 3.16, the p+ substrate / n^- drift junction, J1, is reverse biased, and the depletion layer expands to the n^- drift region. An optimal design in resistivity and thickness for the n^- drift region is necessary in obtaining desirable reverse blocking capability. The width of the n^- drift region is equivalent to the sum of depletion width at maximum operating voltage and minority carrier diffusion length. It is important to optimize the breakdown voltage while maintaining a narrow n^- drift region width, as the forward voltage drop increases with an increase in n^- drift region width. The following equation calculates the n^- drift region width:

$$d_1 = \sqrt{\frac{2\varepsilon V_b}{qN_D}} + L_p \qquad (3.22)$$

where d_1: n^- drift region width
V_b: maximum blocking voltage
N_D: doping concentration
L_p: minority carrier diffusion length $=\sqrt{D_p \tau_p}$

Processing alternative for reverse blocking
Because the n region surfaces on the emitter side of the device, the uncontrolled field in this region produced by a reverse voltage, causes premature breakdown. To avoid this, the first processing step is to surround each IGBT die region on the wafer by a deep boron p-well which is selectively driven in from the emitter side. The collector side is mechanically ground to about 100µm, so as to expose to boron diffusion. The remaining processes are essentially as for the conventional NPT IGBT, which results in a structure as shown in figure 3.18.
The reverse bias scl is modified and silicon nitride passivation of the emitter surface and an n-channel field stop results in a controlled scl profile, as shown dashed in figure 3.18. Other than increased processing complexity (hence costs) minimal on-state voltage - switching speed compromise arises. Effectively, a device with the performance lagging by one technology generation is achieved.
Reverse blocking capability extension to the desirable PT IGBT structure is problematic since the n-buffer region is of a higher concentration than the n-substrate. Thus the formed pn junction will have a significantly lower avalanche breakdown voltage level, as predicted by equation 2.3.

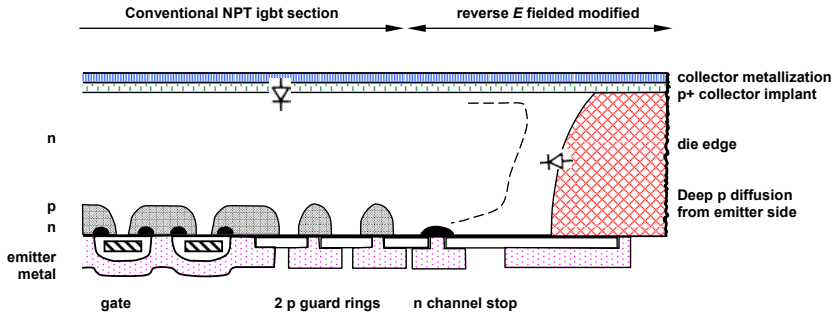


Figure 3.18. *Reverse voltage blocking NPT-IGBT structure.*

3.2.5 *Forward conduction characteristics*

Structurally, the IGBT can be viewed as a serial connection of the MOSFET and PiN diode. Alternatively, it is sometimes considered a wide base pnp transistor driven by the MOSFET in a Darlington configuration. The former view can be used to interpret the behaviour of the device, but the latter better describes the IGBT.
The width of the undepleted n- drift region does not change rapidly with the increase in the collector voltage due to the high concentration of the buffer layer, but maintains the same width as the n+ buffer layer for all collector voltages. This results in a constant value of the pnp transistor's current gain. Additionally, the n+ buffer layer reduces the injection efficiency of the p+ substrate / n+ buffer junction, J1. This reduces the current gain of the pnp transistor. Also, the collector output resistance can be increased with electron irradiation to shorten the minority carrier lifetime, which reduces the diffusion length. The IGBT saturated collector current expression involves the MOSFET current given by equation (3.7), giving:

$$I_d = \frac{1}{1-\alpha_{pnp}} \times \tfrac{1}{2}\mu \frac{W_c}{L_c} C_a (V_{gs} - V_{Th})^2 \qquad \text{(A)} \qquad (3.23)$$

Transconductance in the active region is obtained by differentiating the drain current with respect to V_{ge}. The IGBT's saturated collector current and transconductance are higher than those of the power MOSFETs of the same aspect ratio (W_c /L_c). This is because the pnp transistor's current gain α_{pnp} is significantly less than 1.

$$g_{fs} = \frac{1}{1-\alpha_{pnp}} \times \mu \frac{W_c}{L_c} C_a (V_{gs} - V_{Th}) \qquad \text{(mho)} \qquad (3.24)$$

3.2.6 *PT IGBT and NPT IGBT comparison*

Generally, faster switching speed is traded for higher on-state losses, and vice versa.

The N^+ buffer layer improves turn-off speed by reducing minority carrier injection and by increasing the recombination rate during the switching transition. In addition, latch-up characteristics are improved by reducing the PNP transistor current gain. The trade-off is that the on-state voltage increases. However, the thickness of the N^- drift region can be reduced with the same forward voltage blocking capability because the N^+ buffer layer improves the forward voltage blocking capability. As a result, the on-state voltage can be decreased. Hence, the PT-IGBT has superior trade-off characteristics as compared to the NPT-IGBT in switching speed and forward conduction voltage. Most IGBTs are PT-IGBTs. The IGBT static forward and reverse blocking capabilities for both types are similar because these characteristics are determined by the same N^- drift layer thickness and resistance. The reverse-blocking voltage of PT-IGBTs that contain the N^+ buffer layer between the P^+ substrate and N^- drift region is lowered to tens of volts due to the heavy doping regions bounding J1.

Table 3.1: PT versus NPT IGBTs

IGBT TYPE	PT IGBT	NPT IGBT
conduction loss (same switching speed)	Lower $v_{ce(sat)}$ Decreases slightly with temperature A slight positive temperature co-efficient at high current densities allows parallel connection.	Higher $v_{ce(sat)}$ Increases with temperature Suitable for parallel connection
switching speed (same on-state loss)	Faster switching due to high gain and reduced minority carrier lifetime	
short circuit rating		More rugged due to wider base and low pnp gain
turn-on switching loss	Largely unaffected by temperature	Largely unaffected by temperature
turn-off switching loss	Loss increases with temperature but start lower than NPT devices	Virtually constant with temperature

3.2.7 *The junction field effect transistor (JFET)*

The field effect for a FET may be created in two ways:

- A voltage signal controls charge indirectly using a capacitive effect as in the MOSFET, section 3.2.2.
- In a junction FET (JFET), the voltage dependant scl width of a junction is used to control the effective cross-sectional area of a conducting channel. If the zero bias voltage cuts off the channel then the JFET is *normally off*, otherwise if a reverse bias is needed to cut-off the channel, the JFET is termed *normally on*.

The electrical properties of SiC make the JFET a viable possibility as a power switch. Two normally on JFET structures are shown in figure 3.19, where it is seen how the scl layer decreases the channel width as the source to gate voltage reverse bias increases. In SiC, the channel has a positive temperature coefficient, $R_{on} \propto T^{2.6}$, hence parallel connection is viable. Natural current saturation with a positive temperature coefficient means lengthy short-circuit currents of over a millisecond can be sustained. Although the channel is bidirectional, in the biased off-state an integral fast, robust pn body diode is inherent as seen in figure 3.19b. The natural off-state properties of the MOSFET make the SiC variant more attractive than the JFET. The simpler JFET structure has revived interest in its SiC fabrication.

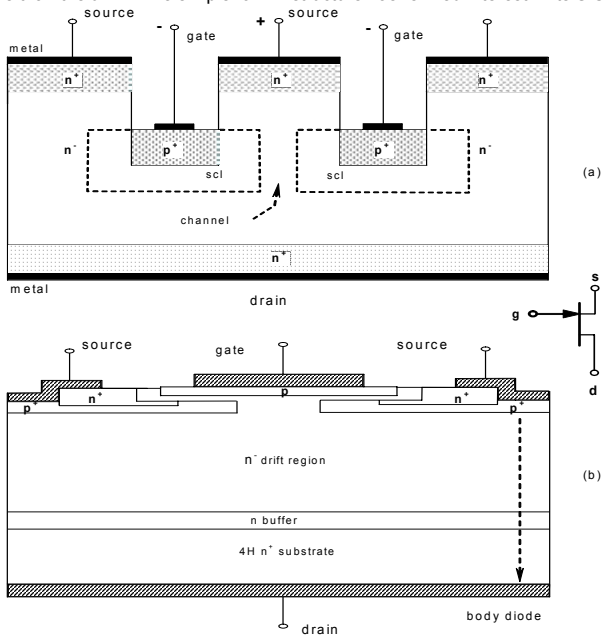


Figure 3.19. *Cross-section of the SiC vertical junction field effect transistor: (a) trench gate with channel shown and (b) variation incorporating a pn body diode.*

3.3 Thyristors

The name thyristor is a generic term for a bipolar semiconductor device which comprises four semiconductor layers and operates as a switch having a latched on-state and a stable off-state. Numerous members of the thyristor family exist. The simplest device structurally is the silicon-controlled rectifier (SCR) while the most complicated is the triac.

3.3.1 *The silicon-controlled rectifier (SCR)*

The basic SCR structure and doping profile in figure 3.20 depicts the SCR as three pn junctions J1, J2, and J3 in series. The contact electrode to the outer p-layer is called the ***anode*** and that to the outer n-layer is termed the ***cathode.*** With a ***gate*** contact to the inner p-region, the resultant three-terminal, 4 layer thyristor device is technically called the silicon-controlled rectifier (SCR).

A low concentration n-type silicon wafer is chosen as the starting material. A single diffusion process is then used to form simultaneously the p1 and p2 layers. Finally, an n-type layer, n1, is diffused selectively into one side of the wafer to form the cathode. The masked-out areas are used for the gate contact to the p1 region. To prevent premature breakdown at the surface edge, bevelling is used as in figure 3.1, to ensure that breakdown will occur uniformly in the bulk.

A number of observations can be made about the doping profile of the SCR which relate to its electrical characteristics.

The anode and cathode would both be expected to be good emitters of minority carriers into the n2 and p1 regions respectively because of their relative high concentrations with respect to their injected regions.

The n2 region is very wide, typically hundreds of micrometres, and low concentration, typically less than 10^{14}/cc. Even though the hole lifetime may be very long, 100µs, the base transport factor for hole minority carriers, b_{tn2} is low. The low-concentration provides high forward and reverse blocking capability and the associated reverse-biased scl's penetrate deeply into the n2 region. Gold lifetime killing or electron irradiation, most effective in the n2 region, is employed to improve the switching speed by increasing the number of carrier recombination centres.

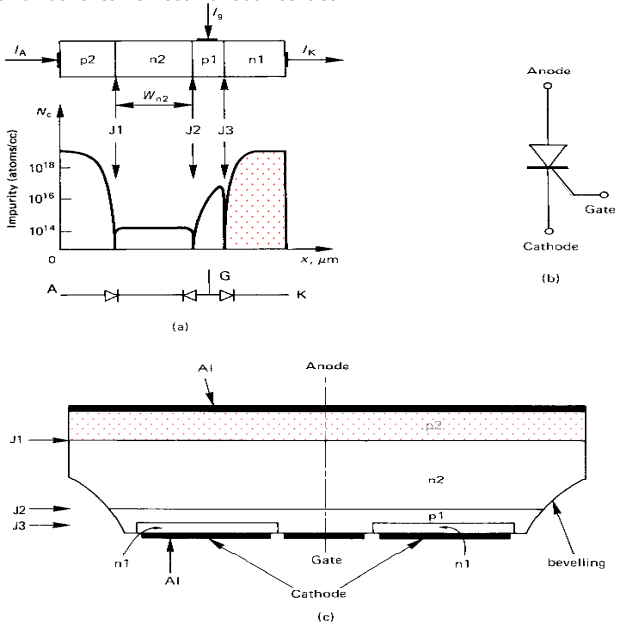


Figure 3.20. *The silicon-controlled rectifier, SCR:*
(a) net impurity density profile; (b) circuit symbol; and (c) cross-sectional view.

The two-transistor model of the SCR shown in figure 3.21 can be used to represent the p2-n2-p1-n1 structure and explain its electrical and thermal characteristics. Transistor T_1 is an npn BJT formed from regions n2-p1-n1 while T_2 is a pnp BJT formed from SCR regions p2-n2-p1.

The application of a positive voltage between anode and cathode does not result in conduction because the SCR central junction J2 is reverse-biased and *blocking*. Both equivalent circuit transistors have forward-biased emitter junctions and with reverse-biased collector junctions, both BJT's can be considered to be cut off.

3.3.1i - SCR turn-on

It is evident from figure 3.21c that the collector current of the npn transistor provides the base current for the pnp transistor. Also, the collector current of the pnp transistor along with any gate current I_G supplies the base drive for the npn transistor. Thus a *regenerative* current situation occurs when the loop gain exceeds unity.

The base current of the pnp transistor T_2 with dc current gain α_2 is

$$I_{b2} = (1 - \alpha_2) I_A - I_{co2}$$

which is supplied by the collector of the npn transistor. The current I_{co} is the collector junction reverse

bias leakage current. The collector current of the npn transistor T_1 with a dc current gain of α_1 is given by

$$I_{c1} = \alpha_1 I_K + I_{co1}$$

By equating I_{b2} and I_{c1}

$$(1 - \alpha_2) I_A - I_{co2} = \alpha_1 I_K + I_{co1}$$

Since $I_K = I_A + I_G$

$$I_A = \frac{\alpha_1 I_G + I_{co1} + I_{co2}}{1 - (\alpha_1 + \alpha_2)} = \frac{\alpha_1 I_G + I_{co1} + I_{co2}}{1 - G_1} \qquad \text{(A)} \qquad (3.25)$$

where $\alpha_1 + \alpha_2$ is called the *loop gain, G_1*.

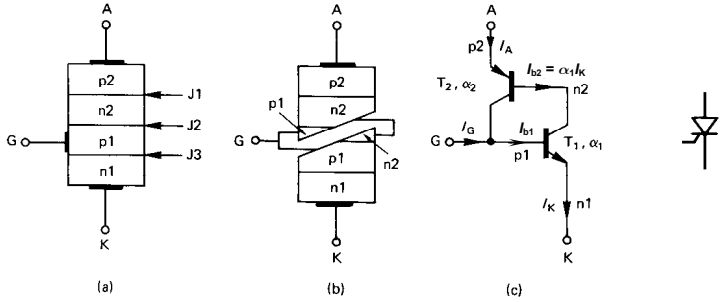


Figure 3.21. *Cross-section of the SCR showing its model derivation:*
(a) schematic of the SCR cross-section; (b) the division of the SCR into two transistors; and (c) the npn-pnp two-transistor model of the basic SCR.

At high voltages, to account for avalanche multiplication effects, the gains are replaced by $M\alpha$, where M is the avalanche multiplication coefficient in equation 3.15. Hence, G_1 becomes $M_1\alpha_1 + M_2\alpha_2$. By inspection of equation (3.25) it can be seen that a large anode current results when $G_1 \rightarrow 1$, whence the circuit regenerates with each transistor driving its counterpart into saturation. All junctions are forward-biased and the total device voltage is approximately that of a single pn junction, with the anode current limited by the external circuit. The n2-p1-n1 device acts like a saturated transistor and provides a remote contact to the n2 region. Therefore the device behaves essentially like a p-*i*-n diode (p2-*i*-n1), where the voltage drop across the *i*-region is inversely proportional to the recombination rate. Typical SCR static *I-V* characteristics are shown in figure 3.22.

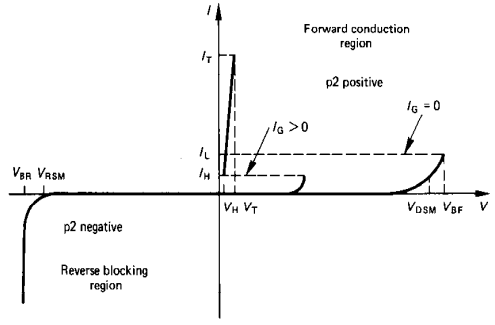


Figure 3.22. *The silicon-controlled rectifier static I-V characteristics.*

At low current levels, α_1 and α_2 are small because of carrier recombination effects, but increase rapidly as the current increases. The conventional gate turn-on mechanism is based on these current gain properties. External gate current starts the regeneration action and the subsequent increase in

anode current causes the gains to increase, thus ensuring a high loop gain, whence the gate current can be removed. The *I-V* characteristics in figure 3.22 show this property, where a minimum anode current I_{Latch} is necessary for the loop gain to increase sufficiently to enable the SCR to latch on by the regeneration mechanism.
The SCR can be brought into conduction by a number of mechanisms other than via the gate (other than the light triggered SCR used in high-voltage dc converters).

- If the anode-cathode voltage causes avalanche multiplication of the central junction, the increased current is sufficient to start the regenerative action. The forward anode-cathode breakover voltage V_{BF} is dependent on the central junction J2 avalanche voltage and the loop gain according to

$$V_{BF} = V_b (1 - \alpha_1 - \alpha_2)^{1/m} \quad \text{(V)} \qquad (3.26)$$

  where the avalanche breakdown voltage, at room temperature, for a typical SCR p^+n central junction J2 is given by equation (2.3)

$$V_b = 5.34 \times 10^{13} \times N_D^{-3/4} \quad \text{(V)} \qquad (3.27)$$

  where N_D is the concentration of the high resistivity n2 region when $10^{13} < N_D < 5\text{x}10^{14}$ /cc.
- Turn-on can also be induced by means of an anode-to-cathode *applied dv/dt* where the peak ramp voltage is less than V_{BF}. The increasing voltage is supported by the central blocking junction J2. The associated scl width increases and a charging or displacement current flows according to $i = d(Cv)/dt$. The charging current flows across both the anode and cathode junctions, causing hole and electron injection respectively. The same mechanism occurs at the cathode if gate current is applied; hence if the terminal *dvldt* is large enough, SCR turn-on occurs.
- The forward SCR leakage current, which is the reverse-biased pn junction J2 leakage current, doubles approximately with every 8K temperature rise. At elevated temperatures, the thermally generated leakage current (in conjunction with the gains increasing with temperature and current) can be sufficient to increase the SCR loop gain such that turn-on occurs.

3.3.1ii - SCR cathode shorts

All SCR turn-on mechanisms are highly temperature-dependent. A structural modification commonly used to reduce device temperature sensitivity and to increase *dv/dt* rating is the introduction of *cathode shorts.* A cross-sectional structure schematic and two-transistor equivalent of the cathode shorting technique are shown in figure 3.23. It will be seen that the cathode metallization overlaps the p1 region, which is the gate contact region. The technique is based on some of the anode forward-blocking current being shunted from the cathode junction via the cathode short. The cathode electron injection efficiency is effectively reduced, thereby decreasing α_1 which results in an increase in the forward voltage-blocking rating V_{BF} and *dv/dt* capability. The holding and latching currents are also increased.

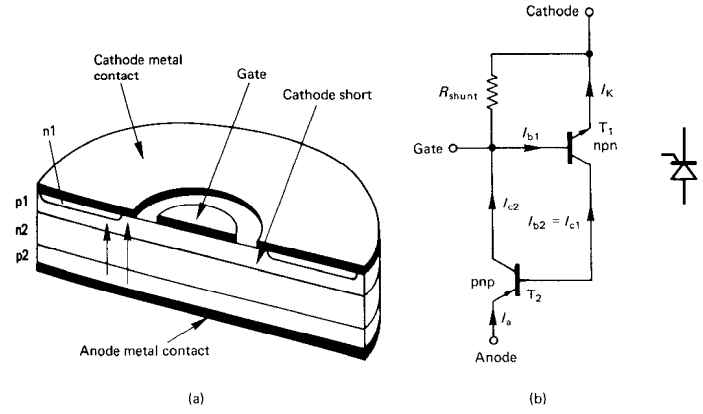


Figure 3.23. *Shorted cathode SCR:*
(a) SCR cross-section showing some anode current flowing through cathode shorts and (b) the SCR two-transistor equivalent circuit SCR with cathode shorts.

The cathode-anode, reverse breakdown voltage V_{BR} is shown in figure 3.22. The anode $p2^+n2$ junction J1 characterises SCR reverse blocking properties and V_{BR} is given by (equation (3.6))

$$V_{BR} = V_b (1 - \alpha_2)^{1/m}$$

If a very high resistivity n2 region, N_{Dn2}, is used (in conjunction with low temperature) and breakdown is due to punch-through to J2, then the terminal breakdown voltage will be approximated by (equation (2.2))

$$V_{PT} = 7.67 \times 10^{-16} N_{Dn2} W_{n2}^2$$

where W_{n2} is the width of the n2 region. This relationship is valid for both forward and reverse SCR voltage breakdown arising from punch-through.

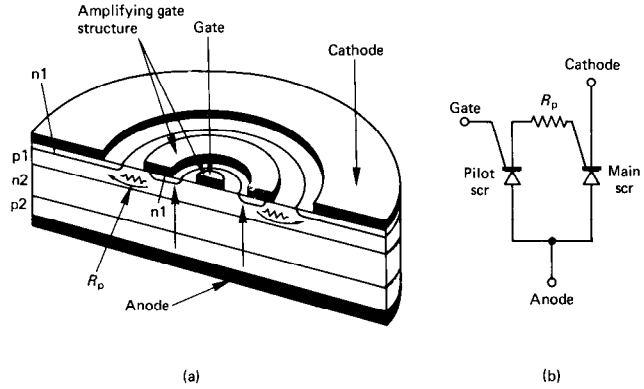


Figure 3.24. *The amplifying gate SCR:*
(a) cross-section of the structure and (b) two-SCR equivalent circuit.

3.3.1iii - SCR amplifying gate

At SCR turn-on, only a small peripheral region of the cathode along the gate region conducts initially. The conducting area spreads at about 50m/s, eventually encompassing the whole cathode area. If at turn-on a very large anode current is required, that is a high *initial di/dt,* a long gate-cathode perimeter is necessary in order to avoid excessively high, localised initial cathode current densities. The usual method employed to effectively enlarge the SCR initial turn-on area is to fabricate an integrated *amplifying gate,* as shown in figure 3.24. A small gate current is used to initiate the *pilot SCR,* which turns on very rapidly because of its small area. The cathode current of this pilot SCR provides a much larger gate current to the main SCR section than the original gate triggering current. Once the main device is fully on, the pilot device turns off if the gate current is removed.

An important property of the SCR is that once latched on, the gate condition is of little importance. The regenerative action holds the device on and SCR turn-off can only be achieved by reducing the anode current externally to a level below which the loop gain is significantly less than unity.

3.3.2 The asymmetrical silicon-controlled rectifier (ASCR)

The doping profiles and cross-sectional views comparing the asymmetrical SCR and conventional SCR are shown in figure 3.25. In each case the electric field ξ within the p1n2 junction reverse-bias scl is shown and because the n2 region is lightly doped, the scl extends deeply into it. The scl applied reverse-bias voltage is mathematically equal to the integral of the electric field, ξ (area under the curve).
If, in the conventional SCR, the scl edge reaches the $p2^+$ layer, then punch-through has occurred and the SCR turns on. To prevent such a condition and to allow for manufacturing tolerances, the $n2^-$ region is kept thick with the unfortunate consequence that on-state losses, which are proportional to n2 layer thickness, are high.
In the case of the ASCR, a much thinner $n2^-$ region is possible since a highly doped n layer adjacent to the $p2^+$ anode is utilised as an *electric field stopper.* The penalty paid for this layer construction is that in the reverse voltage blocking mode, the $n2p2^+$ junction avalanches at a low voltage of a few tens of volts. Thus the ASCR does not have any usable repetitive reverse-blocking ability, hence the name asymmetrical SCR. By sacrificing reverse-blocking ability, significant improvements in lower on-state voltage, higher forward-blocking voltage, and faster turn-off characteristics are attained.

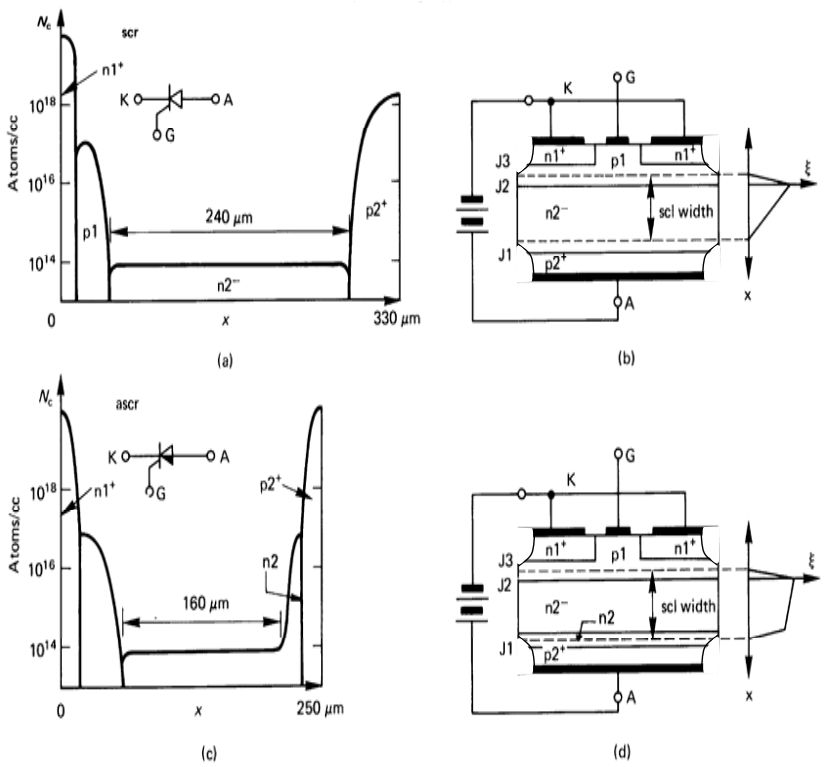


Figure 3.25. *Doping profile, cross-section, and the electric field of J2 in the forward biased off-state for: (a) and (b) the conventional SCR; (c) and (d) the asymmetrical SCR.*

3.3.3 *The reverse-conducting thyristor* (RCT)

The RCT is electrically equivalent to an SCR in anti-parallel with a diode, but both are integrated into the same wafer. The reason for integrating the SCR and diode is to minimise external interconnecting lead inductance. The circuit symbol, cross-sectional wafer view, and typical doping profile are shown in figure 3.26.

Since no reverse voltage will be applied to the RCT there is only the cathode-side deep p-diffused layer. This and the ASCR n-region type field stopper result in low forward voltage characteristics. As in the ASCR case, the highly n-type doped anode end of the wide n-region also allows higher forward voltages to be blocked. Both anode and cathode shorts can be employed to improve thermal and *dv/dt* properties. As shown in figure 3.26a, an amplifying gate can be used to improve initial *di/dt* capability.

The integral anti-parallel diode comprises an outer ring and is isolated from the central SCR section by a diffused guard ring, or a groove, or by irradiation lifetime control techniques. The guard ring is particularly important in that it must confine the carriers associated with the reverse-blocking diode to that region so that these carriers do not represent a forward displacement current in the SCR section. If the carriers were to spill over, the device *dv/dt* rating would be reduced - possibly resulting in false turn-on.

Gold or irradiation lifetime killing can be employed to reduce the turn-off time without significantly increasing the on-state voltage.

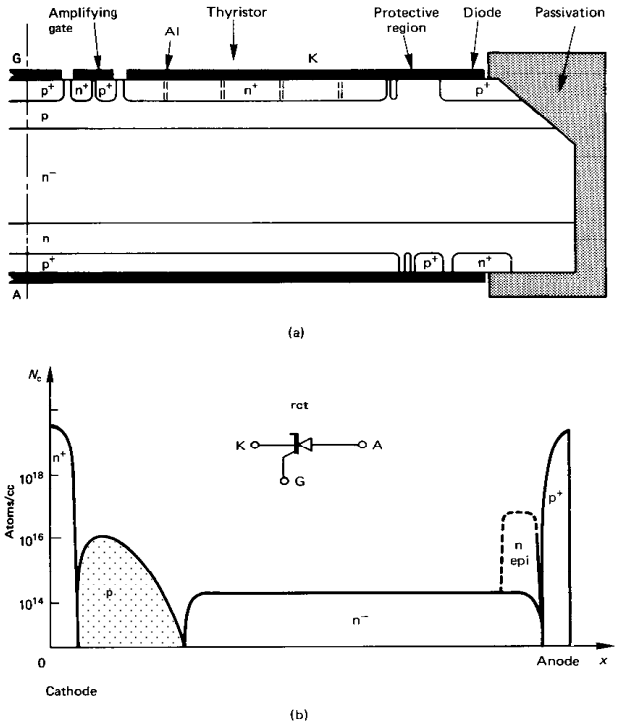


Figure 3.26. *Reverse conducting thyristor with an amplifying gate structure: (a) cross-section of the structure and (b) typical doping profile of the SCR section.*

3.3.4 *The bi-directional-conducting thyristor* (BCT)

Two anti-parallel connected SCRs can be integrated into one silicon wafer, as shown in figure 3.27. As a result of integrated symmetry, both devices have near identical electrical properties. The mechanical feature different to the triac, is that there are two gates – one on each side of the wafer. Also, unlike the triac, the two SCR sections are physically separated in the wafer to minimise carrier diffusion interaction. The equivalent circuit comprises two SCRs connected in anti-parallel. As such, one device turning off and supporting a negative voltage, represents a positive *dv/dt* impressed across the complementary device, tending to turn it on. Also, any charge carries which diffusion from the SCR previously on, exasperate the *dv/dt* stress on the off SCR.

The two central amplifying gate structures are as for the RCT, in figure 3.26a. A separation of a few minority carrier lateral diffusion lengths, along with an increased density of cathode shorts along the separating edge of each cathode and in the amplifying gate region close to the anode of the complementary SCR, enhance the physical separation. The amplifying gate fingers are angled away from the separation regions to minimise the shorting effect of the complementary SCR anode emitter shorting.

The on-state voltage of each SCR is fine tuned, match for on-state loss, using electron irradiation.

3.3.5 *The gate turn-off thyristor* (GTO)

The gate turn-off thyristor is an SCR that is turned on by forward-biasing the cathode junction and turned off by reverse-biasing the same junction, thereby preventing the cathode from injecting electrons into the p1 region. Other than its controlled turn-off properties, the GTO's characteristics are similar to the conventional SCR. The basic structure and circuit symbol are shown in figure 3.28.

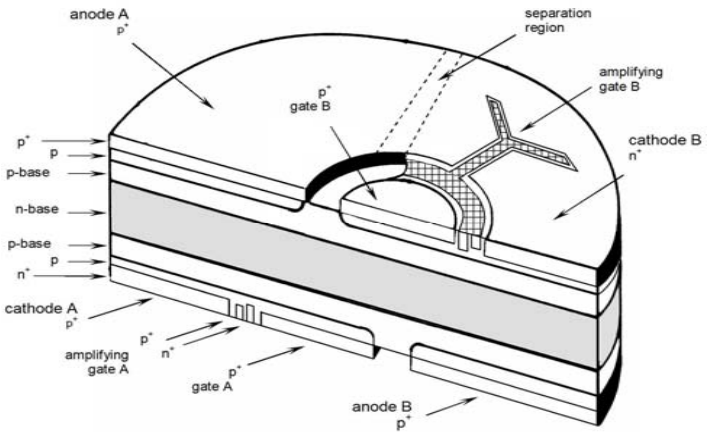


Figure 3.27. *Cross-section structure of the bidirectional conducting phase-control SCR with an amplifying gate structure.*

***3.3.5i* - GTO turn-off mechanism**

In the on-state, due to the high injection efficiency of junctions J1 and J3, the central p-base is flooded with electrons emitted from the n-cathode and the central n-base is flooded with holes emitted from the p-anode. If a reverse gate current flows from the cathode to the gate, with a driving voltage tending to reverse bias the gate-cathode junction – then p-base holes are extracted from the gate, suppressing the cathode junction from injecting electrons. Eventually the cathode junction is cut-off and the pnp transistor section, now without base current turns off, thereby turning off the GTO.

The turn-off mechanism can be analyzed by considering the two-transistor equivalent circuit model for the SCR shown in figure 3.21c. The reverse gate current I_{GQ} flows from the gate and is the reverse base current of the npn transistor T_1. The base current for transistor T_1 is given by $I_B = \alpha_2 I_A - I_{GQ}$, where $I_{GQ} = -I_G$. The reverse base current in terms of the gain of T_1 is $I_{RB} = (1-\alpha_1) I_K$. The GTO as a three terminal device must satisfy $I_A = I_K + I_{GQ}$ and to turn-off the GTO, $I_B < I_{RB}$. These conditions yield

$$(\alpha_1 + \alpha_2 - 1) I_A < \alpha_2 I_{GQ}$$

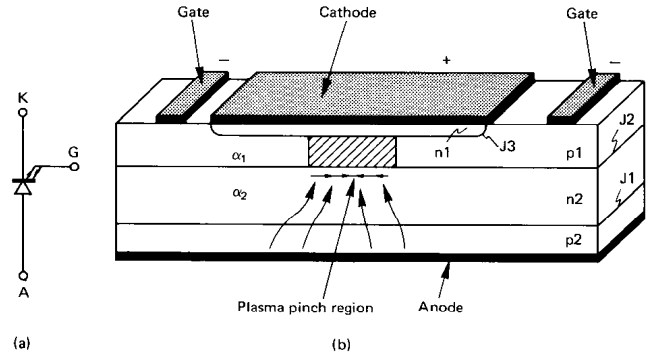


Figure 3.28. *The gate turn-off thyristor:*
(a) circuit symbol and (b) the basic structure along an interdigitated finger showing plasma focussing in the p1 region at the cathode junction at turn-off.

The turn-off gain of the GTO, β_Q, is defined as the ratio of anode current I_A to reverse gate current I_{GQ}, that is

$$\beta_Q \equiv I_T / I_{GQ} < \alpha_1 / (\alpha_1 + \alpha_2 - 1) \quad (3.28)$$

Thus for high turn-off gain it is important to make α_1 for the npn section as close to unity as possible, while α_2 of the pnp section should be small. A turn-off current gain of 5 is typical.

During the turn-off process, the conducting plasma is squeezed to the centre of the cathode finger, since the lateral p1 region resistance causes this region to be last in changing from forward to reverse bias. This region has the least reverse bias and for reliable GTO operation, the final area of the squeezed plasma must be large enough to prevent an excessive current density. Device failure would be imminent because of localised overheating.

The doping profile is characterised by a low p1 region sheet resistance and an inter-digitated cathode region to ensure even distribution of the reverse bias across the cathode junction at turn-off. Both turn-off and temperature properties are enhanced by using an anode shorting and defocusing technique as shown in figure 3.29a, but at the expense of any reverse-blocking capability and increased on-state voltage.

The shown two-level cathode and gate metallization used on large-area devices allow a flat metal plate for the cathode connection. As with the conventional SCR, a reverse conducting diode structure can be integrated, as shown in figure 3.29b.

3.3.6 The gate commutated thyristor (GCT)

GTO frequency limitations and the need for an external parallel connected capacitive turn-off snubber (to limit re-applied *dv/dt*), have motivated its enhancement, resulting in the gate commutated thyristor, GCT. As shown in figure 3.29c, a number of processing and structural variations to the basic GTO result in a more robust and versatile high power switch.

- *n-type buffer*
  An n-type buffer layer allows a thinner n-drift region. A 40% thinner silicon wafer, for the same blocking voltage, reduces switching losses and the on-state voltage. An integral reverse conducting diode is also possible, as with the conventional SCR and GTO.
- *transparent emitter*
  A thin lightly doped anode p-emitter is used instead of the normal GTO anode shorts. Some electrons pass through the layer as if the anode were shorted and recombine at the anode contact metal interface, without causing hole emission into the n-base. Effectively, a reduced emitter injection efficiency is achieved without anode shorts. Consequently, gate current triggering requirements are an order of magnitude lower than for the conventional GTO.
- *low inductance*
  A low inductance gate structure, contact, and wafer assembly (<2μH) allow the full anode current to be shunted from the gate in less than 1μs, before the anode voltage rises at turn-off.

3.3.6i - GCT turn-off

Unity turn-off gain, $\beta_Q = 1$, means the modified GTO turns off as a pnp transistor without base current, since the cathode junction is cut-off. Without npn BJT regenerative action, the pnp transistor rapidly traverses the linear region, thus eliminating the need for a capacitive turn-off snubber in the anode circuit. The high reverse gate current results in a short saturation delay time, enabling the accurate turn-off synchronisation necessary for devices to be series connected.

3.3.6ii - GCT turn-on

With high gate current, turn-on is initially by npn BJT action, not SCR regeneration. The pnp transistor section is inoperative since the carriers in the n-base are initially ineffective since they require a finite time to transit the wide n-base.

The SCR on-state regenerative mechanism is avoided at both turn-off and turn-on switching transitions thereby yielding a device more robust than the GTO thyristor.

As with the GTO, an inductive series turn-on snubber is still required to cope with the initial high *di/dt* current. The GCT switch is thermally limited, rather than frequency limited as with the conventional GTO.

Electron irradiation trades on-state voltage against switching performance.

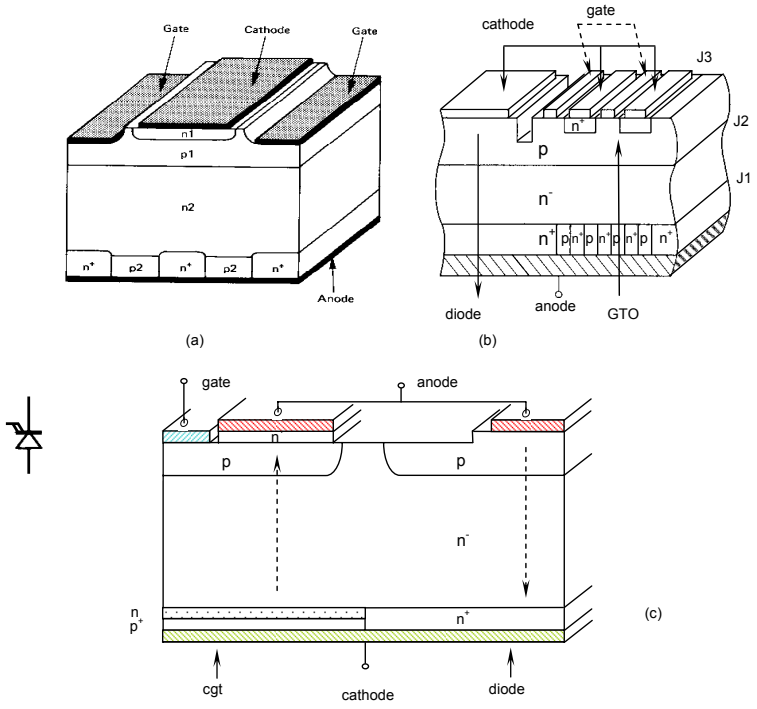


Figure 3.29: *GTO structure variations:*
(a) schematic structure of GTO finger showing the anode defocusing shorts, n^+;
(b) an integrated diode to form a reverse conducting GTO; and
(c) the reverse conducting gate commutated thyristor GCT.

3.3.7 *The light triggered thyristor (LTT)*

The light triggered thyristor is series connected in HVDC applications. Five inch wafers, after 16 major processing steps (as opposed to 10 for the conventional high voltage thyristor), offer 8kV ratings with on-state voltages of 2.3V at 3000A, with surge ratings of up to 63kA. Turn-off time is 350μs, and turn-on requires about 40mW of light power for 10μs, with a half microsecond rise time. The light causes the generation of hole-electron pairs and these free charges create a change in the electrical characteristics of the semiconductor region. Consequently a current flows across the exposed junction which is equivalent to gate current. Because of the low turn-on energy, multiple cascaded amplifying gates are laterally integrated to achieve modest initial current rises limited to 300A/μs. Reapplied voltages are limited to 3500V/μs.
A temperature dependant over voltage protection mechanism is also integrated into the wafer, the characteristics of which suffer from a wide production spread.

3.3.8 *The triac*

Pictorial representations of the triac are shown in figure 3.30. The triac is a thyristor device that can switch current in either direction by applying a low-power trigger pulse of either polarity between the gate and main terminal M1. The main terminal I-V characteristics, device symbol, and four trigger modes for the triac are shown in figure 3.31.

The triac comprises two SCR structures, p1-n1-p2-n2 and p2-n1-p1-n4 which utilise the n3 and p2 regions for turn-on. It should be noted that n2-p2, p1-n4, and p2-n3 are judiciously connected by terminal metallizations, but are laterally separated from their associated active parts.

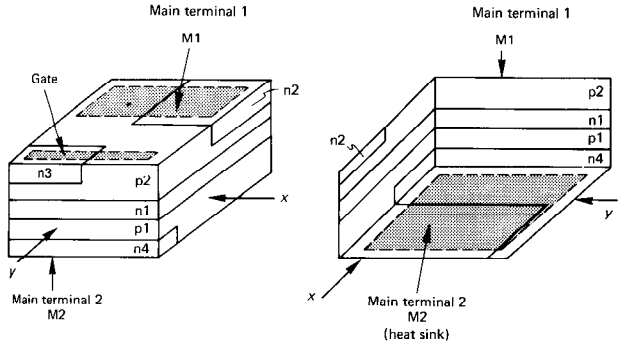


Figure 3.30. *Two views of the typical triac structure, showing main terminal M1 and M2, and a single gate.*

The four different trigger modes of the triac are illustrated in figure 3.32 and the turn-on mechanism for each mode is as follows.

(a) M2 positive, I_g positive (Mode I)
The main terminal M2 is positive with respect to M1 and gate current forward-biases the p2-n2 junction, J3. The active main SCR section is p1-n1-p2-n2. Turn-on is that for a conventional SCR, as shown in figure 3.32a.

(b) M2 positive, I_g negative (Mode II)
In figure 3.32b, M2 is positive with respect to M1 but negative gate voltage is applied. Junction J4 is now forward-biased and electrons are injected from n3 into p2. A lateral current flows in p2 towards the n3 gate and the auxiliary SCR section p1-n1-p2-n3 turns on as the gain of the n3-p2-n1 transistor section increases. Current flow in this auxiliary SCR results in a current flow across J3 into n2, hence piloting the SCR p1-n1-p2-n2 into conduction.

(c) M1 positive, I_g negative (Mode III)
Figure 3.32c shows the bias condition with M2 negative with respect to M1 and the gate negative with respect to M1 such that J4 is forward-biased and electrons are injected from n3 into the p2 region. The potential in n1 is lowered, causing holes to be injected from p2 into the n1 layer which provide base current for the p2-n1-p1 transistor section. This brings the p2-n1-p1-n4 SCR into conduction.

(d) M1 positive, I_g positive (Mode IV)
When M2 is negatively biased with respect to M1 and the gate is positively biased such that J3 is forward-biased, as in figure 3.32d, electrons are injected from n2 to p2 and diffused to n1. This increases the forward bias of J2 and eventually the SCR section p2-n1-p1-n4 comes into full conduction.

The various turn-on mechanisms are highly reliant on the judicious lateral separation of the various contacts and regions. The main advantage of the triac lies in the fact that two anti-parallel SCR's in the one silicon structure can be triggered into conduction from the one gate. Because of the need for extra structure layers, hence processing steps, some conventional SCR characteristics are sacrificed and poor device area utilisation results. Two anti-parallel SCR's therefore tend to be more robust than a triac but unlike the BCT device in section 3.3.4, only one gate drive circuit is needed for the triac.

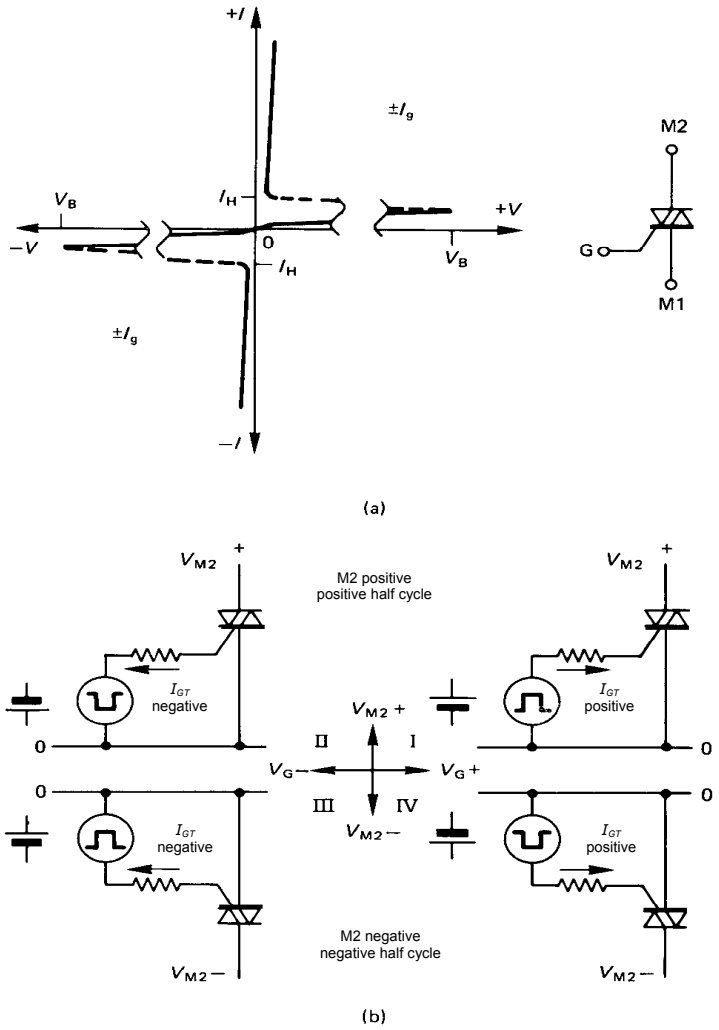


Figure 3.31. *The triac:*
(a) I-V characteristics and circuit symbol and (b) its four firing modes.

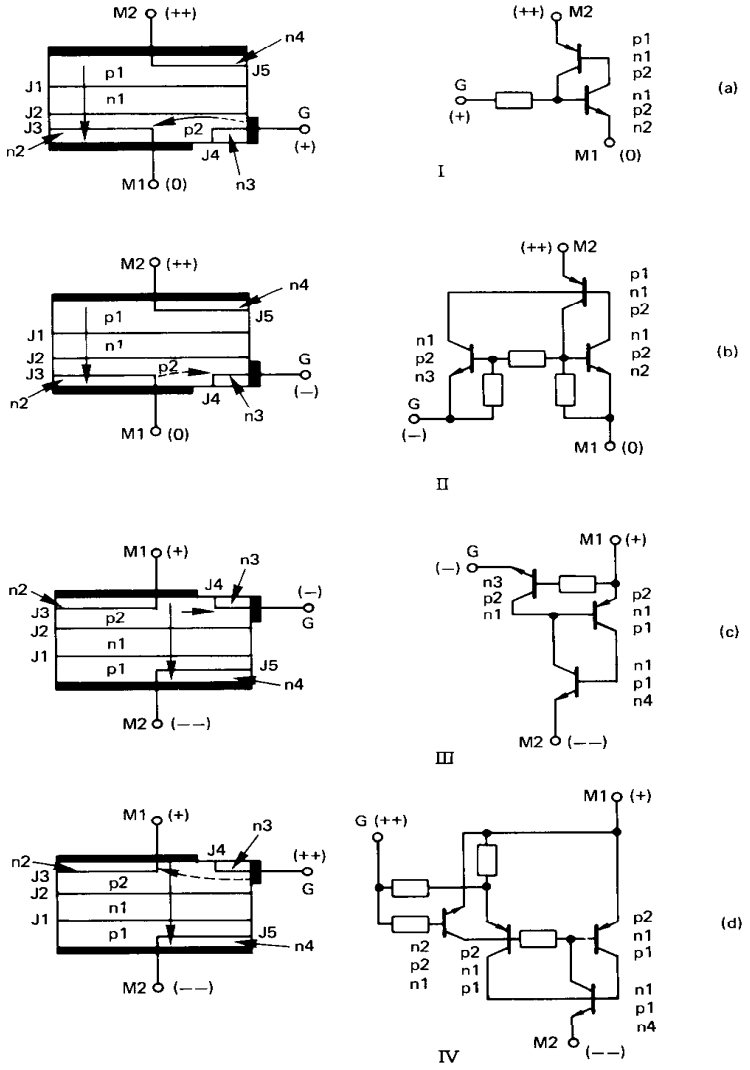


Figure 3.32. *Current flow for the four different turn-on triggering modes of the triac.*

3.4 Power packages and modules

Power thyristors are usually encapsulated as a floating disk in a ceramic package with Cu connection disks, as in figure 5.45. This offers the following features compared with high current IGBT modules.

- increased reliability with power cycling failure decreased by a factor of 10
- lower packaging connection and internal inductance
- explosion rated and stable short circuit failure mode
- suitable for liquid immersion
- lower thermal resistance due to double sided cooling

Advantageous features of high current IGBT modules are an electrically isolated base plate (based on $A\ell_3N_4$ or $A\ell_20_3$)) and low cost connections and heatsink mounting (see figure 5.5). The relative features of aluminium oxide and aluminium nitride substrates can be found in Chapter 5. No isolated pressure clamping arrangement is necessary with flat pack IGBT modules.

The emergence of SiC power switching devices has presented packing challenges. Package internal substrate and base plate assemblies currently prevent the high temperature capabilities of SiC from being exploited at junction temperatures above 300°C.

Details of high temperature die and substrate attachment can be found in Chapter 5.27.

Reading list

Baliga, B. J., *Modern Power Devices,* John Wiley-Interscience, 1987.

Blicher, A., *Thyristor Physics,* Springer, New York, 1976.

Ghandhi, S. K., *Semiconductor Power Devices,* John Wiley-Interscience, New York, 1977.

Grafham, D. R. *et al., SCR Manual,* General Electric Company, 6th edition, 1979.

Van Zeghbroeck, B.*, Principles of Semiconductor Devices,* *http://ece-www.colorado.edu/~bart/book*, 2004.

blank

4

Electrical Ratings and Characteristics of Power Semiconductor Switching Devices

Semiconductor device characteristics and ratings are primarily concerned with electrical and thermal properties. The thermal properties and cooling design aspects are similar for all power switching semiconductor devices. A common, unified thermal design approach is applicable since manufacturers use the concept of a semiconductor device being thermally represented by one *virtual junction*. This virtual junction is considered as the point source of all losses, which comprise on-state and off-state losses as well as switch-on and switch-off losses and any control input loss.

Not only are the power dissipation characteristics similar for all semiconductor devices, but many similarities exist in the area of maximum device ratings.

4.1 General maximum ratings of power switching semiconductor devices

The maximum allowable limits of current, applied voltage, and power dissipation are defined as the *maximum ratings* for that device. These absolute maximum ratings are important and the device must not experience a condition under which any one limit is exceeded if long life and reliability are to be attained. Generally, at worst, the device should experience only one near maximum rating at any instant.

Ratings are dependent on the materials used, the structure, the design, the mount, and the type of processes employed. The one property inherent in these physical features is temperature dependence and its interaction on electrical properties. Maximum ratings are therefore generally based on the variation of electrical characteristics that arises from the created variations. Because of this close correlation between properties, different ratings cannot be considered independently. Also, ratings are highly dependent on the device external circuit conditions.

This interdependence of device properties and the effects of external circuit conditions are no more evident than during *thermal runaway* - a condition to be avoided. Such a condition can occur in all devices that have bipolar junctions. For example, with the diode, thyristor, and the MOSFET's parasitic diode; reverse recovery current increases junction temperature. The reverse recovery charge increases with temperature, thus increasing junction power dissipation and further raising the junction temperature. This endless increasing of temperature and recovery charge results in thermal runaway and eventual device destruction. A similar thermal runaway condition occurs in the bipolar transistor and devices employing BJT mechanisms, like the thyristor and the IGBT. Here, collector current causes an increase in temperature which increases the conductivity of the bipolar transistor. More current then flows, further increasing the device temperature. If external circuit conditions allow, thermal runaway occurs, eventually resulting in irreversible device damage.

Figure 4.1 shows the electrical operating bounds of common semiconductor power switches, where the general trend is the higher the I-V ratings the slower the possible switching frequency, (because of increased losses associated with attaining higher sustaining voltages), hence increased junction temperature. High-frequency low-power switching applications are dominated by the MOSFET or possibly trench-gate IGBTs while high-power low-frequency switching applications are dominated by thyristor type devices or possibly IGBT modules. Rectifying or fast recovery diodes, as appropriate, are available with matched I-V ratings for all the switch device types in figure 4.1.

4.1.1 Voltage ratings

The absolute voltage limit is characterised by a sharp increase in leakage current when the device has at least one junction reverse-biased. The most commonly experienced voltage-limiting mechanism is that of avalanche multiplication, as considered in chapters 2.2.2 and 3.1.2. Since leakage current increases significantly with increased temperature, as shown in figure 2.2 and given by equation (2.7), the absolute repetitive voltage rating must be assigned such that thermal runaway does not occur. Most voltage ratings, because of historic reasons, are characterised at an impractical case temperature of 25°C. The leakage current at rated voltage and 25°C varies from hundreds of microamperes for low voltage devices (less than 600V) to milliamperes for high voltage devices (> 3.3kV).

4.1.2 Forward current ratings

The forward current ratings are usually specified after consideration of the following factors.

- Current at which the junction temperature does not exceed a rated value.
- Current at which internal leads and contacts are not evaporated.
- External connector current-handling capabilities.

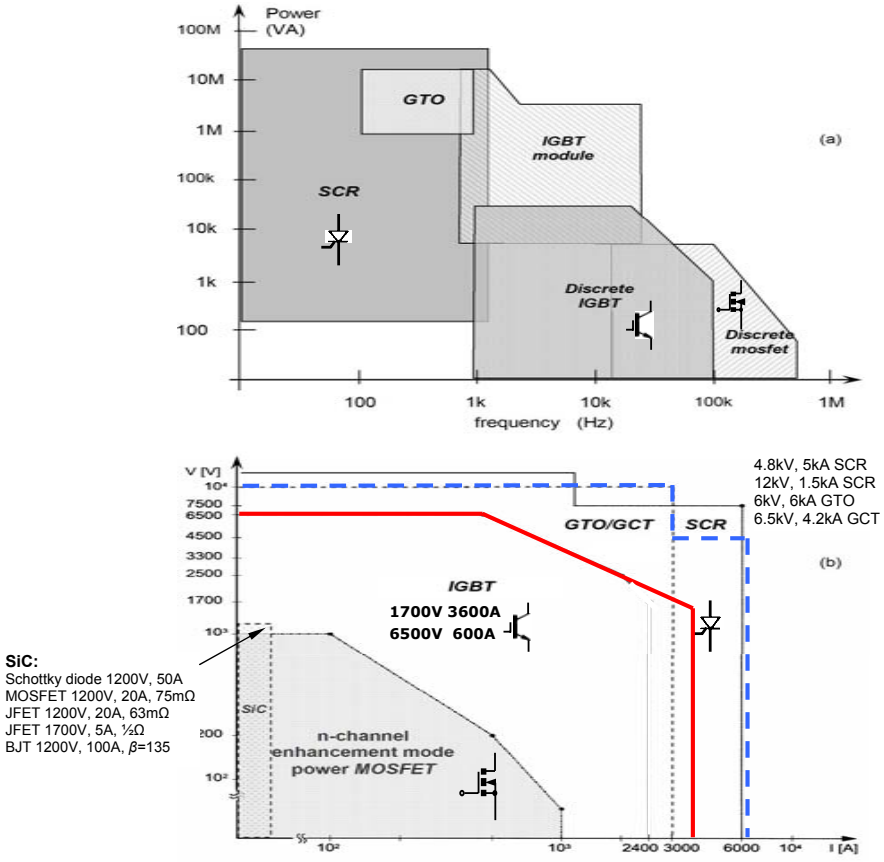


Figure 4.1. *Electrical rating bounds for power switching silicon devices, where (a) frequency related losses limit upper power through-put and (b) voltage is restricted by silicon limitations while current is bounded by packaging and die size constraints.*

4.1.3 Temperature ratings

The maximum allowable junction temperature $\hat{T}_j$, is dependent on the quality of the materials used and the type of junction, and is traded off against the reduced reliability that arises from deterioration and accelerated service life. The higher the junction temperature, the higher the rate of deterioration. The relationship between service life L_t in hours, and the junction temperature T_j (K) is approximated by

$$\log_{10} L_t \approx A + B/T_j \qquad (4.1)$$

where A and B are constants which are related to the device type.

4.1.4 Power ratings

Power dissipated in a semiconductor device is converted into thermal energy which produces a temperature rise. The major parameters limiting the maximum allowable power dissipation $\hat{P}_d$ are the maximum allowable junction temperature and the device case temperature T_c. These parameters are related to one another by the *thermal resistance* R_θ, according to

$$\hat{P}_d = \frac{\hat{T}_j - T_c}{R_{\theta j\text{-}c}} \quad \text{(W)} \qquad (4.2)$$

The virtual junction to case thermal resistance $R_{\theta\,j\text{-}c}$ is a physical value representing the ratio, junction temperature rise per unit power dissipation. Thermal resistance is a measure of the difficulty in removing heat from the junction to the case. Most maximum power values are specified at a case temperature of 25°C, and are derated linearly to zero as the case-operating temperature increases to $\hat{T}_j$, which is typically a maximum of 175°C for silicon power switching devices.

4.2 The fast-recovery diode

Static I-V diode characteristics were considered in chapter 2 and chapter 3.1. In low-frequency applications the only problem posed by a rectifier is heat dissipation, which can be readily calculated if the current waveform is known. On the other hand, calculation of losses in rectifiers for high-frequency application requires knowledge of device switching phenomena. The forward and reverse recovery characteristics are the most important fast-recovery bipolar pn diode electrical switching properties.

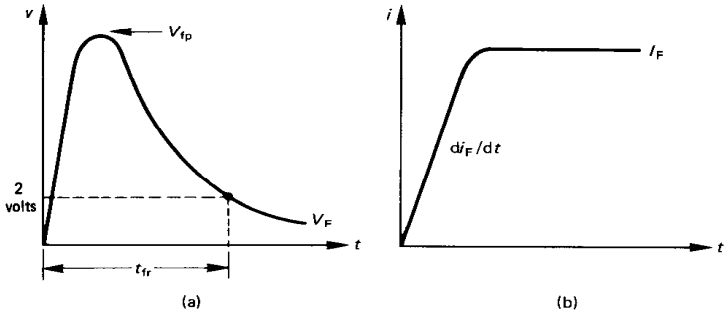


Figure 4.2. *Diode forward recovery measurement:*
(a) specification of forward recovery time, t_{fr} and peak forward voltage, V_{fp} and
(b) diode anode current test waveform.

4.2.1 Turn-on characteristics

During the forward turn-on period of a rectifier, an overshoot voltage is impressed in a forward bias direction across the diode as the forward current increases. The forward recovery characteristics of time t_{fr} and peak forward voltage V_{fp} are measured as shown in figure 4.2, with a specified increase in forward current di_F/dt, rising to a maximum forward current level I_F.

Two mechanisms predominate and contribute to the forward voltage overshoot phenomenon. The first mechanism is resistive, while the second is inductive.

Initially at turn-on, the device resistance is dominated by the ohmic resistance of the low-concentration n-region. As the concentration of the injected minority carriers increases, the n-region becomes conductively modulated and the associated ohmic voltage drop decreases significantly. These charging effects contribute to a minor initial capacitive component which serves to clamp V_{fp} initially to zero.

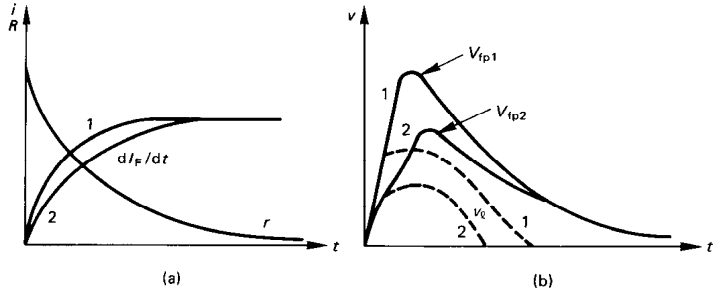


Figure 4.3. *Diode forward turn-on characteristics for two initial anode di/dt cases:*
(a) forward current and effective change in resistive component, r and
(b) anode voltage and voltage contribution v_t, as a result of die inductance.

The forward di_F/dt causes a voltage drop across the internal device inductance. This inductance comprises both the diode wafer internal inductance and the bonding and connection inductance. In bipolar power devices, the inductance of the wafer predominates. Any inductance contribution to the forward transient voltage ceases when the steady-state current level I_F is reached, as shown in figure 4.3. It will be seen that the peak forward transient voltage increases as di_F/dt increases. The resistive component predominates at low di_F/dt.

As with most minority carrier based power semiconductor characteristics, the turn-on phenomenon is significantly worsened by an increase in junction temperature. That is, both t_{fr} and V_{fp} are increased with increased temperature. Although a pre-reversed biased junction condition does not significantly prejudice the turn-on characteristics, if the junction is pre-forward biased slightly, the turn-on transitional phase can be significantly reduced. The Schottky diode, a majority carrier device, does not suffer from forward turn-on transient effects. Package inductances dominate at turn-on.

4.2.2 Turn-off characteristics

When a forward-conducting bipolar junction diode is abruptly reverse-biased, a short time elapses before the device actually regains its reverse blocking capabilities. Most importantly, before the diode does regain blocking ability, it may be considered as a short circuit in its normally blocking direction.

During forward conduction there is an excess of minority carriers in each diode region and the holes in the n-region and electrons in the p-region must be removed at turn-off. The attempted reverse bias results in a reverse current flow as shown in figure 4.4.

The total *recovery charge* Q_Σ is given by

$$Q_\Sigma = \tau I_F \qquad (4.3)$$

where I_F is the forward current before switching. In the usual p^+n diode, the excess minority holes in the n-region are more dominant. The lifetime τ is therefore the hole lifetime τ_h. Since carrier lifetime increases with temperature, recovery charge increases with temperature.

The recovery charge Q_Σ has two components, one due to internal excess charge natural recombination and the other, the *reverse recovery charge* Q_R, due to the reverse diode current shown in figure 4.4.

The excess charges reside in the neutral scl regions of the diode that border the junction. The excess charge concentration is largest at the scl edge on the n-side, reducing to zero well before the cathode contact.

Turn-off is initiated at t_f and the reverse recovery current i_{rr} commences. The rate of rise of this current is determined solely by the external inductance L of the switching circuit and the circuit applied reverse voltage E, according to

$$\frac{dI_F}{dt} = -\frac{E}{L} \quad \text{(A/s)} \qquad (4.4)$$

Until the time t_o, the diode carries forward current and is forward-biased. When the current reverses, the forward voltage drop decreases slightly but the device still remains positively biased. The external circuit inductance L supports the voltage E. The excess carrier concentrations now begin to reduce as holes leave via the junction, in providing the reverse current, i_{rr}. Holes are therefore extracted first and quickest at the edge of the scl.

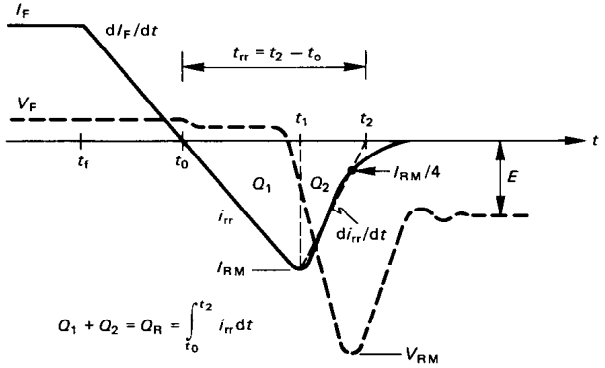


Figure 4.4. *Diode voltage and current during reverse recovery at turn-off.*

At time t_1, the hole concentration at the scl edge reaches zero, the charge Q_1 has been removed (plus natural recovery) and charge Q_2 remains. The reverse current now reduces rapidly since insufficient holes exist at the scl edge. The scl widens quickly, as it is charged. That is, the diode regains its ability to support reverse voltage and at the maximum reverse current I_{RM}, dI_F/dt reduces to zero.

Since $dI_F/dt = 0$, the voltage across the circuit inductance L drops rapidly to zero and E is applied in reverse bias across the diode. Between t_1 and t_2 the rate of change of reverse current di_{rr}/dt is high and, in conjunction with L, produces a reverse voltage overshoot to V_{RM}. After time t_2, di_{rr}/dt reduces to zero, the circuit inductance supports zero volts, and the diode blocks E.

In specifying the *reverse recovery time*, $t_{rr} = t_2 - t_o$, the time t_2 is defined by projecting i_{rr} through $¼I_{RM}$ as shown in figure 4.4. The reverse recovery time t_{rr} and peak reverse recovery current I_{RM}, at high magnitude dI_F/dt such that $Q_R \approx Q_\Sigma$, are approximated by

$$t_{rr} \approx 2.8 \times 10^{-6} V_b \sqrt{I_F / |dI_F/dt|} \quad \text{(s)}$$
$$\text{and} \quad I_{RM} \approx 2.8 \times 10^{-6} V_b \sqrt{I_F \times |dI_F/dt|} \quad \text{(A)} \qquad (4.5)$$

where the avalanche breakdown voltage for a step junction, V_b, is given by equation (2.3). The reverse recovery charge Q_R is therefore given by

$$Q_R = ½ I_{RM} t_{rr} = 3.92 \times 10^{-12} V_b^2 I_F \quad \text{(C)} \qquad (4.6)$$

that is, the reverse recovery charge is proportional to the forward current, as shown in figure 5.9a for dI_F/dt >100 A/µs.

Figure 4.5 illustrates *snap-off* and *soft recovery* diode properties (S_r) which are characterised by the recovery di_{rr}/dt magnitude. The higher the value of di_{rr}/dt, the higher is the induced diode overshoot V_{RM} and it is usual to produce soft recovery diodes so as to minimise voltage overshoot V_{RM}, resulting from inductive ringing.

Reverse recovery properties are characterised for a given temperature, forward current I_F, and dI_F/dt as shown in figure 5.9.

4.2.3 Schottky diode dynamic characteristics

Being a minority carrier device, the Schottky barrier diode, both in silicon and silicon carbide, is characterised by the absence of forward and reverse recovery, plus the absence of any temperature influence on switching.

Forward recovery traits tend to be due to package and external circuit inductance.

Reverse recovery is dominated by the barrier charging – a capacitive effect, which increases slightly with increased temperature, reverse di/dt, and I_F. The barrier charge requirements are significantly less than the highly temperature dependant minority carrier charge Q_Σ, associated with the bipolar pn junction diode. Unlike the pn diode, as Schottky junction charging occurs, the junction reverse bias voltage begins to increase immediately. Turn-off voltages are well controlled, less snappy, as the scl capacitor barrier junction acts like a capacitive turn-off snubber, as considered in chapter 8.3.

Whereas the transient performance is virtually independent of temperature, the static forward and reverse I-V characteristics are highly temperature dependant. In the case of silicon carbide, the reverse leakage current increases by 4% /K, the reverse breakdown voltage decreases by -4% /K while a 0.45% /K increase in on-state voltage means die can be readily parallel connected. In contrast, it will be noticed in figure 2.2 that reverse breakdown voltage and leakage current of a bipolar junction diode, both have a positive temperature co-efficient.

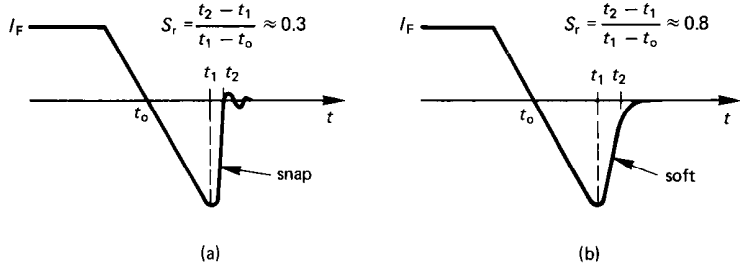


Figure 4.5. *Comparison of fast recovery diode di_{rr}/dt characteristics of: (a) short current tail, producing snap-off (low S_r) and (b) gradual current tail, producing soft recovery (high S_r).*

4.3 The bipolar, high-voltage, power switching npn junction transistor

The electrical properties of the high-voltage power switching npn transistor are related to and dominated by the wide low-concentration n⁻ collector region employed to obtain high-voltage characteristics in all semiconductor devices. Many of the limitations and constraints on the MOSFET, IGBT, and the different thyristors are due to their parasitic bjt structures, which introduce undesirable BJT characteristics and mechanisms. It is therefore essential to understand the electrical characteristics and properties of the BJT if the limitations of other switching semiconductor devices are to be appreciated.

4.3.1 Transistor ratings

4.3.1i – BJT collector voltage ratings

The breakdown voltage ratings of a transistor can be divided into those inherent to the actual transistor (V_{ceo}, V_{cbo}) and those that are highly dependent on the external base circuit conditions (V_{cer}, V_{ces}, V_{cev}).

Figure 4.6 shows the various voltage breakdown modes of the BJT, which are defined as follows.

V_{cbo} Collector to base voltage-current characteristics with the emitter open; that is, I_e = 0, where $V_{(BR)cbo}$ is the collector to base breakdown voltage with I_e = 0 and the collector current I_c specified as I_{cbo}.

V_{ceo} Collector to emitter characteristics with the base open circuit such that the base current I_b = 0, where $V_{(BR)ceo}$ is the collector to emitter breakdown voltage with I_b = 0 and I_c specified as I_{ceo}.

V_{ces} Collector to emitter characteristics with the base shorted to the emitter such that V_{be} = 0, where $V_{(BR)ces}$ is the collector to emitter breakdown voltage with I_c specified as I_{ces}.

V_{cer} Collector to emitter characteristics with resistance R between the base and the emitter such that R_{be} = R, where $V_{(BR)cer}$ is the collector to emitter breakdown voltage with I_c specified as I_{cer}.

V_{cev} Collector to emitter characteristics with reverse base to emitter bias V_{eb} = X, where $V_{(BR)cex}$ is the collector to emitter breakdown voltage with I_c specified as I_{cex}.

Each breakdown voltage level and its relative magnitude can be evaluated.

1 – BJT $V_{(BR)cbo}$ - maximum collector-base voltage with the emitter open circuit

The $V_{(BR)cbo}$ rating is just less than the voltage V_b, where the base to collector junction breaks down because of avalanche multiplication, as illustrated in figure 4.6. The common base avalanche breakdown voltage V_b is determined by the concentration of the collector n-region, N_c /cc, and as its resistivity increases, V_b increases according to (equation (2.3))

$$V_b = 5.34\times10^{13} \times N_c^{-3/4} \quad \text{(V)} \tag{4.7}$$

It can be assumed that $V_{(BR)cbo} \approx V_b$.

2 – BJT $V_{(BR)ceo}$ - maximum collector-emitter voltage with the base open circuit

Avalanche multiplication breakdown of a common emitter connected transistor occurs at a collector voltage V_a when the common emitter amplification factor β becomes infinite. The gain β, from equation 3.4 and accounting for avalanche multiplication, is defined by

$$\beta = \frac{\alpha_0 M}{1-\alpha_0 M} \tag{4.8}$$

where M is the avalanche multiplication factor, which is collector junction voltage V_{cb} dependent, according to (equation 3.17)

$$M = 1/\left(1-\frac{V_{cb}}{V_b}\right)^m \tag{4.9}$$

The factor m is empirically determined and is between 2 and 4 for the collector-base doping profile of the high-voltage silicon npn transistor. The common base current amplification factor α_0 is for a voltage level well below any avalanche.

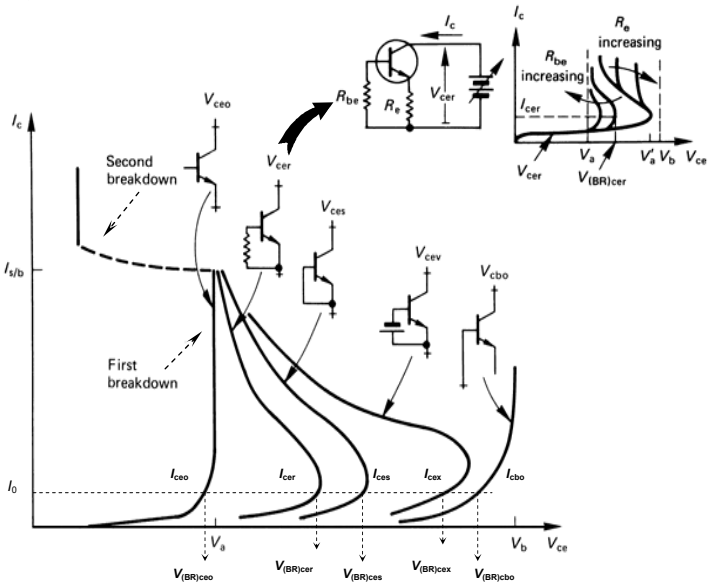


Figure 4.6. *Relative magnitudes of npn transistor collector voltage breakdown characteristics, showing first and second breakdown.*

At high V_{cb} voltages, near V_a, avalanche multiplication causes a high injection of hole carriers. Thus no base current is required and a $\beta \rightarrow \infty$ condition effectively occurs. With such conditions, equation (4.8) indicates that $\alpha_0 M \rightarrow 1$ which, upon substitution into equation (4.9), yields

$$V_a = V_b \sqrt[m]{1-\alpha_0} \approx V_{(BR)ceo} \quad \text{(V)} \tag{4.10}$$

V_a becomes the common emitter avalanche breakdown voltage $V_{(BR)ceo}$ which is commonly called the collector emitter sustaining voltage, $V_{ceo(sus)}$.

It can be shown that (see figure 4.6)

$$V_{(BR)cbo} > V_{(BR)cex} > V_{(BR)ces} > V_{(BR)cer} > V_{(BR)ceo} \tag{4.11}$$

With low-gain BJTs, V_a is almost V_b in value, but with high-gain devices V_b may be 2 to 3 times that of V_a. Notice in figure 4.6 that negative resistance characteristics occur after breakdown, as is the case with all the base circuit-dependent breakdown characteristics.

The inserted diagram in figure 4.6 shows how base-emitter resistance affects collector-emitter voltage breakdown. Importantly, the breakdown voltage increases as the base-emitter resistance decreases. This is because the injection efficiency of the emitter is reduced. This shorting feature is exploited extensively in alleviating parasitic problems in the MOSFET, IGBT, and thyristor devices, and is discussed in the respective device sections.

4.3.1ii – BJT safe operating area *(SOA)*

The safe operating area represents that electrical region where a transistor performs predictably and retains a high reliability, without causing device destruction or accelerated deterioration.

Deterioration or device destruction can occur when operating within the absolute maximum device ratings, as a result of second breakdown (s/b) or excessive thermal dissipation. Typical SOA characteristics are shown in figure 4.7. These collector characteristics are for a single pulse, of a given duration, such that the transistor operates in the linear region and at a case temperature of 25°C. The dc or continuous operation case has the most restrictive SOA curve, while a short single pulse of 1µs duration enables the full device I-V ratings to be exploited.

The SOA is basically bounded by the maximum collector $\hat{I}_c$ and the collector emitter breakdown voltage $V_{(BR)ceo}$. In figure 4.7 it will be seen that four distinct operating region limits exist, viz., A to D.

Ⓐ Maximum collector current which is related to allowed current density in the leads and contacts and the minimum gain of the transistor. The maximum lead current is given by $I = K_w d^{2/3}$ where the diameter d is in mm and K_w depends on the type and length of wire. For lengths greater than 1mm, K_w = 160 for both copper and silver.

Ⓑ Maximum thermal dissipation, which is related to the absolute maximum junction temperature $\hat{T}_j$, and the thermal resistance or impedance from the virtual junction to the case. In this thermally limited region, the collector power loss is constant and $I_c = P V_c^{-1}$. Thus the thermal limit gradient is -1, when plotted on logarithmic axes as in figure 4.7.

Ⓒ Limit of forward second breakdown (s/b). This breakdown occurs when the local current density is too high and a hot spot is created which causes thermal runaway. The physical causes of the high current concentration phenomenon are a fall in electrical potential or instability of lateral temperature distribution in the base area. These occur as a consequence of base-width concentration non-uniformity, a faulty junction or improper chip mounting. A typical s/b characteristic is shown in figure 4.6, and is characterised by a rapid drop in collector voltage to the low-impedance area after s/b. The s/b SOA limit can be modelled by $I_{s/b} = PV^{-n}$, where *n,* the gradient in figure 4.7, ranges from 1.5 to 4 depending on the fabrication processes and structures that have been employed. S/b, with a forward-biased base emitter is usually characterised by a short circuit at the emitter periphery, since this area is more forward-biased than central regions because of lateral base resistance effects. S/b, with a reverse-biased base-emitter junction, occurs in the central emitter region because of current focussing to that area as a result of the same lateral base resistance effects.

Ⓓ Maximum collector voltage under worst case conditions. In switching applications the $V_{(BR)ceo}$ limit can be exceeded provided suitable base conditions exist. At turn-off, when the collector current has fallen below I_{cex}, the collector supporting voltage can be increased from $V_{(BR)ceo}$ to $V_{(BR)cex}$ if the proper reverse bias base emitter junction conditions exist. The SOA together with this small extension area form the *reverse bias SOA*. Turn-on in switching applications can take place from a $V_{(BR)cex}$ condition, provided the collector current rise time is very short, usually much less than 1µs. As the rise time value decreases, the current that can be switched at turn-on increases. Under such conditions a significant portion of $\hat{I}_c$ can be switched from $V_{(BR)cex}$. The SOA together with this large switch-on extension area form the *forward bias SOA,* as shown in figure 4.7.

The SOA is usually characterised for a case temperature of 25°C. In practice much higher case temperatures are utilised and then the power and s/b SOA limits are modified with the aid of the derating curves of figure 4.8. At a given case temperature, above 25°C, power derating is greater than s/b derating. No derating is necessary for case temperatures below 25°C.

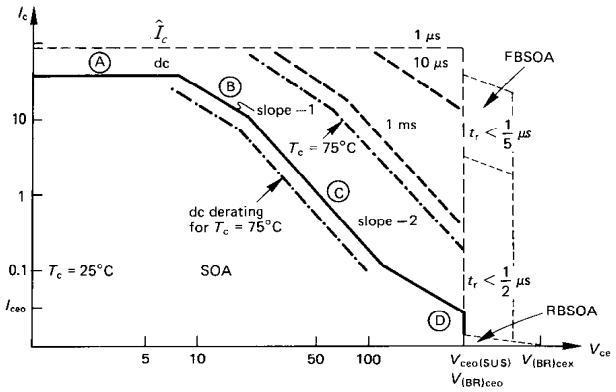


Figure 4.7. *Safe operating area (SOA) bounds of an npn high-voltage power switching transistor including forward and reverse bias SOA. Temperature derating for a case temperature of 75°C is shown.*

Figure 4.7 shows the derating, according to figure 4.8, of the dc and 1ms operating loci when the case temperature is increased from 25°C to 75°C. Figure 4.8 indicates that the power limit line (B) is derated to 71.5 per cent, while the s/b limit line (C) is reduced to 80 per cent. The slope of the 10μs single pulse limit line indicates that no s/b component exists, thus only power derating need be employed. This is because the pulse period of a few microseconds is short compared to the die thermal constant, whence the rate of local heating is too brief to disperse and cause second breakdown.

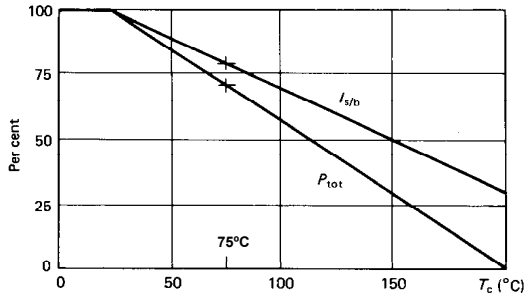


Figure 4.8. *Power and second breakdown derating versus case temperature.*

It is important to note that when a transistor is employed in a switching application, where the device is either cut-off or hard-on, the full SOA bounded by $\hat{I}_c$ and $V_{(BR)ceo}$ can usually be exploited. As indicated in figure 4.7, provided the collector switching times are of the order of a microsecond or less, no power or s/b derating need be factored. Design is based on total power losses, such that the maximum allowable junction temperature, $\hat{T}_j$ is not exceeded. For high reliability and long device lifetime only one electrical limit, either $\hat{I}_c$ or $V_{(BR)ceo}$, should be exploited in a given application.

4.3.2 Transistor switching characteristics

If a current pulse is supplied into the base of a common emitter connected transistor, as shown in figure 3.8, the resultant collector current waveform is as shown in figure 4.9. The collector voltage waveform is essentially collector load circuit dependent and therefore is not used to characterise transistor switching.

4.3.2i – BJT turn-on time: $t_{on} = t_d + t_{ri}$
Turn-on consists of a delay time t_d followed by a current rise time t_{ri}.
The delay time corresponds mainly to the charging of the base-emitter junction diffusion capacitance. The turn-on delay time can be significantly reduced by increasing the applied rate and magnitude of the forward base current I_{bf}.
The current rise time is related to the effective base zone width and, as the base charge increases because of the base current, the collector current increases.

4.3.2ii – BJT turn-off time: $t_{off} = t_s + t_{fi}$
In order to cut-off a transistor from the saturated state, all the accumulated charges must be neutralised or removed from the base and from the lightly doped n^- region of the collector. The turn-off process is started by removing the forward base current I_{bf}, and applying the reverse base current I_{br}. The excess minority carriers, namely holes, in the collector n^- region are progressively reduced in the process of providing the collector current. The excess minority carriers in the base are removed by the reverse base current. The reverse base current does not influence the collector n^- region recombination process. The period after the cessation of positive base current until the transistor enters the linear region is termed storage or *saturation time*, t_s. Generally, and undesirably, the larger the forward gain β_f, the greater the saturation time, t_s.

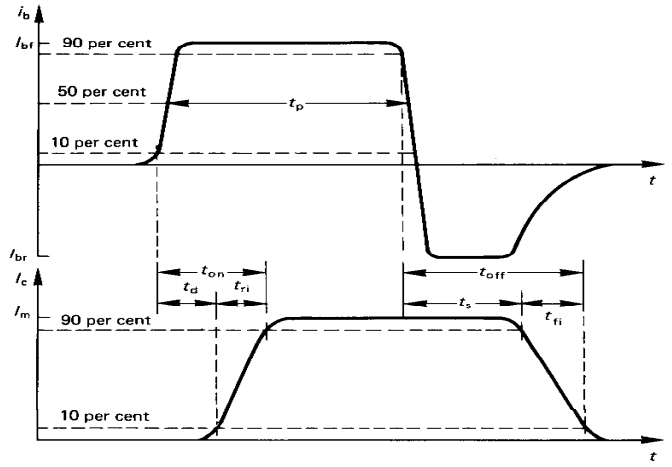


Figure 4.9. *Defining transistor base and collector current switching times for turn-on and turn-off.*

Optimal turn-off occurs when the emitter junction cuts off, as a result of I_{br}, just as the collector junction cuts off and enters the linear region. Thus the collector current fall time can be decreased by increasing the reverse base current immediately after the collector junction has cut off, which monimises the current fall time t_{fi}.

In switching applications, operation in the linear region is to be avoided, or at least traversed rapidly, because of the associated high device power losses. Although in the saturated state, with $I_{bf} >> I_c / \beta_f$, gives minimum forward gain and losses, this state is not conducive to a rapid turn-off transition to the cut-off region. In switching applications, in order to increase turn-off speed (decrease t_s and t_{fi}), the transistor may be held in the quasi-saturation region by reducing and controlling the forward base current magnitude such that the device is on the verge of saturation, $I_{bf} \approx I_c / \beta_f$, but is not in the linear region. The quasi-saturation on-state losses are slightly higher. In the quasi-saturated on-state the collection n^- region can be considered as extra series collector circuit resistance, which decreases as the neutral base region penetrates and reduces to zero when saturation occurs.

4.3.3 BJT phenomena

Although the BJT is virtually obsolete as a discrete power switching device for new circuit designs, it has been considered in some detail both in this chapter and chapter 3.2.1. This is because its operating electrical mechanisms explain the major limiting electrical operating factors of all controlled power switching devices.

- ***mosfet***: In chapter 3.2.2 the reverse conducting inherent body diode in the MOSFET is part of a parasitic npn transistor. This BJT structure can produce unwanted MOSFET *dv/dt* turn-on. Notice in figure 3.14a that the source metallization overlaps the p^+ well, there-in producing a base to emitter shunting resistor, as shown by R_{be} in figure 3.14b. The emitter shunts perform two essential functions, but inadvertently creates a non-optimal diode.
  - First, the shunt decreases the injection efficiency hence gain of the npn BJT, decreasing the likelihood of a drain *dv/dt* resulting in sufficient Miller capacitance current to turn-on the parasitic BJT, as considered in chapter 3.2.1.

- Second, by decreasing the BJT gain, the npn section voltage rating is increased from V_{ceo} to V_{cer} as considered in section 4.3.1.

- ***igbt***: In figure 3.16d the equivalent circuit of the IGBT has a parasitic pnp-npn thyristor structure. Once again, the emitter metallization (R_{be}) shunts the base to emitter of the npn BJT, helping to avoid latch-up of the SCR section, as modelled by the derivation of equation (3.18). Also the voltage rating of the npn section is increased from V_{ceo} to V_{cer}. Improved thermal stability also results. Judicious profiling of the transistor sections is essential.
- ***gto thyristor***: All the electrical operating mechanisms of the SCR are explainable in terms of BJT mechanisms, including turn-on, turn-off, and thermal stability. Emitter shorts (R_{be} shunts) are used extensively to decrease gain, increase thermal stability, and increase voltage ratings and are essential in providing separation in the bi-directional conducting thyristor, as considered in chapter 3.3.4. The GTO thyristor also uses emitter shorts in order to achieve a stable device at turn-off, as shown in figure 3.28.

An understanding of BJT electrical operating mechanisms is fundamental to the design and operation of semiconductor power switching devices, whether principally bipolar operating devices or unipolar devices which have bipolar parasitic structures.

4.4 The power MOSFET

The main electrical attributes offered by power MOSFETs are high switching speeds, no second breakdown (s/b), and high impedance on and off voltage control. MOSFETs, along with IGBTs, have replaced the bipolar junction transistor due to their superior switching performance and simpler gate control requirements.

4.4.1 MOSFET absolute maximum ratings

The basic enhancement mode power MOSFET structure and electrical circuit symbol are shown in figure 3.11. The SOA bounds shown in figure 4.10 is confined by four outer bounds.

Ⓐ The n^- epitaxial layer concentration and thickness is the key parameter in specifying the drain high-voltage ratings, such as V_{ds} and V_{dg}, which increase with temperature at approximately +0.1 per cent/K, as shown in figure 3.13.

Ⓑ One important rating feature of the power MOSFET is that it does not display the s/b that occurs with the bipolar transistor. Figure 4.10 shows the safe operating area for transistors, with the bipolar junction transistor s/b limitation area shaded. The physical explanation as to why MOSFETs do not suffer from s/b is based on the fact that carrier mobility in the channel decreases with increased temperature at -0.6 per cent/K. If localised heating occurs, the carrier mobility decreases in the region affected and, as a consequence, the localised current reduces. This negative feedback, self-protection mechanism forces currents to be uniformly distributed along the channel width and through the silicon die. This property is exploited when paralleling MOSFET devices. As a result of the enlarged SOA, the power MOSFET is generally a much more robust device than its bipolar counterpart. This region is thermally limited, as defined by $I = P/V^{-1}$ giving the -1 slope on the log-log axes in figure 4.10.

Ⓒ The drain current rating is also related to the epitaxial properties. Its resistance specifies the $I_d^2 R_{ds(on)}$ power loss, which is limited by the junction to case thermal resistance, $R_{\theta\,\text{j-c}}$. The continuous, usable drain current above 25°C is thus given by

$$I_d = \sqrt{\frac{\hat{T}_j - T_c}{R_{ds(on)}\ R_{\theta j\text{-}c}}} \qquad \text{(A)} \qquad (4.12)$$

Ⓓ When the MOSFET is on, with minimum drain voltage at maximum drain current, it operates in the resistive mode where the drain current is given by

$$I_d = \frac{1}{R_{ds(on)}} V_{ds} \qquad \text{(A)} \qquad (4.13)$$

The SOA region at high currents and low voltages is thus characterised by a line of slope 1, on logarithmic axes, as shown in figure 4.10.

The gate to source voltage V_{gs} controls the channel and the higher the value of V_{gs}, the higher the possible drain current. The gate to source is a silicon dioxide dielectric capacitor which has an absolute forward and reverse voltage that can be impressed before dielectric breakdown. Typical absolute maximum voltage levels vary from ±10V to ±40V, as the oxide layer thickness increases and capacitance advantageously decreases.

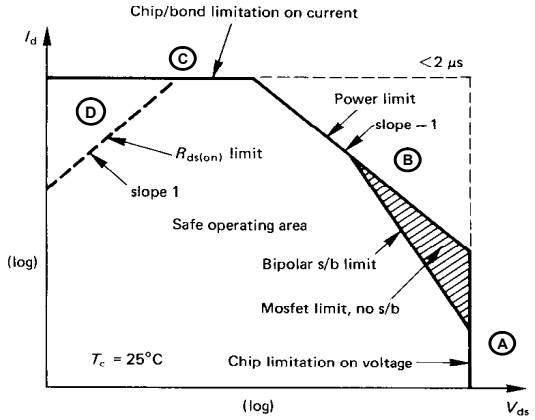


Figure 4.10. *The safe operating area of the power MOSFET, which does not suffer second breakdown.*

4.4.2 Dynamic characteristics

The important power MOSFET dynamic characteristics are inter-terminal voltage-dependent capacitance and drain current-switching times. The various MOSFET capacitances are dominant in specifying switching times.

4.4.2i – MOSFET device capacitances

Figure 4.11 shows an equivalent circuit for the power MOSFET, extracted from figure 3.14, which includes three inter-terminal, non-linear voltage-dependent capacitances C_{gd}, C_{gs}, and C_{ds}. The magnitudes are largely determined by the size of the chip and the cell topology used. Therefore higher current devices inherently have larger capacitances. Electrically, these capacitances are strongly dependent on the terminal drain-source voltage.

Manufacturers do not generally specify C_{gd}, C_{gs}, and C_{ds} directly but present input capacitance C_{iss}, common source output capacitance C_{oss}, and reverse transfer capacitance C_{rss}. These capacitances, as a function of drain to source voltage, are shown in figure 4.12a. The manufacturers' quoted capacitances and the device capacitances shown in figure 4.12b are related according to

$$C_{iss} = C_{gs} + C_{gd}; \qquad C_{ds} \text{ shorted} \qquad \text{(F)} \qquad (4.14)$$

$$C_{rss} = C_{gd} \qquad \text{(F)} \qquad (4.15)$$

$$C_{oss} = C_{ds} + \frac{C_{gs} \cdot C_{gd}}{C_{gs} + C_{gd}}; \qquad C_{gs} \text{ shorted} \qquad (4.16)$$

$$\approx C_{ds} + C_{gd} \qquad \text{(F)}$$

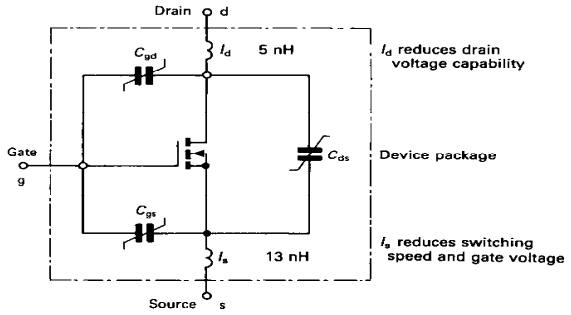


Figure 4.11. *MOSFET equivalent circuit including terminal voltage dependent capacitance and inductance for the TO247 package.*

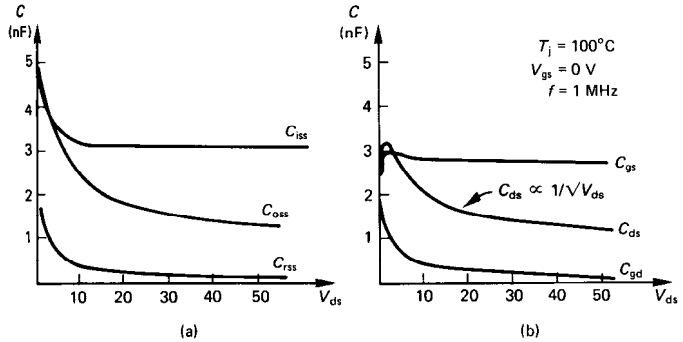


Figure 4.12. *MOSFET capacitance variation with drain-to-source voltage: (a) manufacturers' measurements and (b) inter-terminal capacitance values.*

The measurement frequency is usually 1 MHz and any terminals to be shorted are connected with large, high-frequency capacitance, so as to present a short circuit at the measurement frequency.

Device capacitances are predominant in specifying the drain current switching characteristics, particularly C_{gd} with its large capacitance variation at low drain voltage levels.

4.4.2ii – MOSFET switching characteristics

The simple single-ended MOSFET circuit with an inductive load L_L in figure 4.13, can illustrate how device capacitances influence switching. The MOSFET gate is driven from a voltage source whose output impedance is represented by R_g, which also includes any MOSFET gate series internal resistance. The dc input resistance of a power MOSFET is in excess of 10^{12} Ohms and when used as a switch, the power required to keep it on or off is negligible. However energy is required to change it from one state to the other another, as shown in figure 4.14. This figure shows the relationship between gate charge, gate voltage, and drain current for a typical MOSFET. The initial charge Q_{gs} is that required to charge the gate-source capacitance and Q_{gd} is that required to supply the drain-gate Miller capacitance. For a given gate charging current, switching speed is proportional to gate voltage. The gate charge required for switching, and hence switching speed, is not influenced significantly by the drain current magnitude, and not at all by the operating temperature. The switching speed is directly related to time delays in the structure because of the channel transit time of electrons. External to the device, the switching time is determined by the energy available from the drive circuit. A gate drive design example based on gate charge requirement is presented in chapter 7.1.2.

The switching transients can be predicted for an inductive load, when the load is the parallel inductor and diode, with no stray unclamped inductance, as shown in figure 4.13. It is assumed that a steady load current I_L flows. The various turn-on and turn-off periods shown in figure 4.15 are related to the sequential charging periods shown in figure 4.14. Any gate circuit inductance is neglected.

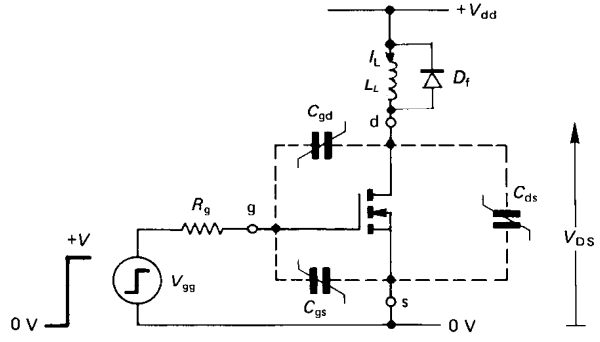


Figure 4.13. *MOSFET basic switching circuit used to demonstrate current switching characteristics.*

***1 – MOSFET** turn-on*

Period I - turn-on delay, $t_{d\,on}$

The gate voltage rises exponentially to the gate threshold voltage V_{TH} according to equation (4.17), that is

$$V_{gs}(t) = V_{gg}\,[1 - e^{-t/C_{in}R_g}] \qquad \text{(V)} \qquad (4.17)$$

where C_{in}, the gate input capacitance is approximated by $C_{gd} + C_{gs}$, or C_{iss}. The drain voltage remains unchanged, that is, it supports the supply voltage V_{dd} and virtually no MOSFET drain current flows. The turn-on delay time is given by

$$t_{d\,on} = C_{in}R_g\;\ell n\left(1 - \frac{V_{Th}}{V_{gg}}\right)^{-1} \qquad \text{(s)} \qquad (4.18)$$

Equations (4.17) and (4.18) can be modified to account for a negative initial gate voltage (as presented in Appendix 4.1), a condition which increases the turn-on delay time, but increases input noise immunity.

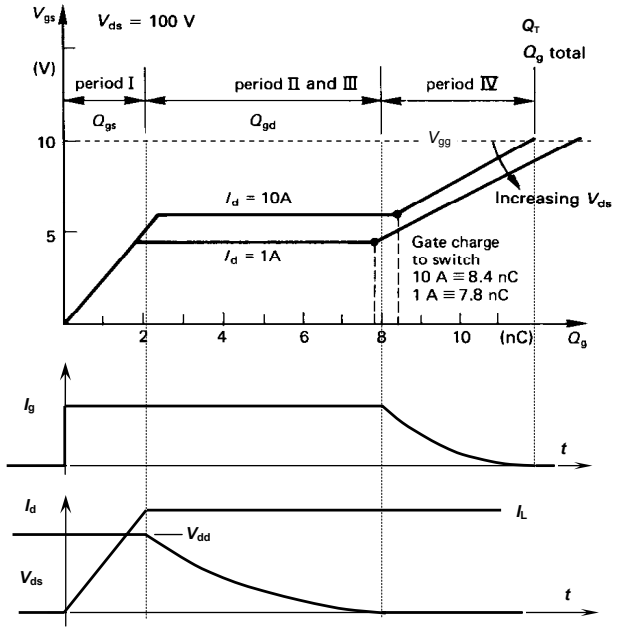


Figure 4.14. *Typical relationships between gate charge, voltage, and current and magnitude of drain current and voltage being switched.*

Period II - current rise, t_{ri}

Drain current commences to flow in proportion to the gate voltage as indicated by the transconductance characteristics in figure 3.12a. The gate voltage continues to rise according to equation (4.17). The drain voltage is clamped to the rail voltage V_{dd} and the drain current rises exponentially to the load current level I_L, according to

$$I_d(t) = g_{fs}\left(V_{gg} - V_{Th}\right)[1 - e^{-t/R_gC_{in}}] \qquad \text{(A)} \qquad (4.19)$$

The current rise time t_{ri} can be found by equating $I_d = I_L$ in equation (4.19).

$$t_{ri} = R_gC_{in}\ell n\left(\frac{g_{fs}\left(V_{gg} - V_{Th}\right)}{g_{fs}\left(V_{gg} - V_{Th}\right) - I_L}\right) \qquad (4.20)$$

Period III - voltage fall, t_{fv}

When the drain current reaches the load current level, the drain voltage will fall from V_{dd} to the low on-state voltage. This decreasing drain voltage produces a feedback current via C_{gd} to the gate, which must

be provided by the gate drive. This feedback mechanism is called the *Miller effect* and the effective gate input capacitance increases to $C_{in} = C_{iss} + (1 - A_v)C_{gd}$ where $A_v = \Delta V_{ds} / \Delta V_{gs}$. For a constant load current, from figure 3.12a, the gate voltage remains constant at

$$V_{gs} = V_{Th} + I_L / g_{fs} \quad \text{(V)} \quad (4.21)$$

as shown in figure 4.15b.
Since the gate voltage is constant, the Miller capacitance C_{gd} is charged by the constant gate current

$$I_g = \frac{V_{gg} - V_{gs}}{R_g} = \frac{V_{gg} - (V_{Th} + I_L / g_{fs})}{R_g} \quad \text{(A)} \quad (4.22)$$

and the rate of change of drain voltage will be given by

$$\frac{dV_{gd}}{dt} = \frac{dV_{ds}}{dt} = \frac{I_g}{C_{gd}} \quad \text{(V/s)} \quad (4.23)$$

that is

$$V_{ds}(t) = V_{dd} - \frac{I_g}{C_{gd}} t \quad \text{(V)} \quad (4.24)$$

The drain voltage decreases linearly in time and the voltage fall time is decreased by increasing the gate current. Assuming a low on-state voltage, the voltage fall time t_{fv} is given by

$$t_{fv} = V_{dd} C_{gd} / I_g = V_{dd} C_{gd} R_g / (V_{gg} - V_{Th}) \quad \text{(s)} \quad (4.25)$$

Period IV
Once the drain voltage reaches the low on-state voltage, the MOSFET is fully on and the gate voltage increases exponentially towards V_{gg}.

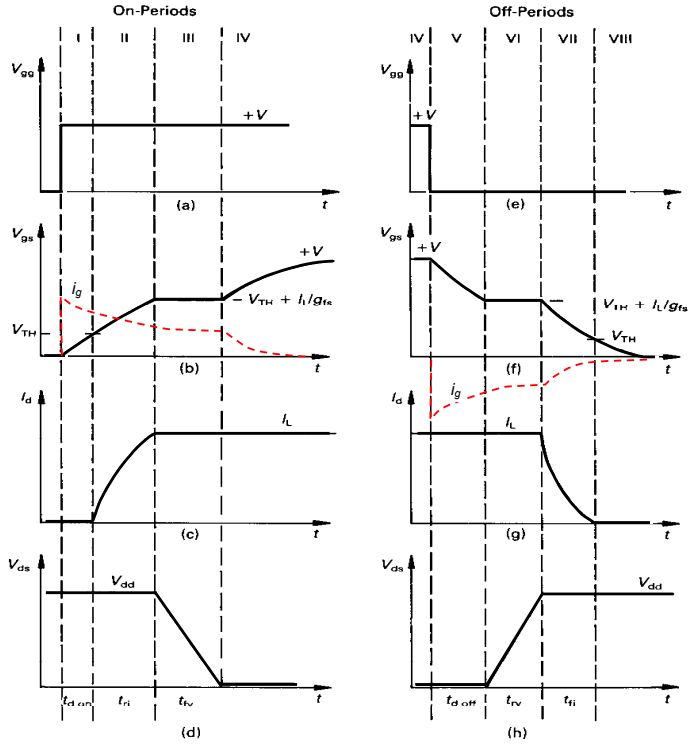


Figure 4.15. *Distinct switching periods of the MOSFET with an inductive load at: (a) (b) (c) (d) comprising turn-on; (e) (f) (g) (h) forming turn-off.*

2 – MOSFET turn-off

Period V - turn-off delay, $t_{d\,off}$
The MOSFET is fully on, conducting the load current I_L, and the gate is charged to V_{gg}. The gate voltage falls exponentially from V_{gg} to $V_{Th} + I_L / g_{fs}$ according to

$$V_{gs}(t) = V_{gg} e^{-t/R_g C_{in}} \quad \text{(V)} \quad (4.26)$$

in a time given by

$$t_{d\,off} = R_g C_{in} \ell n \frac{V_{gg}}{V_{Th} + I_L / g_{fs}} \quad \text{(s)} \quad (4.27)$$

This delay time can be decreased if a negative off-state gate bias is used. The drain conditions are unchanged.

Period VI - voltage rise, t_{rv}
The drain voltage rises while the drain current is fixed to the load current level, I_L. Accordingly the gate voltage remains constant and the gate current is given by

$$I_g = \frac{V_{Th} + I_L / g_{fs}}{R_g} \quad \text{(A)} \quad (4.28)$$

This current discharges the Miller capacitance according to

$$\frac{dV_{ds}}{dt} = \frac{dV_{dg}}{dt} = \frac{I_g}{C_{gd}} \quad \text{(V/s)} \quad (4.29)$$

Thus

$$V_{ds}(t) = \frac{I_g}{C_{gd}} t \quad \text{(V)} \quad (4.30)$$

where the low on-state voltage has been neglected.
The voltage rise time t_{rv} is given by

$$t_{rv} = \frac{C_{gd} V_{dd}}{I_g} \quad \text{(s)} \quad (4.31)$$

and is decreased by increasing the gate reverse current magnitude. The drain voltage rises linearly to the dc supply V_{dd}.

Period VII - current fall, t_{fi}
When the drain voltage reaches the supply rail, the load current in the MOSFET begins to decrease, with load current being diverted to the diode D_f.
The gate voltage decreases exponentially according to

$$V_{gs}(t) = (V_{Th} + I_L / g_{fs}) e^{-t/R_g C_{in}} \quad \text{(V)} \quad (4.32)$$

and is mirrored by the drain current

$$I_d(t) = (I_L + g_{fs} V_{Th}) e^{-t/R_g C_{in}} - g_{fs} V_{Th} \quad \text{(A)} \quad (4.33)$$

The current fall time t_{fi} is given by $I_d = 0$ in equation (4.33) or when the gate-source voltage reaches the threshold voltage, that is, from equation (4.32)

$$t_{fi} = R_g C_{in} \ell n \left(1 + \frac{I_L}{g_{fs} V_{Th}}\right) \quad \text{(s)} \quad (4.34)$$

Period VIII – off-state
The MOSFET drain is cut-off and the gate voltage decays exponentially to zero volts according to

$$V_{gs}(t) = V_{Th} e^{-t/R_g C_{in}} \quad \text{(V)} \quad (4.35)$$

Based on the total gate charge Q_T delivered by the gate source V_{gg}, shown in figure 4.14, the power dissipated in the MOSFET internal gate resistance, hence contributing to device losses, is given by

$$P_G(R_{int}) = V_{gg} Q_T f_s \frac{R_{g\,int}}{R_{g\,int} + R_{g\,ext}} \quad \text{(W)} \quad (4.36)$$

Example 4.1: ***MOSFET drain characteristics***

A power mosfet with
C_{gs}=1nF, C_{gd}=200pF, g_{fs}=4 and the threshold voltage is 3V,
is used to switch a 200V dc 20A load. If the gate is sourced from a 15V voltage source via a 10Ω gate resistance, what is the maximum rate of rise of drain current and voltage?

Solution

From equation (4.21), during turn-on:

$$i_D = g_{fs}\left(V_{gs} - V_{Th}\right)$$

whence

$$\frac{di_D}{dt} = g_{fs}\frac{dV_{gs}}{dt}$$

From equation 4.23

$$\left(C_{gs} + C_{gd}\right)\frac{dV_{gs}}{dt} = I_g = \frac{V_{gg} - V_{gs}}{R_g}$$

Combining these equations

$$\frac{di_D}{dt} = g_{fs}\frac{dV_{gs}}{dt} = g_{fs}\frac{I_g}{\left(C_{gs} + C_{gd}\right)} = g_{fs}\frac{V_{gg} - V_{gs}}{\left(C_{gs} + C_{gd}\right)R_g}$$

The maximum drain di/dt occurs at the gate threshold voltage, that is

$$\left|\frac{di_D}{dt}\right|_{\max} = \frac{g_{fs}}{\left(C_{gs} + C_{gd}\right)}I_g^{\max} = \frac{g_{fs}}{\left(C_{gs} + C_{gd}\right)R_g}\left(V_{gg} - V_{gs}^{\min}\right)$$

$$= \frac{g_{fs}}{\left(C_{gs} + C_{gd}\right)R_g}\left(V_{gg} - V_{Th}\right)$$

$$= \frac{4}{(1\text{nF} + 200\text{pF}) \times 10\Omega}(15\text{V} - 3\text{V}) = 4\text{kA/}\mu\text{s}$$

From equation (4.23)

$$\frac{dV_{gd}}{dt} = \frac{dV_{ds}}{dt} = \frac{I_g}{C_{gd}} = \frac{V_{gg} - V_{gs}}{C_{gd}R_g} \qquad \text{(V/s)}$$

The maximum drain dv/dt occurs at the gate threshold voltage, that is

$$\left|\frac{dV_{ds}}{dt}\right|_{\max} = \frac{I_g^{\max}}{C_{gd}} = \frac{V_{gg} - V_{gs}^{\min}}{C_{gd}R_g}$$

$$= \frac{V_{gg} - V_{Th}}{C_{gd}R_g}$$

$$= \frac{15\text{V} - 3\text{V}}{200\text{pF} \times 10\Omega} = 6\text{kV/}\mu\text{s}$$

♣

4.5 The insulated gate bipolar transistor

4.5.1 IGBT switching

The IGBT gate charge characteristics for switching and the switching waveforms are similar to those of the MOSFET, shown in figures 4.14 and 4.15 respectively, whilst the *I-V* on and off state characteristics are similar to the BJT. The collector switching characteristics depend on the injection efficiency of the collector p^+ emitting junction. The higher the injection efficiency, the higher the pnp transistor section gain and the lower the on-state voltage. The poorer the injection efficiency, the more the characteristics resemble a MOSFET.
The turn-on waveforms and mechanisms are essentially those for the MOSFET shown in figure 4.15. Figure 4.16 shows IGBT turn-off which has components due to MOSFET and BJT action. As with the MOSFET, distinct turn-off stages exist when switching an inductive load.

Period V - turn-off delay, $t_{d\,off}$
The gate voltage falls to a level determined by the gate threshold, V_{TH}, the forward transconductance, g_{fs} and the MOSFET section current level.

Period VI - voltage rise, t_{rv}
As the collector voltage rises the collector current remains constant, hence the gate voltage remains constant while charging the Miller capacitance. For a high gain pnp section the voltage rise time is virtually independent of gate resistance, while for an IGBT closely resembling a MOSFET the voltage rise is gate current magnitude dependent.

Period VII - current fall, t_{fi}
The current falls in two stages, the first, phase VII, due to MOSFET action, as are the previous two phases (periods V and VI). As with the conventional MOSFET the current falls rapidly as the MOSFET section current, shown in figure 4.16b reduces to zero.

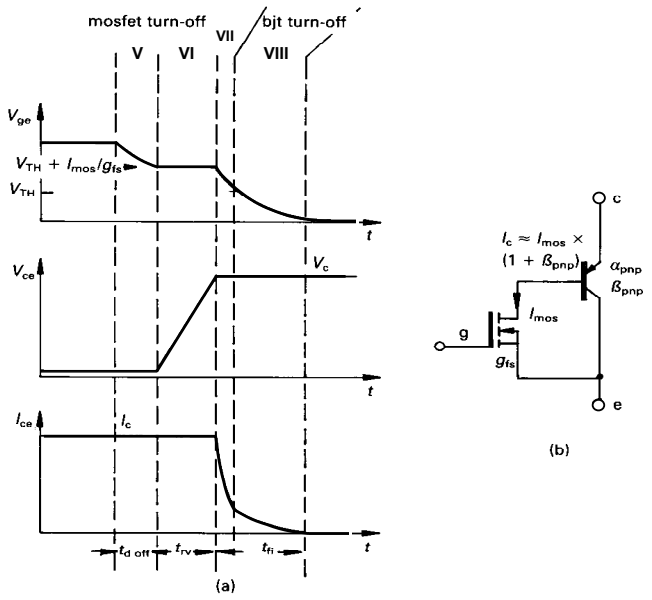


Figure 4.16. *IGBT: (a) turn-off waveforms and (b) equivalent circuit during turnoff.*

Period VIII – current tail time
With the gate voltage at the threshold level, the pnp transistor section turns off in a V_{ceo} mode, phase VIII. A relatively low-magnitude, lengthy current tail results which is dependent on the pnp transistor section minority carrier lifetime in the n base and the injection efficiency of the p^+ collector region.

The switching frequency and current rating of an IGBT are both limited by the minimum of the package dissipation limit (as with any other semiconductor device) and a factor solely dependant on the switching times at turn-on and turn-off. As the switching frequency increases, the current rating decreases. The MOSFET upper frequency is restricted solely by losses, that is, temperature.

4.5.2 IGBT short circuit operation

Under certain electrical conditions the IGBT may be subjected to short circuits, and safely turned off with out damage. Two different short circuit conditions are characterised:

I. IGBT turn-on into a pre-existing load short circuit
II. Subsequent to IGBT turn-on, a short circuit load condition occurs during the on-state period

I. Pre-existing short circuit at turn-on
The collector electrical characteristics are determined by the gate drive parameters and conditions. As the collector voltage falls, the collector current *di/dt* is determined by the stray inductance, characterised at less than 25μH. In this fault mode the IGBT is characterised for up to ten times the rated current, provided the IGBT is turned off within 10us, but at a slower rate than normal.

II. Short circuit arising during the normal on-period

When a load short circuit occurs during the IGBT on period, the collector current rises rapidly and is determined by the supply voltage V_s and stray inductance L_s according to $di_{rise}/dt = V_s/L_s$. The collector voltage de-saturates and as the collector voltage rise towards the supply V_s the resultant *dv/dt* produces a Miller capacitance charging current, which flows into the gate circuit. Depending on the gate drive impedance, the gate voltage rises, which adversely allows higher collector current.

When turn-off is initiated, by reducing the gate voltage to below the threshold level, the resultant collector current fall produces a high voltage across the stray inductance, $V_{Ls} = L_s\, di_{fall}/dt$, which adds to the collector voltage which is already near the supply rail V_s. Because of this over voltage, this mode of short circuit turn-off is more severe than turning off from a pre-existing short circuit.

The maximum allowable short circuit current at turn-off is dependant on the gate voltage and reduces from ten times rated current at a gate voltage of 18V down to five times rated current at 12V. The short circuit must be commutated within 10us at a slower than normal rate so as to ensure the over voltage due to stray inductance remains within rated voltage limits. Repetitive short circuits are restricted to a frequency of less than one Hertz and can only accumulate to 1000 before device deterioration accelerates; both mechanical bonding and electrical.

Stress during the fault period can be reduced if the gate voltage is clamped so that it cannot rise during the Miller capacitance charge period. A Zener diode (plus a reverse series diode if reverse gate bias is used) across the gate to emitter provides low inductance gate voltage clamping, but the Zener standby to clamping voltage ratio of 1:1.4 limits clamping effectiveness. The preferred method is to clamp the gate to the gate supply voltage by a Schottky diode between the gate (diode anode) and gate positive supply (diode cathode). Judicious gate supply ceramic capacitance decoupling will minimise loop inductance which otherwise would deteriorate clamping effectiveness.

A difficulty arises when attempting to utilise the 10µs short circuit capabilities of the IGBT. To improve device robustness, short circuit turn-off is staged, or slowed down. It is prudent to utilise the over current capability of the IGBT in order to reduce nuisance tripping or to briefly ignore capacitor charging which are not true faults. A difficulty arises when a demand pulse is significantly less than 10µs. The gate drive must be able to cater for sub 10µs pulses with normal turn-off yet differentiate 10µs delayed slow turn-off when a short circuit fault is serviced.

Table 4.1: The IGBT Characteristics Comparison with the BJT and MOSFET

Features	BJT	MOSFET	IGBT
Drive Method	Current	Voltage	Voltage
Drive Circuit	Complex	Simple	Simple
Input Impedance	Low	High	High
Drive Power	High	Low	Low
Switching Speed	Slow (µs)	Fast (ns)	Middle
S.O.A.	Narrow	Wide	Wide
Saturation Voltage	Low	High	Low
Series Avalanche Operation	excellent	poor	poor

4.6 The thyristor

Most of the thyristor ratings and characteristics to be considered are not specific to only the silicon-controlled rectifier, although the dynamic characteristics of the gate turn-off thyristor are considered separately.

4.6.1 SCR ratings

The fundamental four layer, three junction thyristor structures and their basic electrical properties were considered in chapter 3.3.

4.6.1i - SCR anode ratings

Thyristors for low-frequency application, such as in 50-60 Hz and 300-360Hz ac supply systems, are termed *converter-grade* thyristors. When a higher switching frequency is required, so-called *gate commutated* devices like the GTO and GCT are applicable. Such devices sacrifice voltage and current ratings for improved self-commutating capability.

The repetitive peak thyristor voltage rating is that voltage which the device will safely withstand in both the forward off-state V_{DRM}, and reverse direction V_{RRM}, without breakdown. The voltage rating is primarily related to reverse leakage or forward blocking current I_{RRM} and I_{DRM} respectively, at a given junction temperature, usually 125°C. Since forward blocking current doubles with every 10K rise in junction temperature T_j, power dissipation increases rapidly with T_j, which may lead to regenerative thermal runaway, turning the device on in the forward direction.

Current related maximum ratings reflecting application requirements include

- peak one cycle surge on-state current I_{TSM}
- repetitive and non-repetitive *di/dt*
- I^2t for fusing.

The maximum junction temperature can be exceeded during non-recurrent over-current cycles. The maximum non-repetitive on-state surge current is generally quoted for one 10 millisecond sinusoidal period at $\hat{T}_j$. Any non-recurrent rating can be tolerated only a limited number of times before failure results. Such non-recurrent ratings are usually specified to allow fuse and circuit breaker short-circuit protection. The I^2t rating for a 10ms period is another parameter used for fuse protection, where I is non-repetitive rms current. When used in 60Hz systems, the ratings are specifies with respect to 8.33ms

If the device is turned on into a fault, the initial current-time relationship, *di/dt*, during turn-on must be within the device's switching capability. In cases where the initial *di/dt* is rapid compared with the active plasma area-spreading velocity of 50 µm/µs, local hot spot heating will occur because of the high current densities in those areas that have started to conduct.

A repetitive *di/dt* rating is also given for normal operating conditions, which will not lead to device deterioration. This repetitive *di/dt* rating will be specified for a given initial blocking voltage and peak forward current. Certain gate drive conditions are specified and the device must survive for 1000 hours.

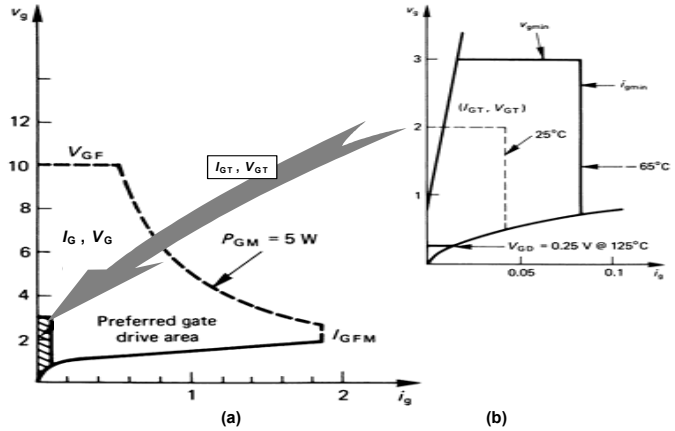


Figure 4.17. *Thyristor gate ratings illustrating: (a) the preferred operating region and (b) minimum gate requirements and their temperature dependence.*

4.6.1ii - SCR gate ratings

The gate ratings usually specified are

- peak and mean gate power, P_{GM} and P_G
- peak forward and reverse gate to cathode voltage, V_{GFM} and V_{GRM}
- peak forward gate current, I_{FMG}.

These gate ratings are illustrated in figure 4.17. The peak gate power rating is obtained by using a low duty cycle pulse, with a mean power that does not exceed P_G. The reverse gate voltage limit, V_{GRM}, is specified by the avalanche voltage breakdown limit of the reverse-biased gate-to-cathode junction. Figure 4.17 not only shows limit ratings, it also indicates the preferred gate voltage and current, and the minimum requirements which will ensure turn-on at different junction temperatures.

4.6.2 Static characteristics

The static anode voltage-current characteristics of a thyristor are similar to those of a diode. Gate commutated thyristors tend to have higher on-state voltages for a given current than comparable converter-grade devices. This higher on-state voltage is one of the trade-offs in improving the switching performance.

4.6.2i - SCR gate trigger requirements

Below a certain gate voltage, called the *gate non-trigger voltage* V_{GD}, the manufacturer guarantees that no device will trigger. This voltage level is shown in figure 4.17b. The hatched insert area in figure 4.17a (figure 4.17b) contains all the possible minimum trigger values (I_{GT}, V_{GT}) for different temperatures, that will result in turn-on. The gate requirements (I_{GT}, V_{GT}) have a negative temperature coefficient as indicated in figure 4.17b. To ensure reliable turn-on of all devices, independent of temperature, the trigger circuit must provide a dc signal (I_G, V_G) outside the shaded area. This area outside the uncertainty area, but within the rating bounds, is termed the *preferred gate drive* area.

An increase in anode supporting voltage tends to decrease the gate drive requirements. But if the gate signal is a pulse of less than about 100µs, the turn-on (I_G, V_G) requirement is increased as the pulse duration is decreased. The gate current increase is more significant than the voltage requirement increase. Typically, for a pulse reduced from 100µs to 1µs, the voltage to current increase above the original requirement is 2:10 respectively. This increased drive requirement with reduced pulse time is accounted for by the fact that some of the initial gate p-region charge recombines. When the free charge reaches a certain level the device triggers. Thus, to get the required charge into the gate in a relatively short time compared with the recombination time requires higher current, and hence higher voltage, than for dc triggering.

4.6.2ii – SCR holding and latching currents

If the on-state anode current drops below a minimum level, designated as the *holding current* I_H, the thyristor reverts to the forward blocking state. This occurs because the loop gain of the equivalent circuit pnp-npn transistors falls below unity and the regenerative hold-on action ceases. The holding current has a negative temperature coefficient; that is, as the junction temperature falls, the device holding current requirement increases. The holding current is typically about 2% of the rated anode current, and increases as switching performance is improved (and on-state voltage increases).

A somewhat higher value of anode on-state current than the holding current is required for the thyristor to latch on initially ($I_{Latch} > I_H$). If this higher value of anode *latching current* I_{Latch} is not reached, the thyristor will revert to the blocking state as soon as the gate trigger signal is removed. After latch-on, however, the anode current may be reduced to the holding current level, without turn-off occurring. These two static current properties are shown in the *I-V* characteristics in figure 3.22. With inductive anode circuits, it is important to ensure that the anode current has risen to the latching current level before the gate turn-on signal is removed. Continuous gate drive avoids this inductive load problem but at the expense of increased thyristor gate power losses.

4.6.3 Dynamic characteristics

The main thyristor dynamic characteristics are the turn-on and turn-off switching intervals, which are associated with the anode and gate circuit interaction.

4.6.3i – SCR anode at turn-on

Turn-on comprises a delay time t_d and a voltage fall time t_{fv}, such that the turn-on time is $t_{on} = t_d + t_{fv}$.

The turn-on delay time for a given thyristor decreases as the supporting anode voltage at turn-on is increased. The delay time is also decreased by increased gate current magnitude. The gate p-region width dominates the high gate current delay time characteristics while carrier recombination is the dominant factor at low gate current levels.

The anode voltage fall time is the time interval between the 90 per cent and 10 per cent anode voltage levels. The associated anode current rise characteristics are load dependent and the recurrent *di/dt* limit must not be exceeded.

As introduced in chapter 3.3.1, a thyristor can be brought into conduction by means of an anode impressed *dv/dt*, called *static dv/dt* capability, even though no gate external current is injected. The anode voltage ramp produces a displacement current according to $i = dQ/dt$ as the central junction scl charges and its width increases. The resultant displacement current flows across the cathode and anode junctions causing minority carrier emission and, if sufficient in magnitude, turn-on occurs. Static *dv/dt* capability is an inverse function of device junction temperature and is usually measured at $\hat{T}_j$.

4.6.3ii – SCR anode at turn-off

As analysed in chapter 3.3.1, once a thyristor is turned on, it remains latched-on provided

- the holding current remains exceeded
- it is forward biased.

If the supply voltage is ac, a thyristor will turn off after the supply voltage has reversed and the anode current attempts to reverse. The thyristor is thus reverse-biased and this turn-off process is called *line commutation* or *natural commutation,* as defined in chapter 6.3.4.

If the supply voltage is dc and the load is a series *L-C* resonant circuit, the anode current falls to zero when the capacitor is charged. The load current falls below the holding current level and the SCR turns off. This is termed *load commutation,* which is a form of load resonant switching as defined in 6.3.3.

In thyristor applications involving dc supplies and resistive/inductive loads, a thyristor once on will remain on. Neither the supply nor the load is capable of reducing the anode current to below the holding current level, or producing a reverse bias across the thyristor. Such a thyristor can be turned off only if the anode current is interrupted or forced below the holding current level. External circuitry, called a *commutation circuit,* is employed to accomplish turn-off, by reverse-biasing the thyristor and reducing the anode current to near zero. This external turn-off approach, now obsolete, is called thyristor *forced commutation*. A topological variation of the forced commutated circuitry method is called *resonant link commutation*. The gate turn-off thyristor eliminates the need for this external commutation circuitry since the GTO can be commutated from its gate using reverse gate current.

4.7 The gate turn-off thyristor

In essence, the gate turn-off (GTO) thyristor has similar ratings and characteristics to those of the conventional converter grade SCR, except those pertaining to turn-off. Both GTO turn-on and turn-off are initiated from the gate, hence the power-handling capabilities of the GTO gate are much higher than those of SCR devices.

4.7.1 Turn-on characteristics

Because of the higher p1 gate region concentration, the GTO thyristor holding current level and gate trigger requirements are somewhat larger than those of the conventional SCR. Higher anode on-state voltages also result.

At low anode current levels, a steep trailing edge at the end of the gate on-pulse may cause the GTO to unlatch even though the anode current is above the dc holding current level. For this reason, together with the fact that the cathode comprises many interdigitated islands, a continuous, dc gate on-drive is preferred. Continuous gate current prevents any cathode islands from falling out of conduction should the anode current be reduced to near the holding current level. If cathode islands should turn off prematurely and the anode current subsequently rise, the GTO no longer has its full current handling capability and it could overheat specific islands, leading to device destruction.

With very high voltage GTO's, turn-on is like that of a high voltage npn transistor which has low gain, limiting the initial rate of rise of anode current, until the regenerative latching action has occurred. Hence an initial, high current of up to six times the steady-state gate requirement is effective for a few microseconds.

4.7.2 Turn-off characteristics

Before commencing turn-off, a minimum on-time of tens of microseconds must be observed so that the principal current may distribute uniformly between the cathode islands. This is to ensure that all cells conduct, such that turn-off occurs uniformly in all cells, rather than being confined to a few cells, where the current to be commutated may be higher than individual cells can survive.

The anode current of a GTO in the on-state is normally turned off via a low voltage source, negative gate current, I_{RG}. The negative gate current I_{GQ}, which is just sufficient to turn-off the on-state current I_T is defined as the minimum turn-off current. Turn-off amplification (equation 3.21) is defined as

$$\beta_Q = I_{TGQ} / I_{GQ} \qquad (4.37)$$

where β_Q is related to the internal construction of the GTO thyristor.

Figure 4.18 illustrates typical gate and anode turn-off waveforms for the GTO. Application of reverse gate current causes the anode current to reduce after a delay period t_s. This delay time is decreased as the reverse gate current di_{GQ}/dt increases; that is, as I_{RGM} increases and t_s decreases. Increased anode on-state current or junction temperature increases the delay time and turn-off time.

The reverse gate current prevents cathode injection and the anode current rapidly falls to the storage current level, I_{tail}. The subsequently slow current fall time, t_{tail}, is due to charges stored in regions other than the gate and cathode that are not influenced by the reverse gate current and must decrease as a result of natural recombination, producing a decaying principal anode current. Anode n^+ shorts are used to accelerate the recombination process, reducing both storage current and storage time, but at the expense of reverse blocking ability and on-state voltage. Avalanche of the cathode junction (typically -20V) is acceptable during turn-off for a specified time. Reverse gate bias should be maintained in the off-state in order to prevent any cathode injection.

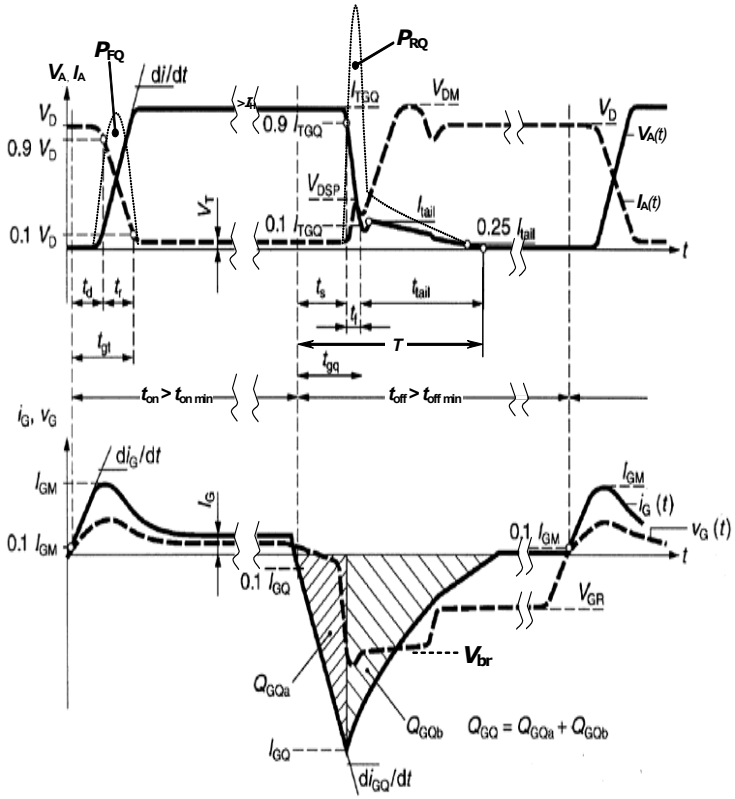


Figure 4.18. *Schematic representation of GTO thyristor turn-off waveforms.*

After turn-off some dispersed charges still exist. A minimum off-time of the order of tens of microseconds is needed for these charges to recombine naturally. This time increases with increased blocking voltage rating. If a turn-on were to be initiated before this recombination is complete, the area of un-combined charge will turn-on first, resulting in a high *di/dt* in a confined area, which may cause a hot spot and possibly destruction.

During the storage and fall time, power loss P_{RQ} occurs as illustrated in figure 4.18 and is given by

$$P_{RQ} = \frac{1}{T}\int_0^T V_A(t)\, I_A(t)dt \qquad \text{(W)} \qquad (4.38)$$

where $T = t_{gq} + t_{tail}$. The cathode junction loss, due to the gate turn-off reverse current can also be incorporated, which may become significant as the turn-off gain reduces to unity.

The actual anode voltage turn-off waveform is dependent on the load circuit. Care is needed in preventing excessive loss at turn-off, which can lead to device destruction. One technique of minimising turn-off loss is to increase the rate at which the reverse gate current is applied. Unfortunately, in reducing the turn-off time, the turn-off current gain β_Q is decreased, from typically 25 to 3. The anode turn-off voltage $V_A(t)$ in figure 4.18 assumes a capacitive turn-off snubber is used. Such a capacitive switching aid circuit is not essential with the GCT, which uses unity reverse gain at turn-off, as considered in chapter 3.3.5.

4.8 Appendix: Effects on MOSFET switching of negative gate drive

The effects of negative gate voltage on MOSFET turn-on and turn-off delays, which were analysed in section 4.4.2, are given by

$$V_{gs}(t) = (V_{gg} - V_{gg-})\,[1 - e^{-t/C_{in}R_g}] + V_{gg-} \qquad \text{(V)} \qquad (4.39)$$

$$t_{d\,on} = R_g C_{in}\, \ell n\, \frac{V_{gg} - V_{gg-}}{V_{gg} - V_{TH}} \qquad \text{(s)} \qquad (4.40)$$

$$V_{gs}(t) = (V_{gg} - V_{gg-})\, e^{-t/R_g C_{in}} + V_{gg-} \qquad \text{(V)} \qquad (4.41)$$

$$t_{d\,off} = R_g C_{in}\, \ell n \frac{V_{gg} - V_{gg-}}{V_{TH} + I_L / g_{fs} - V_{gg-}} \qquad \text{(s)} \qquad (4.42)$$

Reading list

See Chapter 3 reading list

Van Zeghbroeck, B., *Principles of Semiconductor Devices,* //ece-www.colorado.edu/~bart/book, 2004.

Power device manufacturers

http://www.infineon.com/eupec/toc.htm
http://www.fujielectric.co.jp/eng/fdt/scd/index.html
http://www.irf.com/indexsw.html
http://www.onsemi.com/
http://www.pwrx.com/
http://www.st.com/stonline/products/families/transistors/transistors.htm
http://www.ixys.com/
http://www.microsemi.com/
http://www.semikron.com/
http://www.fairchildsemi.com/index2.html
http://www.dynexsemi.com/
http://www.westcode.com/products.html
http://www.infineon.com/cgi-bin/ifx/portal/ep/home.do?tabId=1
http://www.abb.com/product/us/9AAC910029.aspx?country=00
http://www.mitsubishichips.com/Global/products/power/index.html
http://www.semicon.toshiba.co.jp/eng/
http://www.nxp.com/products/power_management/index.html

5

Cooling of Power Switching Semiconductor Devices

Semiconductor power losses are dissipated in the form of heat, which must be transferred away from the switching junction, if efficient switching is to be maintained. The reliability and life expectancy of any power semiconductor are directly related to the maximum device junction temperature experienced. It is therefore essential that the thermal design determine accurately the maximum junction temperature from the device power dissipation. Every 10°C junction temperature decrease, doubles device lifetime.

i. Heat load

The heat load may be active or passive, then there is a combination of the two. An active load is the heat dissipated by the device being cooled. It is generally equal to the input power to the device, for example, $P = V \times I = I^2 \times R = V^2 / R$. Passive heat loads are indirect, are parasitic in nature, and may consist of radiation, convection or conduction.

Heat energy, due to a temperature difference, can be transferred by any of, or a combination of, three mechanisms, viz.,

- Convection - heat transferred to a moving fluid which takes the heat away
- Conduction - heat flows through a thermal conducting material, away from the heat source
- Radiation - heat flow by long-wave electromagnetic radiation, e.g. infrared.

Electromagnetic thermal *radiation* heat transfer

When two objects at different temperatures come within proximity of each other, heat is exchanged between them. Electromagnetic wave propagation, radiation is emitted from one object and absorbed by the other. As a result of the temperature difference, the hot object experiences a net heat loss and the cold object undergoes a net heat gain. This is termed thermal radiation.

Radiation heat loads are usually considered insignificant when the system is operated in a gaseous environment since other passive heat loads are usually greater. Radiation loading is usually significant in systems with small active loads and large temperature differences, especially when operating in a vacuum environment, where convection processes are absent.

Electromagnetic thermal ***radiation*** heat loading (for a grey body, $\varepsilon < 1$) is given by

$$P_d = \sigma \varepsilon A (T_1^4 - T_2^4) \qquad (5.1)$$

where P_d is the rate of radiated heat transfer (that is, the power dissipated), W
σ is the Stefen-Boltzmann constant (5.667×10^{-8} W/m^{2}K^4)
ε is a surface property, termed emissivity, $0 \le \varepsilon \le 1$, see Table 5.6 and Appendix 5.25
A is the surface area involved in the heat transfer, m^2
T is absolute temperature, K

The ideal emitter, or black body, is one which gives off radiant energy with $\varepsilon = 1$ in equation (5.1).

The one-dimensional model for general molecular (non-radiation) heat transfer is given by

$$P_d = -\lambda A \frac{\delta T}{\delta \ell} + \rho_m A \ell \frac{\delta T}{\delta t} \quad \text{(W)} \qquad (5.2)$$

where $\delta T = T_2 - T_1$ or ΔT, is the temperature difference between regions of heat transfer
λ is thermal conductivity, W/m K, see Appendix 5.24
ρ_m is density of the heatsink material
c_p is specific heat capacity, J/kg K, such that $\Delta T = W/mc_p$ (W is energy, m is mass)
ℓ is distance (thickness).

Equation (5.2) shows that the thermal power generated P_d is balanced by the stored thermal power (first term on the right hand side) and the thermally dissipated power (second term on the right hand side).

Conduction **heat transfer**

Conductive heat transfer occurs when energy exchange takes place, by direct impact of molecules, from a high temperature region to a low temperature region.

Conductive heat loading on a system may occur through lead wires, mounting screws, etc., which form a thermal path from the device being cooled to the heat sink or ambient environment.

Assuming steady-state heat dissipation conditions, then $\delta T / \delta t = 0$ in equation (5.2).

Conduction through a homogeneous solid, from Fourier's law of heat conduction, is therefore given by

$$P_d = \frac{\lambda}{\ell} A \Delta T \quad \text{(W)} \qquad (5.3)$$

Convection **heat transfer**

When the temperature of a fluid (a gas or liquid) flowing over a solid object differs from that of the object surface, heat transfer occurs. The amount of heat transfer varies depending on the fluid flow rate. Convective heat loads are generally a result of natural (or free) convection. This is the case when gas flow is not artificially created as by a fan or pump (forced convection), but rather occurs naturally from the varying density in the gas caused by the temperature difference between the object being cooled and the gas. Heat transfer processes that involve *change of phase* of a fluid (for example evaporation or condensation) are also considered to be convection.

The convective loading is a function of the exposed area and the difference in temperature between the load and the surrounding gas. Convective loading is usually most significant in systems operating in a gaseous environment with small active loads or large temperature differences.

Convection heat transfer through a fluid or air, under steady-state conditions in equation (5.2), is given by Newton's law of cooling, that is

$$P_d = h A \Delta T \quad \text{(W)} \qquad (5.4)$$

The convection heat transfer coefficient h ($= \lambda / \ell$), W/m^{2}K, depends on the heat transfer mechanism used and various factors involved in that particular mechanism. It is not a property of the fluid.

Natural or *free convection* is essentially still to slightly stirred air with h values ranging from 1 to 25. *Forced convection* is air moved by a fan or other active method, giving h values range from 10 to 100. Values for forced liquid convection are 50 to 20,000, while the range for boiling and condensation is 2,500 to 100,000.

For natural vertical convection in free air, the losses for a plane surface may be approximated by the following empirical formula

$$P_d = 1.35 A \sqrt[4]{\frac{\Delta T^5}{\ell}} = 1.35 A \sqrt[4]{\frac{\Delta T}{\ell}} \Delta T = hA\Delta T \quad \text{(W)} \qquad (5.5)$$

where ℓ is the vertical height in the direction of the airflow and h is of the form

$$h = K \left(\frac{\Delta T}{\ell} \right)^{1/4} \qquad (5.6)$$

Two cases occur for forced airflow, and the empirical losses are

- *for laminar flow*

$$P_d = h A \Delta T = 3.9 \sqrt{\frac{v}{\ell}} A \Delta T \quad \text{(W)} \qquad (5.7)$$

- *for turbulent flow*

$$P_d = h A \Delta T = 6.0 \sqrt[5]{\frac{v^4}{\ell}} A \Delta T \quad \text{(W)} \qquad (5.8)$$

where v is the velocity of the vertical airflow.

Combined *convection* and *conduction* heat transfer
Heat Loss (or gain) - through the walls of an insulated container (combined convection and conduction, estimation) is

$$P_d = A \times \frac{\Delta T}{\ell / \lambda + 1/h}$$

where P_d is the heat lost or gained, W
ℓ is the thickness of insulation, m
λ is the thermal conductivity of the insulation material, W/m K
A is the outside surface area of the container, m^2.
h is the convection heat transfer coefficient of the surface material, W/m^2 K
$\Delta T = T_{o/s} - T_{i/s}$
$T_{o/s}$ is the outside temperature, °C
$T_{i/s}$ is the inside temperature, °C

ii. Transient heating
Some designs require a set amount of time to reach the desired temperature.
The estimated time required to heat (or cool) an object (also known as Newton's Law of Cooling) is

$$t = \frac{m \times c_p \times \Delta T}{P}$$

P is the mean heat added (or being removed) from the object, W, watts
m is the mass (weight) of the object, kg (density x volume)
c_p is the specific heat of the object material, J/kg K
t is the time required to cool down (or heat up) the object in seconds
$\Delta T = T_o - T_f$
T_o is the starting temperature, °C
T_f is the final temperature, °C

$$P = \tfrac{1}{2}\left(P_{t_o} + P_{t_f}\right)$$

P_{to} is the initial heat pumping capacity when the temperature difference across the cooler is zero.
P_{tf} is the heat pumping capacity when the desired temperature difference is reached and heat-pumping capacity is decreased.
Heat loading may occur through one or more of four modes: active, radiation, convection or conduction. By utilizing these equations, the heat load can be estimated.

iii. Thermal resistance
It is generally more convenient to work in terms of thermal resistance, which is defined as the ratio of temperature change to power. Thermal capacity is the reciprocal of thermal resistance. For conduction, from equation (5.4), thermal resistance R_θ is

$$R_\theta = \frac{\Delta T}{P_d} = \frac{1}{hA} = \frac{\ell}{\lambda A} \qquad \text{(K/W)} \qquad (5.9)$$

where the conduction thermal heat transfer coefficient, h, is

$$h = \frac{\lambda}{\ell} \qquad (5.10)$$

The average power dissipation P_d and maximum junction temperature $\hat{T}_j$, in conjunction with the ambient temperature T_a, determine the necessary heat sink, according to equation (5.9)

$$P_d = \frac{\hat{T}_j - T_a}{R_{\theta j\text{-a}}} \qquad \text{(W)} \qquad (5.11)$$

where $R_{\theta\ j\text{-a}}$ is the total thermal resistance from the junction to the ambient air. The device user is restricted by the thermal properties from the junction to the case for a particular package, material, and header mount according to

$$P_d = \frac{\hat{T}_j - T_c}{R_{\theta j\text{-c}}} \qquad \text{(W)} \qquad (5.12)$$

where T_c is the case temperature, K and
$R_{\theta j\text{-c}}$ is the package junction-to-case mounting thermal resistance, K/W.

An analogy between the thermal equations and Ohm's law and Kirchhoff's laws is often made to form models of heat flow. The temperature difference ΔT could be thought of as a voltage drop ΔV, thermal resistance R_θ corresponds to electrical resistance R, and power dissipation P_d is analogous to electrical current I. *[viz., $\Delta T = P_d R_\theta \equiv \Delta V = IR$].* See Table 5.10.

For convection, from equation (5.4), the effective thermal resistance is

$$R_\theta = \frac{\Delta T}{P_d} = \frac{1}{hA} \qquad (5.13)$$

For radiation, from equation (5.1), the effective thermal resistance of radiation is

$$R_\theta = \frac{\Delta T}{P_d} = \frac{1}{\sigma\varepsilon A \times (T_1 + T_2)\left(T_1^2 + T_2^2\right)} = \frac{1}{h_r A} \qquad (5.14)$$

where the radiation heat transfer coefficient, h_r, in W/m^2K, is

$$h_r = \sigma\varepsilon \times (T_1 + T_2)\left(T_1^2 + T_2^2\right) \approx 4 \times \sigma\varepsilon \times T_{mean}^3 \qquad (5.15)$$

where T_{mean} is the arithmetic mean of T_1 and T_2, specifically ½($T_1 + T_2$).

5.1 Thermal resistances

A general thermal dissipation model or thermal equivalent circuit for a mounted semiconductor is shown in figure 5.1. The total thermal resistance from the virtual junction to the open air (ambient), $R_{\theta\,j\text{-a}}$, is

$$R_{\theta j\text{-a}} = R_{\theta j\text{-c}} + \frac{R_{\theta c\text{-a}} \times (R_{\theta c\text{-s}} + R_{\theta c\text{-a}})}{R_{\theta c\text{-a}} + R_{\theta c\text{-s}} + R_{\theta s\text{-a}}} \qquad \text{(K/W)} \qquad (5.16)$$

In applications where the average power dissipation is of the order of a watt or so, power semiconductors can be mounted with little or no heat sinking, whence

$$R_{\theta j\text{-a}} = R_{\theta j\text{-c}} + R_{\theta c\text{-a}} \qquad \text{(K/W)} \qquad (5.17)$$

Generally, when employing heat sinking, $R_{\theta\ c\text{-a}}$ is large compared with the other model components and equation (5.16) can be simplified to three series components:

$$R_{\theta j\text{-a}} = R_{\theta j\text{-c}} + R_{\theta c\text{-s}} + R_{\theta s\text{-a}} \qquad \text{(K/W)} \qquad (5.18)$$

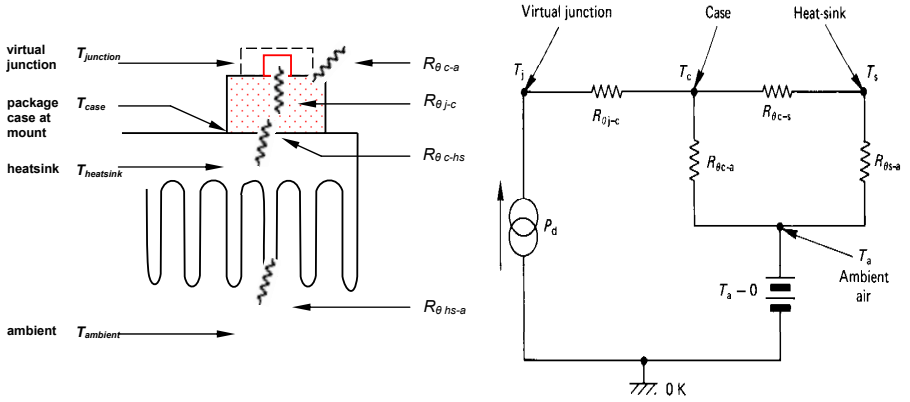


Figure 5.1. *Semiconductor thermal dissipation equivalent circuit.*

5.2 Contact thermal resistance, $R_{\theta\ c\text{-s}}$

The case-to-heat-sink thermal resistance $R_{\theta\ c\text{-s}}$ (case means the device thermal mounting interface surface) depends on the package type, interface flatness and finish, mounting pressure, and whether thermal-conducting grease and/or an insulating material (thermal interface material, TIM) is used. In general, increased mounting pressure decreases the interface thermal resistance, and no insulation but thermal grease results in minimum $R_{\theta\ c\text{-s}}$. Common electrical insulators are mica, aluminium oxide, and beryllium oxide in descending order of thermal resistance, for a given thickness and area. Table 5.1 shows typical contact thermal resistance values for smaller power device packages, with various insulating and silicone grease conditions. Silicon based greases are best, for example Assmann V6515, spread at a thickness of 100µm to 150µm, on both surfaces. Grease in excess of this will be squeezed out under clamping pressure. Initial grease thermal resistance decreases slightly after a few normal deep thermal cycles.

Table 5.1: Typical case-to-heat-sink thermal resistance value for various small packages

Package	*Insulating washer*	$R_{\theta c\text{-}s}$ (K / W) *Silicone grease* *with*	*without*
TO-3	No insulating washer	0.10	0.3
	Teflon	0.7-0.8	1.25-1.45
	Mica (50 -100 µm)	0.5-0.7	1.2-1.5
TO-220	No insulating washer	0.3-0.5	1.5-2.0
	Mica (50 -100 µm)	2.0-2.5	4.0-6.0
TO-247	No insulating washer	0.1-0.2	0.4-1.0
	Mica (50 -100 µm)	0.5-0.7	1.2-1.5
SOT-227 ISOTOP	No insulating washer	0.1-0.2	0.3-0.4
	Mica (50 -100 µm)	0.5-0.7	1.0-1.2

The thermal resistance of a heat-conducting layer is inversely proportion to heat conductivity of the material and in direct ratio to its thickness. If the clamping pressure is increased, the layer thermal resistance falls. In figure 5.2, the exemplary dependence of the gasket thermal resistance per surface unit on pressure is shown. However, with a growth of pressure it is necessary to find an optimum, as the clamping effort should not exceed a package recommended value or introduce differential thermal expansion problems into the clamping arrangement.

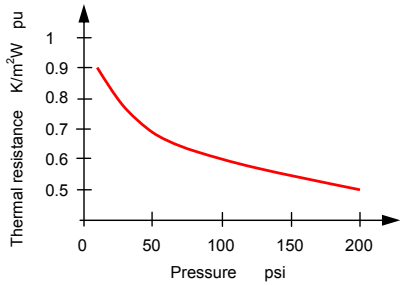


Figure 5.2. Exemplary dependence of the gasket thermal resistance on clamping pressure.

5.2.1 Thermal Interface Materials

To be effective, heatsinks require intimate surface-to-surface contact with the component to be cooled. Unfortunately, irregular surface areas, both on the electronic components and on the heatsink mating surface prevent good contact. Up to 99% of the surfaces are separated by a layer of interstitial air, which is a poor conductor of heat thus presents a thermal barrier. Therefore, a thermally conductive interface material is necessary to fill the interstices and microvoids between the mating surfaces. To ensure that electrical problems are not inadvertently introduced while solving the thermal problems, it is often essential that the thermal interface materials also perform an electrical isolation function. Thermal interface materials TIMs vary widely in terms of their performance (that is, thermal, electrical, and physical properties), their general appearance, and their mode of application. Among the most commonly used classes of thermal interface materials are: thermal greases, cure-in-place thermally conductive compounds, gap filling thermally conductive elastomeric pads, thermally conductive adhesive tapes, and phase change materials, all of which are summarised in Table 5.2 and are briefly described.

- Thermal Greases
  Comprised of thermally conductive ceramic fillers in silicone or hydrocarbon (organic) oils, as shown in Tables 5.2a and b, a thermal grease is a paste, which is applied to at least one of the two mating surfaces. When the surfaces are pressed together, the grease spreads to fill the void. During compression, excess grease squeezes out from between the mated surfaces. Some form of clip or other mounting hardware is needed to secure the joint. Although it is comparatively inexpensive and thermally effective, thermal grease is not an electrical insulator. Disadvantageously, it can be inconvenient to dispense and apply, and requires cleanup to prevent contamination problems.

- Cure-in-Place Thermally Conductive Compounds
  A thermally conductive compound again incorporates thermally conductive ceramic fillers, shown in Table 5.3, but unlike thermal greases, the binder is a rubber material. When first applied, the paste-like compound flows into the interstices between the mating surfaces. Then, when subjected to heat, it cures into a dry rubber film. Besides its thermal properties, this film also serves as an adhesive, allowing a tight, void-free joint without the need for additional fasteners. Thermally conductive compounds can fill larger gaps in situations where thermal greases might ooze from the joint. Although application and performance is similar to that of thermal grease, cleanup is easier, simply involving removal of the excess cured rubber film.

- Thermally Conductive Elastomeric Pads
  A thermally conductive elastomeric pad consists of a silicone elastomer filled with thermally conductive ceramic particles and may incorporate woven glass fibre or dielectric film reinforcement. Typically ranging in thickness from 0.1 - 5 mm and in hardness from 5 to 85 Shore A, these pads provide both electrical insulation and thermal conductivity, making them useful in applications requiring electrical isolation. Thicker pads prove useful when large gaps must be filled. During application, the pads are compressed between the mating surfaces to make them conform to surface irregularities. Mounting pressure must be adjusted according to the hardness of the elastomer to ensure that voids are filled. A mechanical fastener is essential to maintain the joint once it is assembled.

- Thermally Conductive Adhesive Tapes
  A thermally conductive adhesive tape is a double-sided pressure sensitive adhesive film filled with thermally conductive ceramic powder. To facilitate handling, aluminium foil or a polyamide film may support the tape; the latter material also provides electrical insulation. When applied between mating surfaces, the tape must be subjected to pressure to conform to the surfaces. Once the joint is made, the adhesive holds it together permanently, eliminating the need for supplemental fasteners. No bond curing is needed. One limitation of thermally conductive tapes is that they cannot fill large gaps between mating surfaces as well as liquids; hence, the convenience of tape mounting is traded against a nominal sacrifice in thermal performance.

- Phase Change Materials
  Solid at room temperature, phase change materials, shown in Table 5.3, melt (that is, undergo a phase change) as the temperature rises to the 40° to 70°C range. This makes the material (0.13 mm thick in its dry film form) as easy to handle as a pad, while assuring, when subjected to heat during the assembly process, the melt flows into voids between mating surfaces as effectively as a thermal grease. Applying power to the power electronic component introduces the needed heat for the phase change to occur, establishing a stable thermal joint. These materials consist of organic binders (that is, a polymer and a low-melt-point crystalline component, such as a wax), thermally conductive ceramic fillers, and, if necessary, a supporting substrate, such as aluminium foil or woven glass mesh. See section 5.2.2 for further details.

Table 5.2: Thermal Interface Material (TIM) thermal resistances

Interface	Thickness mm	Thermal Conductivity λ W / m.K	Thermal resistance $R_{\theta c\text{-}s}$ K / W	Thermal resistance pu area K cm^2/W
Dry Joint	n/a	n/a	2.9	1.8 - 2
Thermal Grease	0.076	0.7	0.9	0.5 - 1.1
Thermal Compound	0.127	1.2	0.8	0.2 - 0.7
Elastomer	0.254	5.0	1.8	1 - 2
Adhesive Tape	0.229	0.7	2.7	0.5 - 1.5
Eutectic (soldering) Sn(91) Zn(9)		61		0.1

Table 5.3: Thermal Interface material (TIM) parameters

Material parameters	Thermal Conductivity	Resistivity	Temperature range	Thickness		Lifetime
(a) Thermal grease	λ, W/mK	ρ, Ω.cm	°C	mm		
Zinc Oxide/Silicone	0.74	> 5×10^{14}	< 150			5 years @ 25C
$A\ell_2O_3$ / Non-Silicone	2	> 10^{13}	< 150			1 year @ 25°C
AℓN / Non-Silicone	4	> 10^{13}	< 150			1 year @ 25°C
Non-silicone paste	2	> 1×10^{13}	n/a			1 year @ 25°C
Silicone paste	> 4.0	> 1×10^{14}	n/a			1 year @ 25°C
Non-curing paste	4	> 1×10^{13}	n/a			n/a
Non-silicon polysynthetic oils	> 7.5	1Å	-50…+150			n/a
thermal phase change	0.90	1×10^{14}	-50…+200			n/a
Silicone	0.8-1.2	1×10^{13}	-60…+180			5 years
(b) Heat conducting gaskets					Tensile Strength psi	
Silicon/Alumina filled	0.38	1×10^{14}	-60…+150	0,08	3000	
Silicon/Boron Nitride	2.07	1×10^{14}	-60…+200	0,25	1000	
$A\ell_2O_3$ filled	1.7	n/a	< +150	0,10	15…300	
AℓN filled	3.6	n/a	< +150	0,10	15…300	
97% Graphite filled	5	n/a	< +450	0,13	15…300	
(c) Phase changing heat conducting gaskets					Clamping psi	Phase Change Temperature °C
Black/self-adhesive layer	0.7	n/a	< +120	0.077	10…200	< +120
Grey/self-adhesive layer	1.0	5×10^{15}	-60…+125	0.13	20…60	-60…+125
Grey/self-adhesive layer	0.6	5×10^{14}	-60…+125	0.18	50…300	-60…+125
Grey/self-adhesive layer	3	n/a	n/a	0.11	n/a	n/a

5.2.2 Phase Change Gasket Materials (solid to liquid)

The inavertable heat produced by power electronics necessitates a carefully designed thermal path in which all of the thermal resistances are minimized. For the case-to-heatsink interface, this requires that thermal grease be used to minimize the interface resistance. Phase change materials, PCM's, are an alternative to the messy application and migration problems associated with thermal grease.

The term *phase change* describes a class of materials that are solid at room temperature and change to a liquid as temperature increases. This phase change, or melting, occurs in the range of 40 to 70°C. PCM's are composed of a mixture of organic binders, fine particle ceramic fillers for thermal enhancement, and, optionally, a supporting substrate, such as aluminium foil or a woven glass mesh. The organic binder is a blend of a polymer and a low-melt-point crystalline component, such as a wax. The ceramic fillers may be $A\ell_2O_3$, BN, AℓN or Zn0.

The way a PCM performs compared to a dry interface joint and thermal grease is illustrated in figure 5.3, where the case to heatsink temperature difference is plotted against elapsed time after the commencement of power dissipated. The curve representing the dry interface shows rapid thermal equilibrium, at about 13°C. The curve involving the use of thermal grease shows the same rapid rise to thermal steady-state but at a lower temperature difference of 4°C. The thermal grease significantly reduces the interface resistance by eliminating the interstitial air.

The PCM - a 0.1 mm thick dry film - behaves as a combination of the two interfaces. Initially at power up, the cool components give the dry interface behaviour, with the temperature difference rapidly increasing to about 12°C. As the system temperature increases, the PCM melts and the clamping pressure exerted by the clamping mechanism forces the liquid to spread in the thermal joint. As the liquid spreads, the molten PCM displaces the interstitial air and the distance between the surfaces

decreases. Both of these processes act to reduce the thermal resistance of the interface and the temperature difference is seen to decrease rapidly, reaching the performance of thermal grease, 4°C. In effect, the solid PCM film has turned into thermal grease and a grease-like joint has been formed. The next time the thermal load is activated, the interface does not experience the large temperature difference because the void free thermal joint has already been established.

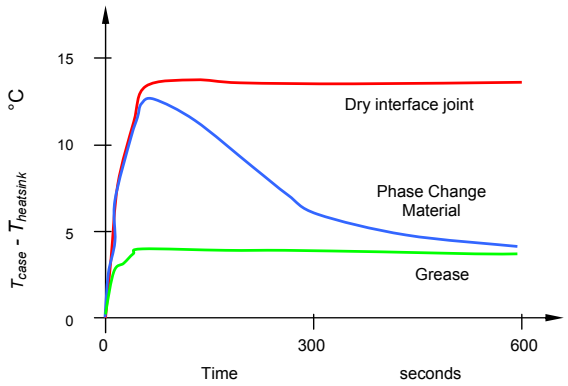


Figure 5.3. *Performance of a PCM, compared to a dry interface joint and thermal grease.*

The thermal resistance across an interface depends on the thermal conductivity of the PCM in the interface and its conductive path length. Thermal conductivity is a function of the type and level of the ceramic filler in the formulation, typically between 0.7 and 1.5 W/m.K. The amount of filler that can be added is limited by the need to keep the viscosity as low as possible to achieve proper flow of the PCM in the interface. The thickness of the interface formed by a PCM is determined by the flatness of the mating surfaces, the clamping pressure, and the viscosity and rheology of the molten PCM. Most commercial surfaces deviate from true flatness by as much as 2µm/mm. This means that the thermal path between a module surface and the heatsink may be as much as 100µm, and more with large heatsinks. Critical applications may require a better surface flatness through additional machining operations to reduce the thermal path.

The viscosity and the rheology of the PCM above its melt point represent another factor determining the thickness of the interface. As the PCM melts in the interface, the pressure applied by the mounting clamps forces the liquid to spread, eliminating the interstitial air and allowing the space between the two surfaces to decease. If the viscosity is high, the low force of the clamps will be insufficient to cause sufficient spreading and the conduction path will be long. Low viscosity on the other hand will allow the liquid to fill most of the joint, resulting in the thinnest joint. Using a stronger mounting force will aid the spreading process, but there is a package limit as to the amount of pressure that can be applied.

Phase change materials offer the same thermal performance as thermal grease without the mess and contamination associated with grease. They can be supplied attached to a heatsink as a dry film. As soon as they are heated above their phase change temperature, they melt and perform as well as, or often better than, thermal grease. Once this interface has been formed, it remains stable until the sink is physically separated from the power module case-mounting surface.

5.3 Heat-sinking thermal resistance, $R_{\theta\,s\text{-}a}$

The thermal resistance for a flat square plate heat sink may be approximated by

$$R_{\theta s\text{-}a} = \frac{3.3}{\sqrt{\lambda\, b}} C_f^{1/4} + \frac{650}{A} C_f \qquad \text{(K/W)} \qquad (5.19)$$

Typical values of heatsink thermal conductance λ in W/K cm at 350 K, are shown in Appendix 5.24 and

b is the thickness of the heat sink, mm

A is the area of the heat sink, cm^2

C_f is a correction factor for the position and surface emissivity of the heat-sink orientation according to Table 5.4.

The correction factor C_f illustrates the fact that black surfaces are better heat radiators and that warm air rises, creating a 'chimney' effect. Equation (5.19) is valid for one power-dissipating device, in the centre of the sink, at a static ambient temperature up to about 45°C, without other radiators in the near vicinity.
In order to decrease thermal resistance, inferred by equation (5.9), finned-type heat sinks are employed which increase sink surface area. Figure 5.6 illustrates graphs of thermal performance against length for a typical aluminium finned heat sink. This figure shows that $R_{\theta s\text{-}a}$ decreases with increased sink length.

Table 5.4: Heatsink correction factor

Surface position C_f	shiny	blackened
vertical	0.85	0.43
horizontal	1.0	0.5

Table 5.5: Fin spacing versus flow and fin length

Fin length (mm) H	75	150	225	300
Airflow (m/s) v	Fin Spacing (mm) S			
natural convection	6.5	7.5	10	13
1.0	4.0	5.0	6.0	7.0
2.5	2.5	3.3	4.0	5.0
5.0	2.0	2.5	3.0	3.5

Table 5.6: Emissivity coefficient of various surface treatments at 100°C

Surface	Emissivity ε
Polished aluminium	0.05
Polished copper	0.07
Rolled sheet steel	0.66
Oxidised copper	0.70
Black anodised aluminium	0.70 - 0.90
Black air-drying enamel	0.85 - 0.91
Dark varnish	0.89 - 0.93
Black oil (organic) paint	0.92 - 0.96
Al_20_3	0.15

If the fin thickness, t, is small relative to the fin spacing, s, the following equation can be used for estimating the thermal resistance of a vertical heat sink in natural convection.

$$R_{\theta hs-a} = \frac{1}{h \times \text{total fin area}} = \frac{1}{h \times (2n_f LH)} \qquad (5.20)$$

where a fin efficiency of unity has been assumed and the number of fins, n_f, is

$$n_f = \frac{W_{hs}}{s + t}$$

Minimal thermal reduction results from excessively increasing base length, H, as shown in figure 5.6b. The maximum distance between fins, s, depends on the fin depth, L, and width of the fins, t, with deep finned heat sinks needing more space between adjacent fins than a shallow design, unless fan cooling is used. The minimum spacing, s is determined by fin depth, L, and airflow. If the fins are packed too closely, the flow through them is significantly reduced and therefore the heat transfer coefficient, h, decreases. The deeper the fins, L, the more space needed between them since a portion of the heat is radiated to adjacent fins, which helps to stabilise the temperature, but does little to dispose of the heat (in figure 5.4a, about 30% of the heat is radiation transferred fin-to-fin, hence not all dissipated).

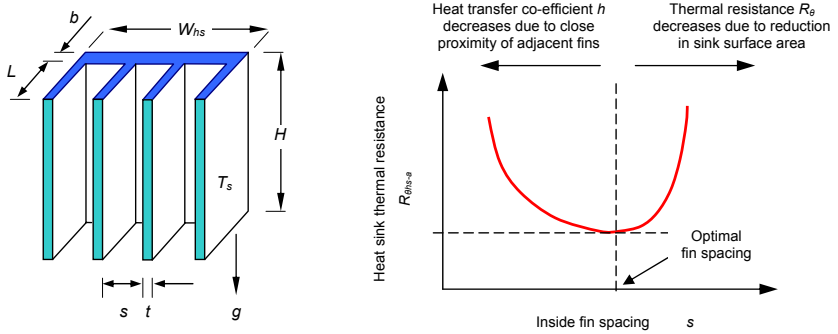


Figure 5.4. *Heat sink dimension parameters and thermal resistance dependence on fin spacing.*

As the base flow height H is increased, the air at the top of a vertical heatsink is hotter than that entering at the bottom. If the fin depth L is increased, there is more mutual radiation between fins, and as the spacing is reduced, mutual radiation increases further. Airflow is also restricted because of the smaller physical area for air to pass, since more of the available space is occupied by the heatsink itself.
The performance of a heatsink is linearly proportional to the base width W_{hs} of the sink in the direction perpendicular to the flow and proportional to the square root of the fin base length H in the direction of the airflow. (The heat transfer coefficient h is inversely related to H). Therefore it is better to increase the width rather than the length, provided the width is not already excessive compared to the length.

Heat transfer coefficient h can be defined in a number of ways. If it is defined referenced to the inlet fluid temperature of the heatsink, the heatsink thermal resistance is calculated by

$$R_{\theta hs-a} = \frac{1}{\eta_f hA} \qquad (5.21)$$

where A is the total surface area of fins and base between fins and
η_f is the fin efficiency, defined as

$$\eta_f = \frac{\tanh(m \times H)}{m_r \times H} \quad \text{and} \quad m_r = \sqrt{\frac{h \times P_\square}{\lambda \times A_x}} \qquad (5.22)$$

where H is the base height of the fin, m
$P_\square$ is the fin perimeter, m
A_x is the fin cross sectional area, m^2
m_f is the mass flow rate, equal to $\rho_f \times V_f \times s \times L$, kg/s
ρ_f fluid density (= $1/\upsilon$ specific volume), kg/m^3
V_f is the velocity between the fins
If the heat transfer coefficient is defined based on the temperature between the fins, the thermal resistance expression involves a heat capacitance component:

$$R_{\theta hs-a} = \frac{1}{\eta hA} + \frac{1}{2m_r c_p} \qquad (5.23)$$

where c_p is the fluid specific heat capacitance at constant pressure, kJ/kg.K.

Estimating radiation heat transfer from an extruded heat sink
The effect of radiation heat transfer (hence emissivity, ε) is important in natural convection, as it can be responsible for up to 40% of the total heat dissipation. Unless the heatsink is facing a hotter surface nearby, it is imperative to have the heat sink surfaces thinly painted or correctly anodised to enhance radiation. In natural convection situations where the convective heat transfer coefficient is relatively low, based on the dimensional parameters in figure 5.4a, the radiation heat transfer from all surfaces of the extruded heat sink can be calculated using

$$R_{\theta hs-a} = \frac{1}{\left\{(n_f - 1)\varepsilon_a S + \varepsilon_a \left[n_f t + 2(L + B)\right]\right\} H\sigma (T_s + T_A)(T_s^2 + T_A^2)} \qquad (5.24)$$

where n_f is the number of fins
ε_a is the apparent emissivity of a channel
T_s is the heatsink surface temperature and
T_A is the ambient temperature

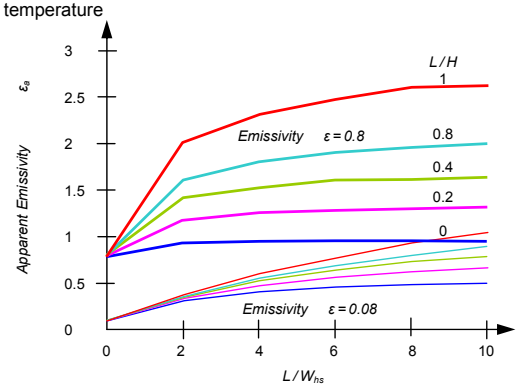


Figure 5.5. *Apparent emissivity ε_a of a channel heatsink of two different surface emissivities for different number of fins and dimensions.*

The apparent emissivity is a function of heat sink dimensions and surface emittances of the sink material, as shown in figure 5.5, for two values of the surface emissivity, namely ε equals 0.08 and 0.8. The apparent emissivity ε_a is based on enclosure theory and assumes a diffuse grey surface and constant surface temperature.

The emissivity coefficient, ε, indicates the radiation of heat from a body according the Stefan-Boltzmann Law, compared with the heat radiation from an ideal black body where the emissivity coefficient is $\varepsilon = 1$. Regardless of the composition of the emitting surface, the microscopic (and macroscopic) roughness of the surface causes differences in emissivity because a rougher surface has a larger emitting area. Generally, the emissivity of most opaque emitting surfaces increases as wavelength becomes shorter. The emissivity coefficient, ε, for some common surface qualities of aluminium and copper can be found in the Table 5.6 and in Appendix 5.25.
The low emissivity coefficients of untreated, polished aluminium and copper means they have surface finishes that limit the radiated heat from a body. Two thin coats of flat white Krylon #1502 (or equivalent) which has an emissivity of 0.96, should be used on all untreated (emissivity-wise) areas.

Unless otherwise stated, the heat sink is assumed anodised black (emissivity of up to 0.97) and vertically mounted with negligible thermal resistance from case to sink. In accordance with the data in Table 5.4, a general derating of 10 to 15 per cent for a bright surface and 15 to 20 per cent in the case of a horizontal mounting position, are usually adopted. Figure 5.6b also shows the improvement effects on dissipation due to the high thermal conductivity (heat spreader effect) of oxidised copper.

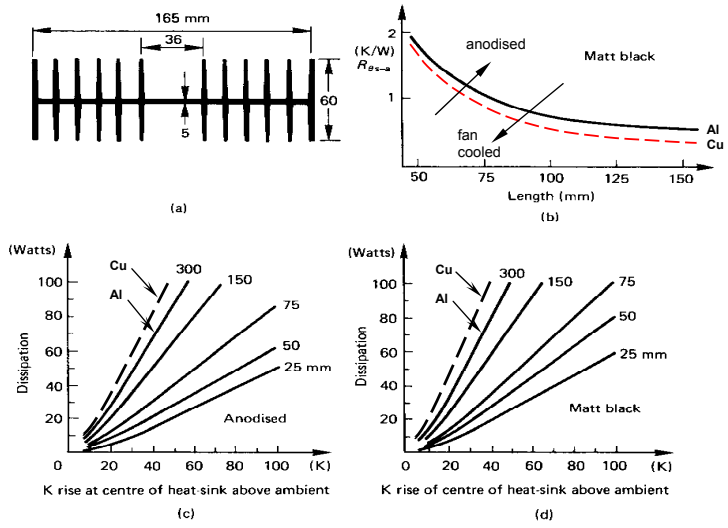


Figure 5.6. *Heat-sink typical data (for aluminium and copper):*
(a) cross-section view; (b) heat-sink length versus thermal resistance for a matt black surface finish;
(c) temperature rise versus dissipation for an anodised finish and different lengths; and
(d) as for (c) but with a matt black surface finish.

Thermal resistance increases with altitude, *z*, above sea level, as air density decreases, according to $R_y(z) = R_{o\,metres} / (1 - 5 \times 10^{-5} z)$. For example: a 1°C/W heatsink degrades to 1.11°C/W at an altitude of 2,000 metres, or 1.18° C/W at 3,000 metres.

The effective sink thermal resistance can be significantly reduced by *forced air cooling*, as indicated in Table 5.5, figure 5.7 and by equations (5.7) and (5.8). If the airflow is

- laminar, heat loss is proportional to the square root of air velocity, equation (5.7);
- turbulent, heat loss is proportional to velocity to the power of 0.8, equation (5.8).

When heatsinks (dissipating a total power of P_{Dtotal}) are vertically stack to share the same vertical natural convention airflow, the air temperature of the flow at the upper heatsink, after passing *n-1* heatsinks, is

$$T_{air} = T_{amb} + \frac{n-1}{n} \frac{c\,v}{P_{D\,total}} \qquad (5.25)$$

The chimney effect results in an airflow velocity *v*, which increases further up the heatsink stack. This and the air density increase results in the upper heatsink being the coolest, even though the passing air is the warmest.

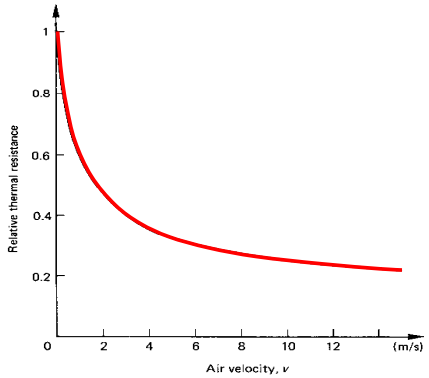


Figure 5.7. *Improved cooling with forced air cooled heat-sink - relative thermal resistance improvement with surface airflow.*

5.4 Modes of power dissipation

For long, >1ms, high duty cycle pulses the peak junction temperature is nearly equal to the average junction temperature. Fortunately, in many applications, a calculation of the average junction temperature is sufficient and the concept of *thermal resistance* is valid.
Other applications, notably switches driving highly reactive loads, may create severe current-crowding conditions which render the traditional concepts of thermal design invalid. In these cases, transistor safe operating area or thyristor *di/dt* limits must be observed, as applicable.
In yet other applications, heat cycling can cause power module faults, hence device failure, due to

- thermal cycling – is associated with large base plate (case) temperature changes, ΔT_c
- power cycling – is associated with large junction temperature changes, ΔT_j

The die is connected to a low thermal impedance substrate, usually utilising copper in the form of so-called *direct copper bonding*, DCB, as shown in figure 5.8a and the forced water cooling effectiveness is shown in figure 5.8c.

Direct copper bonding
Direct copper bonding DCB is a process in which copper (on each side) and a ceramic material, usually either aluminium oxide or aluminium nitride, are fused together at high temperature. By avoiding a thick copper base, the base is thinner and lighter, with lower thermal resistance and much better thermal cycling capabilities. (See Appendix 5.27)
The properties of DCB substrates are

- High mechanical strength and stability
- Good cohesion and corrosion resistance
- High electrical insulation
- Excellent thermal conductivity
- Reliable thermal cycling stability
- Matched thermal expansion coefficient to silicon and gallium arsenide
- Good heat spreading
- Processable, e.g. copper is etchable and millable like a pcb
- Environmentally friendly
- High copper current density

The advantages of DBC substrates are high current carrying capability due to thick copper metallization and a thermal expansion close to the silicon at the copper surface due to high bond strength of copper to ceramic. Normally, DCB has two layers of copper that are directly bonded onto an aluminium-oxide ($A\ell_2O_3$) or aluminium-nitride ($A\ell N$) ceramic base. The DCB process yields a super-thin base and eliminates the need for the thick, heavy copper bases that were used prior to this process. Because power modules with DCB bases have fewer layers, they have a much lower thermal resistance. Because the expansion coefficient matches silicon, they have much better power cycling capabilities (up to 50,000 cycles).

The drawback of standard DCB substrates for high voltage applications is a start of partial discharge at relatively low voltages. Therefore substrates using expensive metal brazing technologies (AMB) are mainly used in high voltage semiconductor modules for traction applications. The initiation voltage for ceramic thickness of 0.63mm is less than 4kV. Main causes for this behaviour are small voids between the copper and ceramic and blurred straight border lines of the copper conductors at the copper/ceramic interface. Precision etching technology can alleviate these disadvantages. The other disadvantage of DCB is its deficiency for thermal shock because of the large residual stress on the substrate surface due to the coefficient of thermal expansion CTE mismatch of alumina and copper.

Thermal cycling

Intermittent equipment operation, start-up, and shutdown in extreme temperature conditions may cause power module thermal stresses due to the different linear expansion temperature co-efficients of the materials associated with the soldered substrate mounting to the copper base plate in multi-chip large area packages (see Tables 5.30 and 5.31 in Appendix 5.24). Large base plate (case) temperature changes in excess of 80K over a few minutes, stress the hard solder bonding between the copper base plate and the insulating substrate (usually $A\ell N$ or $A\ell_2 0_3$), as shown in figure 5.8a. This fatigue leads to eventual crack failure after a finite number of cycles N, approximated by

$$N = \frac{k}{A \times \Delta T^2} \qquad (5.26)$$

where A is the die area and ΔT is the thermal shock temperature change. The constant k depends on the package, type of hard soldering, etc. Large, multiple die IGBT modules suffer from thermal shock limitations and relatively low reliability, because of the sheer large number of die soldered to the substrate over a large base plate copper area in the module.

Figure 5.8b shows how the number of thermal cycles to fracture for DCB substrates varies with copper thickness, when cycled between -40°C to +110°C. For a case temperature change of $\Delta T = 80$K, lifetime can be as low as 3,500 cycles and may only involve powering up and shutting down the associated equipment. Thermal cycling is normally performed by cycling the inactive package between the maximum and minimum storage temperatures. Although Aℓ/SiC is far superior to copper from a differential thermal expansion perspective, its thermal conductivity is only a little better than that of aluminium. Floating silicon wafers in disc type packages suffer to a much lesser extent (an order) from the effects of differential thermal expansion when thermally cycled.

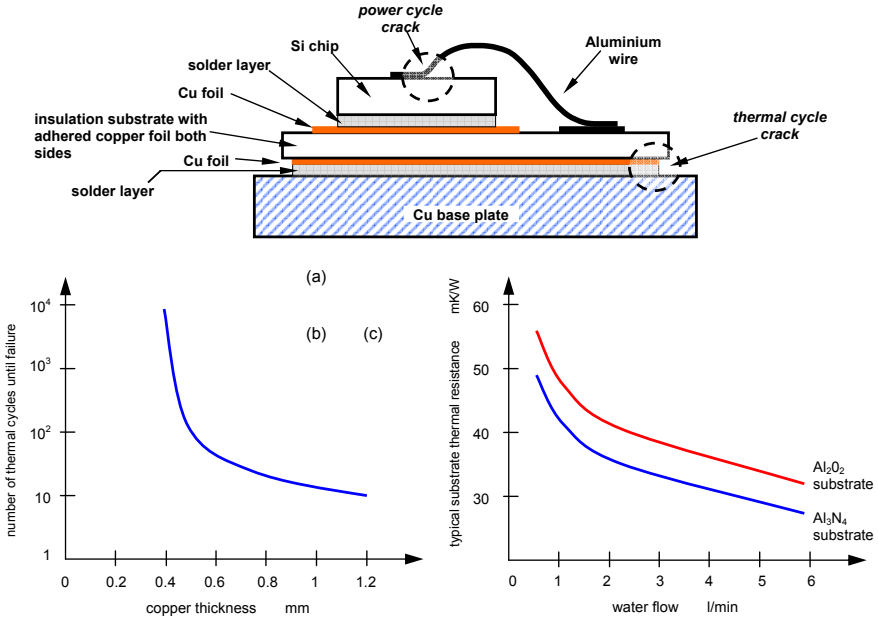


Figure 5.8. *Direct copper bonding:*
(a) sectional view of power module substrate showing boundary regions where power cycle cracking and thermal cycle cracking, occur; (b) copper thickness affect on power failure; and (c) thermal resistance dependence on liquid cooling flow rate and substrate material.

Power cycling

Rapid cycling of the chip junction temperature causes mechanical stress around the silicon chip to aluminium wire bond interface, due to their different linear expansion temperature co-efficients. Eventually a crack occurs on the silicon side of the interface, as indicated in figure 5.8a. Short rapid junction temperature changes, over tens of seconds, of ΔT_j =100K, can lead to failure within 2500 cycles. The number of cycles to failure increases by just over an order for every 10°C decrease in ΔT_j.

In a related thermal application, where the power dissipated in the semiconductor consists of pulses at a low duty cycle, the instantaneous or peak junction temperature, not average temperature, may be the limiting condition. Figure 5.9 shows by comparison such a condition, where the operating frequency, not the maximum power dissipated, is dominant in determining junction temperature. In this case *thermal impedance* $Z_{\theta\,j\text{-}c}$ is used instead of thermal resistance $R_{\theta\,j\text{-}c}$ such that $Z_{\theta\,j\text{-}c} = r(t_p)\ R_{\theta\,j\text{-}c}$, where $r(t_p)$ is the normalising factor yielded from the normalised transient thermal impedance curves for the particular device. Appropriate values for the rectangular power pulse width t_p and duty cycle factor δ are used.

The power devices employed in power electronics are usually used in some form of on/off power pulse waveform mode. The following power waveforms are analysed

- Periodic rectangular power pulses (steady-state thermal response);
- Single rectangular power pulse;
- Composite rectangular superimposed power pulses;
- A burst of rectangular power pulses; and
- Non-rectangular power pulses.

5.4.1 Steady-state response

Large cycle-by-cycle junction temperature fluctuations occur at low frequencies. As frequency increases, thermal inertia of the junction smoothes out instantaneous temperature fluctuations, as shown in figure 5.9b, and the junction responds more to average, rather than peak power dissipation. At frequencies above a kilohertz and duty cycles above 20 per cent, cycle-by-cycle temperature fluctuations usually become small, and the peak junction temperature rise approaches the average power dissipation multiplied by the steady-state junction-to-case thermal resistance, within a few per cent.

Because of thermal inertia (long thermal time constant), the heat sink and package case respond only to average power dissipation, except at ultra low frequencies, < 1Hz. The steady-state thermal conditions for the case-mount and junction (equation (5.12)) are given by

$$P_d = \frac{\hat{T}_j - T_c}{R_{\theta j\text{-}c}} = \frac{T_c - T_a}{R_{\theta c\text{-}s} + R_{\theta s\text{-}a}} \quad \text{(W)} \qquad (5.27)$$

where P_d is the average power dissipation, which is the maximum power multiplied by the on-time duty cycle δ for rectangular power pulses. The difficulty in applying equation (5.27) often lies in determining the average power dissipation.

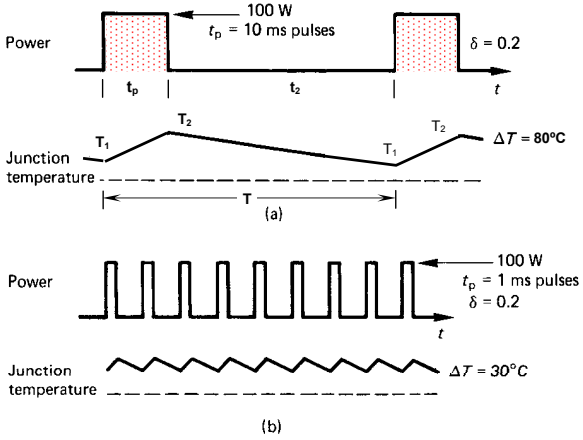


Figure 5.9. *Waveforms illustrating that peak junction temperature is a function of switching frequency: (a) lower switching frequency with 10 ms pulse and a 20 per cent duty cycle and (b) high frequency and 1 ms pulse with a duty cycle the same as in (a).*

5.4.2 Pulse response

When a junction dissipates power associated with a single pulse, the junction temperature increases during the pulse and decays to the original temperature after the energy pulse ceases. The junction temperature variation may vary from an ambient temperature to a level above the normal maximum operating limit, a change of over 150°C. The upper temperature due to the power pulse can cause silicon damage, if the maximum allowable limit is exceeded too often or by a large amount on just one occasion.

Equation (5.2) is valid for one dimensional steady state and transient thermal conditions, and the transient temperature equation is given by the first order solution to

$$P_d = -\frac{\lambda A}{\ell} T + \gamma A \ell \frac{\delta T}{\delta t} \qquad \text{(W)} \tag{5.28}$$

The time domain solution for the temperature rise is

$$\Delta T(t) = \Delta \hat{T} \times \left(1 - e^{-t/\tau}\right) \tag{5.29}$$

where the maximum temperature eventually attained if the power pulse were maintained, above ambient, is

$$\Delta \hat{T} = \frac{P_d \ell}{\lambda A} = \frac{P_d}{h A} = P_d R_\theta \qquad \text{(K)} \tag{5.30}$$

and the thermal time constant

$$\tau = \frac{\gamma \ell^2}{\lambda} = \frac{\text{thermal capacity, J/K}}{\text{power per K, W/K}} \qquad \text{(s)} \tag{5.31}$$

The transient thermal impedance Z_θ is defined as

$$Z_\theta = r(t_p) R_\theta = \frac{\Delta T}{P_d} = \frac{\Delta \hat{T} \times \left(1 - e^{-t_p/\tau}\right)}{\Delta \hat{T} / R_\theta} = \left(1 - e^{-t_p/\tau}\right) R_\theta \tag{5.32}$$

That is, thermal resistance R_θ is modified by the factor $r(t_p)$ to yield transient thermal impedance Z_θ:

$$r(t_p) = \left(1 - e^{-t_p/\tau}\right) \tag{5.33}$$

This one-dimensional solution assumes a homogeneous thermal conducting material with a single point heat source, producing a uniform heat flow path. Since the practical case is far from ideal, manufacturers provide data for dynamic temperature effects based on a concept termed *thermal impedance*. The thermal solution given by equation (5.29) gives acceptable results when applied to solid carbon resistors (being a homogeneous material), as considered in Chapter 25 (specifically, see Example 25.7).

Example 5.1: ***Semiconductor single power pulse capability***

A semiconductor has a thermal capacity (*mc*) of 0.1J/K and a steady state thermal resistance to its case of R_θ = 0.5 K/W. If the junction temperature is not to exceed 125°C in a 25°C ambient, determine the allowable power dissipation, hence transient thermal impedance, as a function of single power pulse duration. Plot the results for five time decades, decreasing from 1s.

Solution

The power dissipation per K is

$$P_d / \text{K} = \frac{1}{R_\theta} = \frac{1}{0.5\,\text{K/W}} = 2\,\text{W/K}$$

From equation (5.31) the thermal time constant τ is given by

$$\tau = \frac{\text{thermal capacity, J/K}}{\text{power per K, W/K}} = \frac{0.1\,\text{J/K}}{2\,\text{W/K}} = 0.05\text{s}$$

After time t_p, the junction temperature rise from 25°C must not exceed 125°C, that is $\Delta T(t_p)$ = 100K, thus equation (5.29) gives

$$\Delta T(t_p) = \Delta \hat{T} \times \left(1 - e^{-t/\tau}\right) = \Delta \hat{T} \times \left(1 - e^{-t_p/0.05\text{s}}\right) = 100\text{K}$$

As a specific example of the procedure, consider a t_p = 10ms energy pulse.

$$\Delta T(10\text{ms}) = \Delta \hat{T} \times \left(1 - e^{-10\text{ms}/0.05\text{s}}\right) = 100\text{K}$$

which yields $\Delta \hat{T}$ = 551.6K. That is, after a long period (>>10ms) the junction temperature would increase by 551.6K. From equation (5.30), this temperature rise corresponds to continuous power of

$$P_d = \frac{\Delta \hat{T}}{R_\theta} = \frac{551.6\,\text{K}}{0.5\,\text{K/W}} = 1103.3\,\text{W}$$

In 10ms the temperature must only rise 100K, hence, from equation (5.32) the transient thermal impedance Z_θ is

$$Z_\theta = r(t_p) R_\theta = \frac{\Delta T}{P_d} = \frac{100\,\text{K}}{1103.3\,\text{W}} = 0.091\,\text{K/W}$$

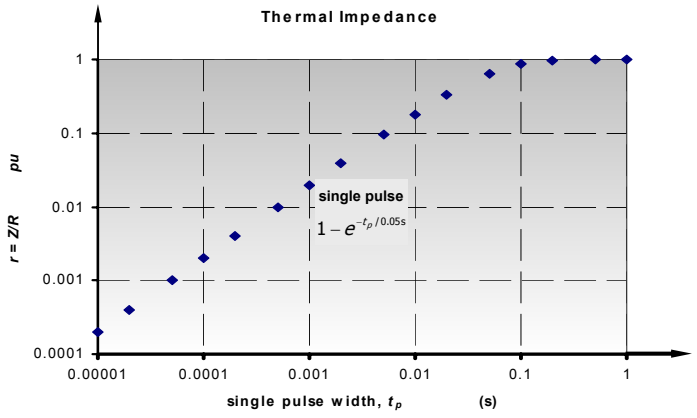
Thus the thermal resistance R_θ is modified, or normalized, by

$$r(10\text{ms}) = \frac{Z_\theta}{R_\theta} = \frac{0.0001\,\text{K/W}}{0.5\,\text{K/W}} = 0.181$$

Table 5.7 shows the normalised thermal impedance factor, $r(t_p)$, for other pulse durations, which are plotted in the accompanying figure. Notice the similarity of the single pulse results given for a practical power device in figure 5.10.

Table 5.7: Single pulse data

t_p	$\Delta \hat{T}$ $t \to \infty$	P_d $= \Delta \hat{T} / R_\theta$	Z_θ $= \Delta T / P_d$	$r(t_p)$ $= Z_\theta / R_\theta$
pulse time	temperature rise	power dissipated	thermal impedance	normalised
s	K	W	K/W	pu
1	100	200	0.5	1
0.1	116	231	0.432	0.86
0.01	552	1103	0.091	0.181
0.001	5050	10100	0.0099	0.0198
0.0001	50050	100100	0.0010	0.0020
0.00001	500050	1000100	0.0001	0.0002



♣

Figure 5.10 shows the thermal impedance curves for a power-switching device, normalised with respect to the steady-state thermal resistance $R_{\theta\,\text{j-c}}$. The curve labelled 'single pulse' shows the rise of junction temperature per watt of power dissipated as a function of pulse duration. The thermal impedance for repetitive pulses *Z*, of duty cycle δ, can be determined from the single pulse value *z* according to

$$Z_\theta(t_p, \delta) = \delta + (1 - \delta) z(t_p) \qquad \text{(K/W)} \tag{5.34}$$

The equation (5.12) becomes

$$P_p = \frac{\hat{T}_j - T_c}{Z_\theta(t_p, \delta)} = \frac{\hat{T}_j - T_c}{r(t_p) R_{\theta j-c}} \quad \text{(W)} \qquad (5.35)$$

Note that the peak power P_p is employed, and then only for thermal analysis from the junction to the case thermal mounting. That is, $Z_{\theta j\text{-}c}$ is the only thermal impedance term that exists. See problem 5.8.

Figure 5.10 shows that at the pulse width minimum extreme, t_p < 1μs, as $z(t_p \to 0) \to 1$ in equation (5.34):

$$\lim_{t_p \to 0} Z_\theta(0, \delta) = \delta R_{\theta j-c} = r(t_p = 0) R_{\theta j-c} \qquad (5.36)$$

that is, $r(t_p \to 0) \to \delta$.

Figure 5.10 also shows that at the pulse maximum extreme, that is, t_p > 1s or continuous power dissipation, as $z(t_p \to 1) \to 0$ in equation (5.34):

$$\lim_{t_p \to \infty} Z_\theta(\infty, \delta) = R_{\theta j-c} \qquad (5.37)$$

that is, $r(t_p \to \infty) \to 1$, independent of duty cycle, that is, for all duty cycles.

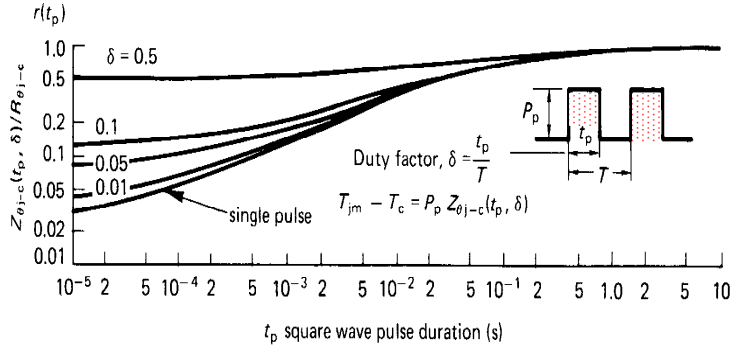


Figure 5.10. *Transient thermal impedance curves; normalised with respect to the steady state thermal resistance, $R_{\theta j\text{-}c}$.*

Example 5.2: ***A single rectangular power pulse***

A semiconductor with a junction to case thermal resistance of 1 K/W absorbs a single 100W power pulse for 20μs. Based on the thermal impedance characteristics in figure 5.10, what is the expected junction temperature rise, assuming the case-mount temperature does not respond to this short pulse?

Solution

The period for a single power pulse is infinite $T \to \infty$, therefore the duty cycle is zero, δ=0.

$$\Delta T_{j-c} = P \times Z_{\theta j-c} = P \times r(t_p) \times R_{\theta j-c}$$

From figure 5.10, for a single 20μs pulse $r(t_p$ = 20μs$)$ = 0.035. The junction temperature change is therefore

$$\Delta T_{j-c} = P \times r(t_p = 20\mu s) \times R_{\theta j-c}$$
$$= 100\text{W} \times 0.035 \times 1\text{K/W} = 3.5\text{K}$$

The peak junction temperature will rise to 3.5K above the case mount temperature at the end of the 100W rectangular power pulse.

♣

The basic single rectangular power pulse approach can be extended to analyse composite rectangular power pulses by algebraic superposition of a series of accumulating rectangular pulses. Because each composite power pulse extends to the end of the temperature-calculating period, any positive rectangular pulse is subsequently cancelled by a negative power pulse. The technique is illustrated in example 5.4.

5.4.3 Repetitive transient response

Minimal temperature variation occurs if the power switching period T is shorter than the junction to case mount thermal time constant, $T < \tau$, whence the concept of steady state thermal resistance is applicable, as presented in 5.4.1. When the relative magnitudes are reversed such that $T > 5\,\tau$, then the temperature effects of the power pulse die away, and the single pulse transient thermal impedance approach presented in 5.4.2 is applicable.

The transition or boundary between junction operation that can be assumed steady state junction temperature operation ($T < \tau$) and that of a series of discrete non-interacting single pulses ($T > 5\,\tau$) can be analysed by extending the one-dimensional thermal transient equation (5.29) in conjunction with figure 5.9a. Figure 5.9a shows how the temperature increases from T_1 to T_2 during the time t_p when power is dissipated, and decreases from T_2 to T_1 during time t_2 when no power is being dissipated by the virtual junction. This increasing and decreasing of the junction temperature occurs cyclically over each period T.

Based on equation (5.29) the junction temperature increases exponentially according to

$$T(t) = \Delta\hat{T} - \left(\Delta\hat{T} - T_1\right) e^{-t/\tau} \qquad (5.38)$$

and decreases exponentially according to

$$T(t) = T_2\, e^{-t/\tau} \qquad (5.39)$$

where the thermal time constant τ and maximum possible junction temperature rise are defined by equations (5.31) and (5.30), respectively. Since these temperature variations are in steady state, the temperature constants T_1 to T_2 can be solve using the boundary conditions. This gives

$$T_2 = \Delta\hat{T}\,\frac{1 - e^{-t_p/\tau}}{1 + e^{-T/\tau}} \quad \text{and} \quad T_1 = T_2\, e^{-t_2/\tau} \qquad (5.40)$$

The junction temperature swing, ΔT is

$$\Delta T_j = T_2 - T_1 = \Delta\hat{T}\,\frac{\left(1 - e^{-t_p/\tau}\right)\left(1 - e^{-t_2/\tau}\right)}{1 + e^{-T/\tau}} \qquad (5.41)$$

The maximum variation in junction temperature occurs for square wave power, that is $t_p = t_2 = ½T$, δ=½:

$$\Delta T_j^{\max} = \Delta\hat{T} \tanh\left(\frac{T}{4\tau}\right) \qquad (5.42)$$

This equation highlights that the magnitude of the temperature change is highly dependant on the power switching frequency $1/T$ relative to the thermal time constant τ of the semiconductor package.

Example 5.3: ***Semiconductor transient repetitive power capability***

A semiconductor with a thermal capacity of 0.02J/K and a thermal resistance from the junction to the case of ½K/W, dissipates 100W at a repetition rate of

i. 50Hz
ii. 300Hz.

By calculating the worst-case junction temperature variation, indicate whether steady-state constant junction temperature-based analysis (a thermal resistance approach) is a valid assumption.

Solution

The long-term junction temperature rise with 100W continuous is given by equation (5.30), which yields

$$\Delta\hat{T} = P_d\, R_\theta = 100\text{W} \times ½\text{K/W} = 50\text{K}$$

The thermal time constant τ is given by equation (5.31), giving

$$\tau = \frac{\text{thermal capacity, J/K}}{\text{power per K, W/K}} = \frac{0.02}{1/½} = 0.01 \quad (\text{s})$$

Worse case temperature variation occurs with a 50% power duty cycle, as given by equation (5.42)

$$\Delta T_j^{\max} = \Delta\hat{T} \tanh\left(\frac{T}{4\tau}\right) = 50\text{K} \times \tanh\left(\frac{T}{4 \times 0.01\text{s}}\right)$$

From this equation:

at 50Hz, T = 20ms, $\Delta T_j^{\max}$ = 23.1K

at 300Hz, T = 3.33ms, $\Delta T_j^{\max}$ = 4.1K

The temperature variation of 4.1K at 300Hz is small compared to the maximum allowable junction temperature, typical between 125°C and 175ºC, thus thermal analysis of this device in a 300Hz application, can be thermal resistance calculation based as presented in 5.4.1. On the other hand, the same device used in a 50Hz application will experience 5.6 times the junction temperature swing. This 23.1K variation represents a significant portion of the allowable junction operating temperature, and could mean a thermal resistance approach is unsafe. A thermal impedance design approach is recommended, as in example 5.2 and 5.4.2.

♣

Example 5.4: ***Composite rectangular power pulses***

A semiconductor with a junction to case thermal resistance of 1 K/W absorbs the composite power pulse shown in figure 5.11. Based on the thermal impedance characteristics in figure 5.12, what is the expected junction temperature rise at indicate times t_x and t_y, assuming the case temperature does not respond to this short pulse? That is, the heatsink-case interface temperature is held constant.
What is the average junction to case temperature rise, in the repetitive case?

Solution

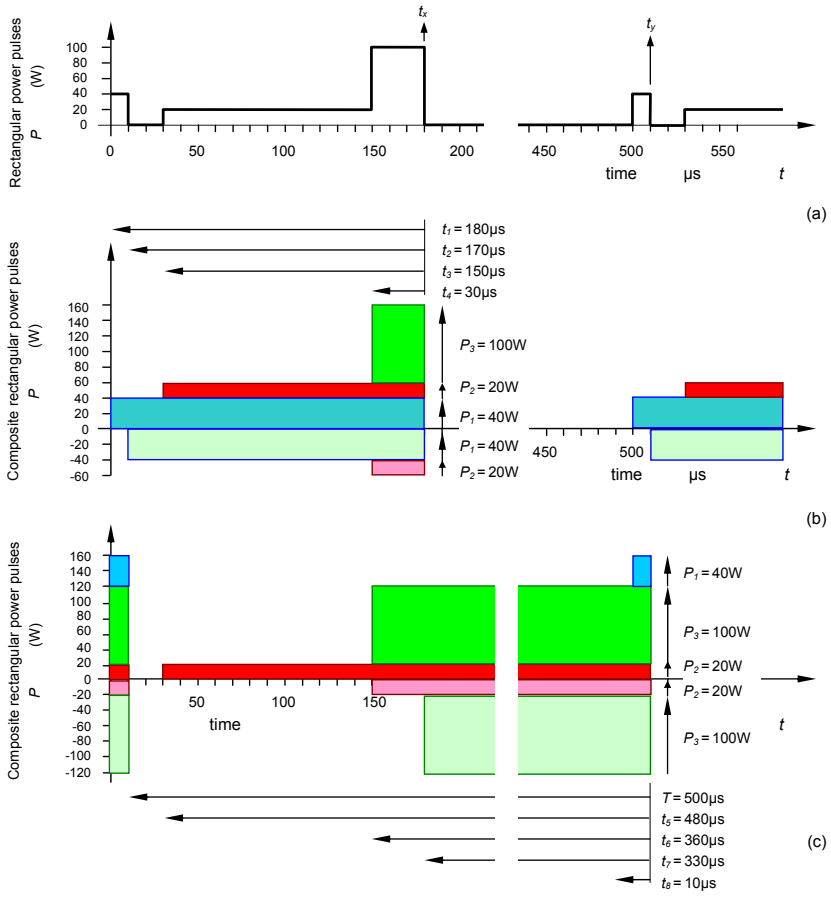


Figure 5.11. *Composite power pulses: (a) original rectangular pulse; (b) composite rectangular pulse, reference t_x; and (c) composite rectangular pulse, reference t_y.*

Table 5.8: Rectangular, composite pulse example data

		t_x				t_y			
Pulse duration		t_{p1}	t_{p2}	t_{p3}	t_{p4}	t_{p5}	t_{p6}	t_{p7}	t_{p8}
		180μs	170μs	150μs	30μs	480μs	360μs	330μs	10μs
single pulse	$\delta = t_p / T$	0	0	0	0	0	0	0	0
$T \rightarrow \infty$	$r(t_p)$	0.06	0.055	0.05	0.025	0.10	0.085	0.08	0.015
period	$\delta = t_p / T$	0.36	0.34	0.30	0.06	0.96	0.72	0.66	0.02
T = 500μs	$r(t_p)$	0.38	0.36	0.35	0.075	0.96	0.73	0.73	0.035

In figure 5.11b, power pulse P_1 = 40W lasts for 180μs, which represents a duty cycle of δ =180μs/500μs = 0.36. The thermal impedance normalised factor of $r(t_{p1}$=80μs) = 0.38 corresponds to δ = 0.45 in figure 5.12.

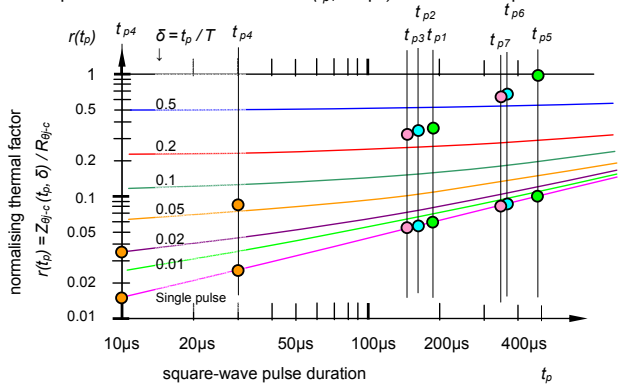


Figure 5.12. *Normalise junction to case-mount thermal impedance characteristics.*

The junction temperature (rise) at t_x, is given by

$$T_{j-c}^{t_x} = P_1 Z_{\theta j-c}^{tp1} - P_1 Z_{\theta j-c}^{tp2} + P_2 Z_{\theta j-c}^{tp3} - P_2 Z_{\theta j-c}^{tp4} + P_3 Z_{\theta j-c}^{tp4}$$

The junction temperature (rise) at t_y, is given by

$$T_{j-c}^{t_y} = P_2 Z_{\theta j-c}^{tp5} - P_2 Z_{\theta j-c}^{tp6} + P_3 Z_{\theta j-c}^{tp6} - P_3 Z_{\theta j-c}^{tp7} + P_1 Z_{\theta j-c}^{tp8}$$

t_x repetitive

$$T_{j-c}^{t_x} = 40 \times Z_{\theta j-c}^{tp1} - 40 \times Z_{\theta j-c}^{tp2} + 20 \times Z_{\theta j-c}^{tp3} - 20 \times Z_{\theta j-c}^{tp4} + 100 \times Z_{\theta j-c}^{tp4}$$
$$= 40 \times 0.38 \times 1\text{K/W} - 40 \times 0.36 \times 1\text{K/W} + 20 \times 0.35 \times 1\text{K/W} - 20 \times 0.075 \times 1\text{K/W} + 100 \times 0.075 \times 1\text{K/W}$$
$$= 13.8\text{K}$$

t_x single pulse

$$T_{j-c}^{t_x} = 40 \times Z_{\theta j-c}^{tp1} - 40 \times Z_{\theta j-c}^{tp2} + 20 \times Z_{\theta j-c}^{tp3} - 20 \times Z_{\theta j-c}^{tp4} + 100 \times Z_{\theta j-c}^{tp4}$$
$$= 40 \times 0.06 \times 1\text{K/W} - 40 \times 0.055 \times 1\text{K/W} + 20 \times 0.05 \times 1\text{K/W} - 20 \times 0.025 \times 1\text{K/W} + 100 \times 0.025 \times 1\text{K/W}$$
$$= 2.7\text{K}$$

t_y repetitive

$$T_{j-c}^{t_y} = 20 \times Z_{\theta j-c}^{tp5} - 20 \times Z_{\theta j-c}^{tp6} + 100 \times Z_{\theta j-c}^{tp6} - 100 \times Z_{\theta j-c}^{tp7} + 100 \times Z_{\theta j-c}^{tp8}$$
$$= 20 \times 0.96 \times 1\text{K/W} - 20 \times 0.73 \times 1\text{K/W} + 100 \times 0.73 \times 1\text{K/W} - 100 \times 0.67 \times 1\text{K/W} + 40 \times 0.035 \times 1\text{K/W}$$
$$= 12.2\text{K}$$

t_y single pulse

$$T_{j-c}^{t_y} = 20\times Z_{\theta j-c}^{tp5} - 20\times Z_{\theta j-c}^{tp6} + 100\times Z_{\theta j-c}^{tp6} - 100\times Z_{\theta j-c}^{tp7} + 100\times Z_{\theta j-c}^{tp8}$$
$$= 20\times 0.10\times 1\text{K/W} - 20\times 0.085\times 1\text{K/W} + 100\times 0.085\times 1\text{K/W} - 100\times 0.08\times 1\text{K/W} + 40\times 0.015\times 1\text{K/W}$$
$$= 1.4\text{K}$$

In the repetitive composite pulse case, the average power dissipated over 500µs is

$$\frac{10\mu s\times 40W + 120\mu s\times 20W + 30\mu s\times 100W}{500\mu s} = 11.6W$$

The average junction to case mounting temperature rise is

$$T_{j-c} = P_{ave}\times R_{\theta j-c} = 11.6W\times 1\text{K/W} = 11.6\text{K}$$

♣

Non-rectangular power pulses

The concept and characterisation of thermal impedance is based on rectangular power pulses. Non-rectangular pulses are converted to equivalent energy, rectangular pulses having the same peak power, P_p, of period t_p, as shown in figure 5.13. The resultant rectangular power pulse will raise the junction temperature higher than any other wave shape with the same peak and average values, since it concentrates its heating effects into a shorter period of time, thus minimising cooling during the pulse. Worse case semiconductor thermal conditions result. Improved thermal accuracy is obtained if each non-rectangular pulse is further sub-divided into numerous equivalent total energy rectangular pulses, as considered in example 5.5.

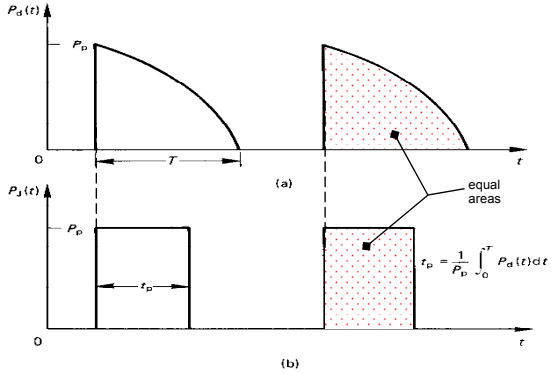


Figure 5.13. *Conversion of non-rectangular power pulse (a) into equivalent rectangular pulse (b).*

Example 5.5: ***Non-rectangular power pulses***

Switch losses are a series of triangular power pulses rising linearly to 100W in 100µs after switch turn-on. If the thermal resistance junction to case mounting is 1 K / W and the thermal impedance characteristics are represented by figure 5.15, calculate the case to junction peak temperature rise for

i. A single pulse, $T\rightarrow\infty$
ii. 50% power duty cycle, $T = 200\mu s$
iii. 10% duty cycle, $T = 100\mu s$

Represent the triangular power pulses by firstly one equivalent rectangular power pulse, secondly two equivalent rectangular power pulses, and compare the predicted peak junction temperature rise results. Assume the case temperature is maintained at a constant temperature.
Where applicable, calculate the average junction to case thermal mounting temperature, $T_{j\text{-}c}$.

Solution

Each saw-tooth power pulse is represented by a single rectangular power pulse, 100W and 50µs duration in figure 5.14, therein fulfilling the requirements of the same maximum power occurring simultaneously in both waveforms and both containing the same energy, area.

Figure 5.14 also shows the two rectangular pulse representations, where successive 50µs portions of the triangle are represent by pulses, 100W, 37.5µs and 50W, 25µs, such that the total area is maintained and the peak junction temperature rise occurs at the end of the power pulse sequence. The two pulses are subsequently decomposed into three equivalent composite rectangular power pulses, which sum at any time to give the original two rectangular pulses.

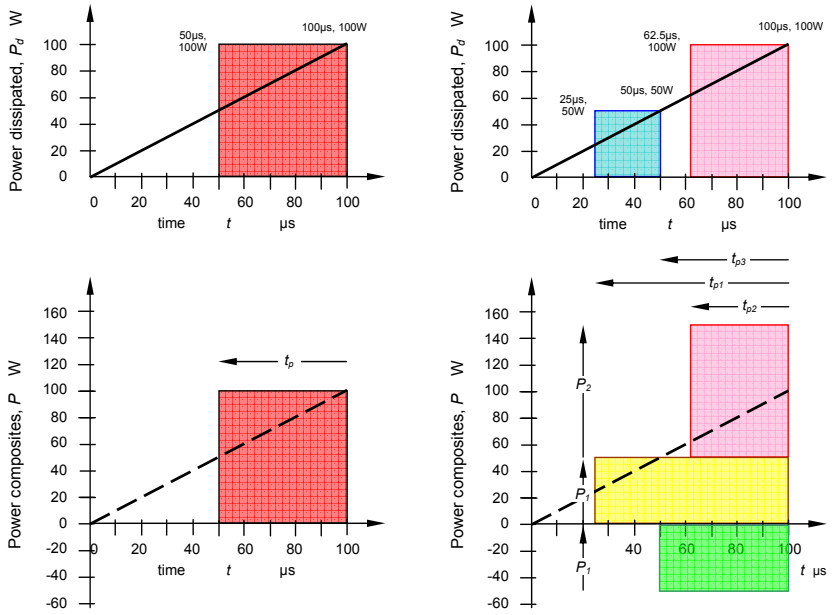


Figure 5.14. *Transient thermal impedance curves; normalised with respect to the steady state thermal resistance, $R_{\theta\ j\text{-}c}$.*

The thermal impedance normalising factor $r(t_p)$ for the applicable device can be read from figure 5.15, using the pulse periods and duty cycles shown in figure 5.14, and are shown in Table 5.9.

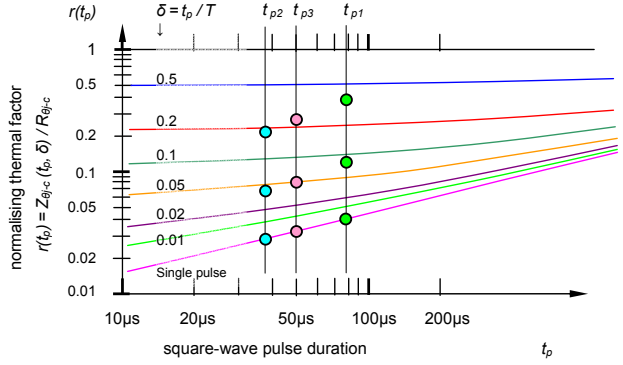


Figure 5.15. *Transient thermal impedance curves; normalised with respect to the steady state thermal resistance, $R_{\theta\ j\text{-}c}$.*

Table 5.9: Non-rectangular, composite pulse example data

		One composite power pulse	Two composite power pulses		
Pulse duration		t_p	t_{p1}	t_{p2}	t_{p3}
		50μs	75μs	37.5μs	50μs
Single pulse $T \rightarrow \infty$	$\delta = t_p/T$	0	0	0	0
	$r(t_p)$	0.045	0.04	0.040	0.045
50% duty cycle T = 200μs	$\delta = t_p/T$	¼	⅜	0.188	¼
	$r(t_p)$	0.32	0.40	0.22	0.32
10% duty cycle T = 1000μs	$\delta = t_p/T$	0.05	0.075	0.0375	0.05
	$r(t_p)$	0.08	0.12	0.066	0.08

For a single pulse rectangular power waveform

$$\Delta T_{j-c} = P \times Z_{\theta j-c} = P \times r(t_p) \times R_{\theta j-c}$$
$$= 100\text{W} \times r(t_p = 50\mu\text{s}) \times 1\text{K/W}$$

For a single pulse, $\delta = 0$

$$\Delta T_{j-c} = 100\text{W} \times r(t_p = 50\mu\text{s}) \times 1\text{K/W}$$
$$= 100\text{W} \times 0.045 \times 1\text{K/W} = 4.5\text{K}$$

For a 50% duty cycle, $\delta = 0.5$

$$\Delta T_{j-c} = 100\text{W} \times r(t_p = 50\mu\text{s}) \times 1\text{K/W}$$
$$= 100\text{W} \times 0.32 \times 1\text{K/W} = 32\text{K}$$

For a 10% duty cycle, $\delta = 0.1$

$$\Delta T_{j-c} = 100\text{W} \times r(t_p = 50\mu\text{s}) \times 1\text{K/W}$$
$$= 100\text{W} \times 0.08 \times 1\text{K/W} = 8\text{K}$$

For a two pulse rectangular power waveform representation

$$\Delta T_{j-c} = P_1 \times Z_{\theta j-c\ t1} - P_1 \times Z_{\theta j-c\ t3} + P_2 \times Z_{\theta j-c\ t2}$$
$$= P_1 \times r(t_{p1}) \times R_{\theta j-c\ t1} - P_1 \times r(t_{p3}) \times R_{\theta j-c\ t3} + P_2 \times r(t_{p2}) \times Z_{\theta j-c\ t2}$$
$$= 50\text{W} \times r(t_{p1} = 75\mu\text{s}) \times 1\text{K/W} - 50\text{W} \times r(t_{p3} = 50\mu\text{s}) \times 1\text{K/W} + 100\text{W} \times r(t_{p2} = 37.5\mu\text{s}) \times 1\text{K/W}$$

For a single pulse, $\delta = 0$

$$\Delta T_{j-c} = 50\text{W} \times r(t_{p1} = 75\mu\text{s}) \times 1\text{K/W} - 50\text{W} \times r(t_{p3} = 50\mu\text{s}) \times 1\text{K/W} + 100\text{W} \times r(t_{p2} = 37.5\mu\text{s}) \times 1\text{K/W}$$
$$= 50\text{W} \times 0.04 \times 1\text{K/W} - 50\text{W} \times 0.045 \times 1\text{K/W} + 100\text{W} \times 0.04 \times 1\text{K/W}$$
$$= 3.75\text{K}$$

For a 50% duty cycle

$$\Delta T_{j-c} = 50\text{W} \times r(t_{p1} = 75\mu\text{s}) \times 1\text{K/W} - 50\text{W} \times r(t_{p3} = 50\mu\text{s}) \times 1\text{K/W} + 100\text{W} \times r(t_{p2} = 37.5\mu\text{s}) \times 1\text{K/W}$$
$$= 50\text{W} \times 0.40 \times 1\text{K/W} - 50\text{W} \times 0.32 \times 1\text{K/W} + 100\text{W} \times 0.22 \times 1\text{K/W}$$
$$= 26\text{K}$$

For a 10% duty cycle

$$\Delta T_{j-c} = 50\text{W} \times r(t_{p1} = 75\mu\text{s}) \times 1\text{K/W} - 50\text{W} \times r(t_{p3} = 50\mu\text{s}) \times 1\text{K/W} + 100\text{W} \times r(t_{p2} = 37.5\mu\text{s}) \times 1\text{K/W}$$
$$= 50\text{W} \times 0.12 \times 1\text{K/W} - 50\text{W} \times 0.08 \times 1\text{K/W} + 100\text{W} \times 0.066 \times 1\text{K/W}$$
$$= 8.6\text{K}$$

The average junction to case temperature for a single pulse is zero, and the average junction temperature is the heatsink/ambient temperature.
The average junction to case temperature during repetitive operation is independent of whether one or two composite rectangular pulses are used to analyse the saw-tooth pulse pulses, since both model the same original power waveform, each having the same waveform area, energy. The junction to case temperature is dependant on the duty cycle, with specifies the average power dissipation.

$$\overline{P}_d = \frac{\text{sawtooth area}}{T}$$
$$= \frac{½ \times 100\text{W} \times 100\mu\text{s}}{T} = \delta \times 50\,\text{W}$$

Thus the average case to junction temperature is

$$T_{j-c} = \overline{P}_d \times R_{\theta j-c}$$
$$= \delta \times 50\text{W} \times 1\text{K/W} = 50 \times \delta\,\text{K}$$

For 50% and 10% duty cycles, this gives average temperature drops of 25K and 5K respectively.

Both rectangular composite power pulse decomposition assumptions produce similar thermal results. At cycle frequencies of 5kHz and 1kHz, together with high duty cycles, the peak junction temperature is marginally higher than the average junction temperature. Using the concept of thermal resistance is adequate under the switching frequency and duty cycle conditions of this problem.

♣

5.5 Average power dissipation

Two commonly used empirical methods for determining power dissipation P_d are

- graphical integration and
- power superposition.

5.5.1 *Graphical integration*

Graphical integration may be formulated by digitally storing a complete cycle of test device voltage and current under limiting steady-state temperature conditions. Each voltage and current time-corresponding pair are multiplied together to give instantaneous values of power loss. Numerical integration techniques are then employed to give the average power dissipation.

5.5.2 *Practical superposition*

This technique is based on substituting a smooth dc voltage source for a complex waveform. A two-pole, two-position switching arrangement is used, which firstly allows operation of the load with the device under test, until the monitored case temperature stabilises. Then, by throwing the switch to the test mode position, the device under test (DUT) is connected to a dc power supply, while the other pole of the switch supplies the normal power to the load to keep it operating at full power level conditions. The dc supply is adjusted so that the semiconductor case temperature remains approximately constant when the switch is thrown to each position for about 10 seconds. The dc source voltage and current values are multiplied together to obtain the average power dissipated.

5.6 Power losses from manufacturers' data sheets

The total power dissipation P_d is the sum of the switching transition loss P_s, the on-conduction loss P_d, drive input device loss P_G, and the off-state leakage loss P_l.
The average total power loss is given by

$$P_d = f_s \int_0^{1/f_s} v(t)\,i(t)\,dt \quad \text{(W)} \qquad (5.43)$$

where f_s is the switching frequency and $v(t)$ and $i(t)$ are the device instantaneous voltage and current over one complete cycle of period $1/f_s$. The usual technique for determining total power loss is to evaluate and sum together each of the individual average power loss components.

5.6.1 *Switching transition power loss, P_s*

Figure 5.16 shows typical power device voltage-current switching waveforms. Normally an exact solution is not required and an approximation based on straight-line switching intervals is usually adequate.
For a resistive load, as derived in Chapter 6

$$P_s = \tfrac{1}{6} V_s I_m \tau f_s \quad \text{(W)} \qquad (5.44)$$

and for an inductive load, as derived in Chapter 6

$$P_s = ½ V_s I_m \tau f_s \quad \text{(W)} \qquad (5.45)$$

where τ is the period of the switching interval (both on and off), and V_s and I_m are the maximum voltage and current levels as shown in figure 5.16. Switching losses occur at both turn-on and turn-off.

5.6.2 Off-state leakage power loss, P_ℓ

During the switched-off period, a small, exponentially temperature dependent current I_ℓ, will flow through the switch. The loss due to this leakage current is

$$P_\ell = I_\ell V_s (1-\delta) \quad \text{(W)} \tag{5.46}$$

where δ is the on-time duty cycle of the switch. Normally P_ℓ is only a small part of the total loss so that the error in neglecting P_ℓ is not significant.

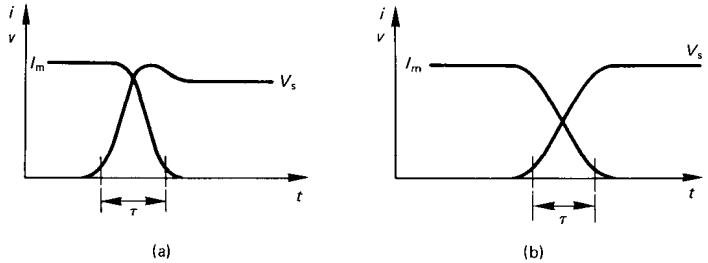


Figure 5.16. *Typical voltage and current at turn-off switching transition for: (a) an inductive load and (b) a resistive load. Current and voltage are interchanged at turn-on.*

5.6.3 Conduction power loss, P_c

The average conduction power loss under a steady-state current condition is given by

$$P_c = \delta I_{on} V_{on} \quad \text{(W)} \tag{5.47}$$

although equation (5.43) is valid in the general case when the integration is performed over the interval corresponding to δ.
The conduction loss for the MOSFET is usually expressed in terms of its on-state resistance (equations (3.14) and (4.12))

$$\begin{aligned} P_c &= \delta\, I^2_{d(rms)} R_{ds(on)} \\ &\simeq \delta\, I^2_{d(rms)} R_{ds(on)}(25°\text{C}) \left\{1 + \frac{\alpha}{100}\right\}^{T_J - 25°\text{C}} \quad \text{(W)} \end{aligned} \tag{5.48}$$

where α is the temperature coefficient of the on-state resistance, which is positive. A linear resistance approximation of equation (5.48) is quite accurate above 25°C if α is small, such that P_c can be approximated by

$$P_c \approx \delta\, I^2_{d(rms)} R_{ds(on)}(25°\text{C}) \{1 + \alpha (T_j - 25°\text{C})\} \quad \text{(W)} \tag{5.49}$$

5.6.4 Drive input device power loss, P_G

A portion of the drive power is dissipated in the controlling junction or, in the case of the MOSFET, in the internal gate resistance. Usually more power is dissipated in the actual external drive circuit resistance. Drive input loss is normally small and insignificant compared with other losses, and can usually be ignored.
Two possible exceptions are:

- One notable exception is in the case of the power GTO thyristor, where continuous gate drive is used to avoid loss of latching or when the holding current is high. The holding current can be 3% of the anode current thus, the gate to cathode junction loss can be included in the total loss calculation for better accuracy. Thus, for a gate junction voltage V_{GC} the gate losses are given by

$$P_g = \delta\, I_G V_{GC} \tag{5.50}$$

  The recovery loss of the gate commutated thyristor (GCT) cathode junction can be included since it is significant because the full anode current is extracted from the gate, thus is involved in recovery of the cathode junction.
- A second exception is the MOSFET and IGBT at high switching frequencies, >50kHz, when the loss in the device, associated with providing the gate charge Q_T is given by equation (4.35):

$$P_G(R_{\text{int}}) = V_{gg} Q_T f_s \frac{R_{G\text{int}}}{R_{G\text{int}} + R_{G\text{ext}}} \quad \text{(W)} \tag{5.51}$$

5.7 Heat-sinking design cases

Heat-sink design is essentially the same for all power devices, but the method of determining power loss varies significantly from device type to device type. The information given in data sheets, in conjunction with the appropriate equation in Table 5.11, allows the designer to calculate power semiconductor thermal rating for a variety of conditions.
Generally, heatsink design is more readily visualised if a thermal equivalent electrical circuit model approach is adopted, as shown in figure 5.1. The equivalence of parameters is shown in Table 5.10. The examples to follow illustrate the approach.

Table 5.10: Thermal equivalent electrical circuit parameters

thermal parameter			thermo-electric model			magnetic model		
temperature drop	degrees Kelvin	ΔT	potential difference	Volts	ΔV	magneto motive force	Amp-turns	$\mathcal{F}$
power dissipated	Watts	P	current flow	Amps	I	flux	Wb	Φ
thermal resistance	K / W	R_θ	Ohm's resistance	Ohms	R	reluctance	Amp-turns/Wb	$\mathcal{R}$

5.7.1 Heat-sinking for diodes and thyristors

At low switching frequencies (<100 Hz), switching loss can be ignored, so that in the case of rectifying diodes or converter-grade thyristors, 50 to 60 Hz, switching loss can usually be ignored. Fast-recovery power diodes switching at less than 500Hz can also have switching losses neglected at low *VA* levels.

***5.7.1i -** Low-frequency switching*
At a given current level I_F and on-time duty cycle δ, on-state power loss can be read directly from the manufacturers' data. Figure 5.17a illustrates loss for square-wave power pulses, while figure 5.17b illustrates loss in the case of half-wave sinusoidal current. Figure 5.17b gives energy loss per cycle, which may be converted to power when multiplied by the sinusoidal pulse frequency.
Thyristor loss due to the current waveform initial rate of rise of current, *di/dt*, can be incorporated and its contribution is added into the manufacturers' conduction loss data for a given device type.

***5.7.1ii -** High-frequency switching*
At device operating frequencies greater than about 100 Hz, fast-recovery diodes are normally employed and at about 500Hz, switching losses must be added to the on-state conduction loss. Diode turn-off loss is usually more significant than turn-on loss. Manufacturers provide maximum reverse recovery charge, Q_R, characteristics as shown in figure 5.18. The reverse recovery charge is a linear function of temperature and between the given junction temperatures of 25°C in figure 5.18a and 150°C in figure 5.18b, interpolation of Q_R is used.
The reverse recovery W.s/pulse, J_R, can be approximated by

$$J_R = V_R Q_R \quad \text{(J)} \tag{5.52}$$

where V_R is the reverse voltage applied to the diode just after turn-off. Losses are lower since the diode only support voltage once peak reverse recovery has occurred. The reverse recovery average power loss is given by

$$P_s = V_R Q_R f_s \quad \text{(W)} \tag{5.53}$$

The total average power loss is the algebraic sum of the steady-state conduction loss and the recovery loss.

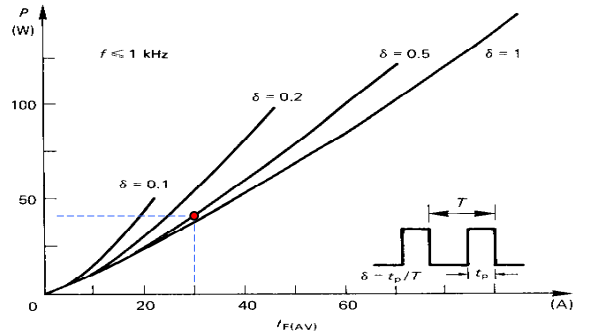


Figure 5.17. *Diode on-state energy loss at low frequency as a function of forward current for: (a) squarewave power pulses and (b) sinusoidal power pulses.*

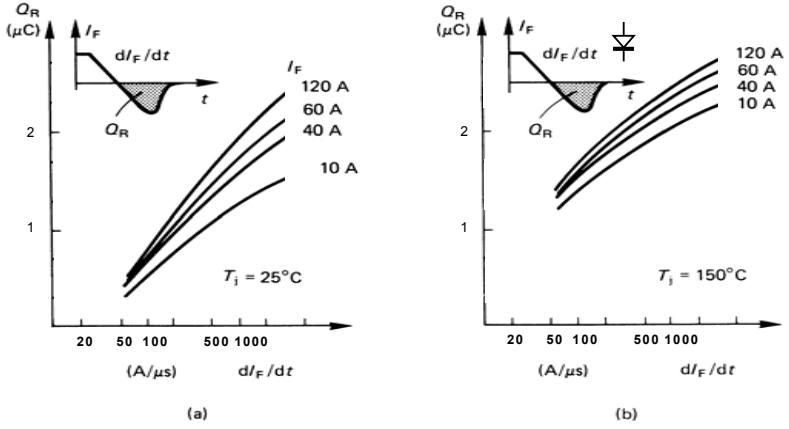
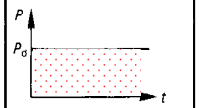
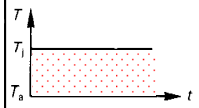
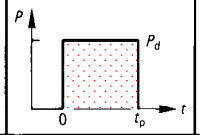
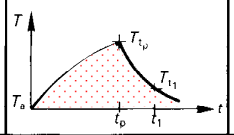
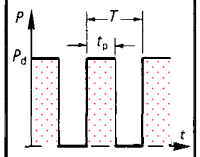
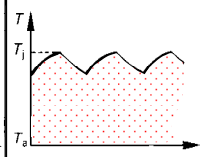
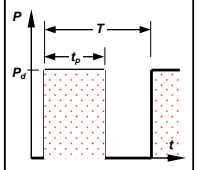
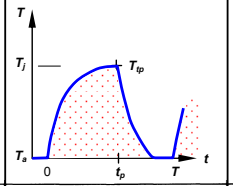
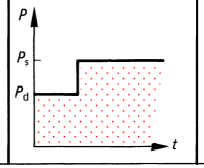
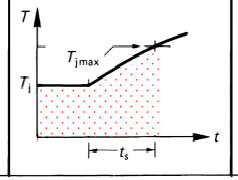


Figure 5.18. *Reverse recovery charge as a function of forward current and dI_F/dt at: (a) 25°C and (b) 150°C junction temperature.*

Table 5.11: Power rating equations based on thermal considerations

Load condition	*Junction power loss waveform*	*Junction temperature rise waveforms*	*Junction temperature solution*
(a) Continuous			$T_j - T_a = P_d R_\theta$
(b) Single pulse			$T_{t_p} - T_a = P_d z_\theta(t_p),\ \delta = 0$ $T_{t_1} - T_a = P_d\{z_\theta(t_1) - z_\theta(t_1 - t_p)\}$
(c) Constant high frequency and constant duty cycle			$\delta = t_p/T$ approximate steady-state solution $T_j - T_a = P_d\{\delta R_\theta + (1-\delta)z_\theta(T + t_p) - z_\theta(T) + z_\theta(t_p)\}$
(d) Constant low frequency and constant duty cycle			$\delta = t_p/T$ $T_{tp} - T_a = P_d z_\theta(t_p)$
(e) Continuous duty followed by a step in power			$T_{jmax} - T_j = P_d R_\theta + (P_s - P_d)z_\theta(t_s)$

Example 5.6: ***Heat-sink design for a diode***

A fast-recovery diode switches 60 A rectangular current pulses at 10kHz. The off-state bias is 400V and the external circuit inductance limits the reverse dI_F/dt to 100A/μs. If the device junction-to-case thermal resistance is 0.7K/W, calculate the minimum heat-sink requirement with a 50 per cent conduction duty cycle, if the maximum ambient temperature is 40°C.

Solution

The steady-state loss given from figure 5.17a is about 40 W when using $I_{F(AV)}$ = 30A for δ = 0.5.
Minimum possible heat-sinking requirements occur when T_j is a maximum, that is 150°C from figure 5.18b. From figure 5.18b, for dI_F/dt = 100 A/μs and I_F = 60A, the maximum reverse recovery charge is 1.3μC. The switching power loss (over estimate) is given by

$$P_s = Q_R V_R f_s$$
$$= 1.3\mu C \times 400V \times 10kHz = 5.2W$$

The total power loss is therefore

$$P_d = 40 + 5.2 = 45.2\text{W}$$

Since the frequency and duty cycle are both high, the concept of thermal resistance is appropriate; that is

$$T_j = T_a + P_d\,(R_{\theta j\text{-c}} + R_{\theta c\text{-a}})$$

Therefore $150\text{V} = 40\text{V} + 45.2\text{A} \times (0.7\Omega + R_{\theta c\text{-a}})$ (in terms of the electrical dual)

whence $R_{\theta c\text{-a}} = 1.73$ K/W

Figure 5.6b shows that a minimum of 50mm length of matt black heat sink is required. This assumes that the case-to-sink thermal resistance is negligible. In order to improve device reliability and lifetime, operation at $\hat{T}_j$ is avoided. A derating of 40 to 50°C significantly reduces junction thermal fatigue and can result in a tenfold improvement in reliability. To restrict $\hat{T}_j$ to 100°C, $R_{\theta\ \text{c-a}}$= 0.7K/W, necessitating 120 mm of the heat sink as characterised in figure 5.6b. The flatness of the $R_{\theta\ \text{s-a}}$ curve means that the effectiveness of the heat sink is diminished and either a wider sink of the same length or a shorter length of a profile offering lower thermal resistance would be more effective in reducing device thermal fatigue.

♣

5.7.2 Heat-sinking for IGBTs

The IGBT conduction loss is related to the gate voltage and the collector current magnitude, which specify the on-state voltage. No simple power loss characteristic is possible, as in figure 5.17 for the diode and thyristor. Fortunately, the power switching IGBT is used in such a way that its on-state collector-emitter voltage is near constant, whence conduction loss is given by

$$P_c = \delta\, v_{ce}\, \bar{I}_c \qquad \text{(W)} \qquad (5.54)$$

Example 5.7: ***Heat-sink design for an IGBT- repetitive operation at a high duty cycle***

A power IGBT is used to switch a 20A, 100V highly inductive load at 10 kHz. The transistor maximum on-state duty cycle is 90 per cent and the device has a junction-to-case thermal resistance of 0.7K/W. The transistor on-state voltage is maintained at 2V and the switch-on and switch-off times are 1 and 2 µs respectively. If the junction temperature is not to exceed 125°C with a maximum ambient temperature of 35°C, what is the minimum heat-sink requirement? Assume that the transistor is in a T0247 package, which is mounted directly on the heat sink but with silicone grease used.

Solution

Since both the duty cycle and switching frequency are high, the peak junction temperature is approximated by the average junction temperature. That is, the concept of thermal resistance is valid.

The on-state power loss is given by

$$P_c = \delta\, v_{ce}\, \bar{I}_c = 0.9 \times 2\text{V} \times 20\text{A} = 36\text{ W}$$

From equation (5.45), the switching losses for an inductive load are

$$P_s = P_{s(\text{on})} + P_{s(\text{off})}$$
$$= \tfrac{1}{2} \times 100 \times 20 \times (1\mu\text{s} + 2\mu\text{s}) \times 10\text{ kHz} = 30\text{ W}$$

Total power losses P_d are 36W+30W = 66 W.

From

$$\hat{T}_j = T_a + P_d(R_{\theta j\text{-c}} + R_{\theta c\text{-s}} + R_{\theta s\text{-a}})$$
$$125°\text{C} = 35°\text{C} + 66\text{W} \times (0.7 + 0.1 + R_{\theta s\text{-a}})$$

Therefore $R_{\theta c\text{-a}} = 0.56$ K/W

The case-to-heat-sink thermal resistance value of 0.1K/W for a T0247 non-insulated case using silicone thermal grease was obtained from Table 5.1. To obtain the minimum heat-sink thermal resistance of 0.56K/W, 150 mm of the heat sink with cross-section shown in figure 5.6a is required. Clearly, a sink profile that has a lower thermal resistance per unit length would be more suitable.

♣

5.7.3 Heat-sinking for power MOSFETs

Switching losses in MOSFETs tend to be low at frequencies below 20 kHz and therefore may be neglected, along with gate and off-state losses. Conduction loss is generally expressed in terms of the on-state resistance as I^2R loss. The first step in the thermal design is to determine the total power dissipation in the device, which is generally dominated by the conduction loss. Determination of this loss is not trivial since, while the power dissipation determines junction temperature, the power dissipation itself is a function of junction temperature, because the on-state resistance increases with temperature, as shown in figure 3.13.

Example 5.8: ***Heat-sink for a MOSFET - repetitive operation at high peak current, low duty cycle***

Find the thermal resistance of the heat sink needed for a MOSFET conducting a repetitive 20A rectangular current waveform. On-time is 10µs, duty cycle is 0.1 per cent, and the maximum ambient temperature is 40°C. Assume $R_{ds\ (\text{on})}$ at 150°C and 20 A is 5 Ohms, and $R_{\theta\ \text{j-c}}$ = 1.5 K/W.

Solution

Since the on-state duty cycle and switching frequency are both low, the peak junction temperature at the end of the on-period will be significantly different from the average junction temperature. The concept of thermal resistance from the junction to the case is therefore invalid; the concept of thermal impedance is used instead.

The peak power per pulse = $P_p = I^2R = 20^2 \times 5\Omega = 2 \times 10^3$ W

Using a thermal impedance basis, the case temperature is given by

$$T_j = T_c + P_p \times Z_{\theta j\text{-c}}$$
$$= T_c + P_p\, r(t_p)\, R_{\theta j\text{-c}}$$

where $r(t_p)$ is the transient thermal impedance factor for the junction-to-case. For a 10 µs pulse from figure 5.10, $r(t_p)$ = 0.03, assuming δ = 0.001≈ a single pulse condition, thus

$$150°\text{C} = T_c + 2 \times 10^3 \times 0.03 \times 1.5 = T_c + 90°\text{C}$$

that is $T_c = 60°\text{C}$

The average junction temperature is

$$T_j = T_c + \bar{P}_d R_{\theta j\text{-c}} = T_c + \delta P_d R_{\theta j\text{-c}}$$
$$= 60°\text{C} + 0.1\% \times 2 \times 10^3\,\text{W} \times 1.5°\text{C/W}$$
$$= 60°\text{C} + 3°\text{C} = 63°\text{C}$$

Although the average junction temperature is only 3°C above the case temperature of 60°C, the peak junction temperature reaches 150°C.

Because of the heat-sink thermal inertia, the concept of thermal resistance and average power are used for calculations involving the heatsink. That is

$$T_c = T_a + \bar{P}_d R_{\theta c\text{-a}} = T_a + \delta P_d R_{\theta c\text{-a}}$$
$$60°\text{C} = 40°\text{C} + 0.001 \times 2 \times 10^3 \times R_{\theta c\text{-a}}$$

thus $R_{\theta c\text{-a}} = 10$ K/W

The heat sink of cross-section shown in figure 5.6a is not suitable in this application, and one of a much smaller surface area is applicable. A heatsink may not be necessary since the package thermal resistance $R_{\theta\ \text{c-a}}$, shown in figure 5.1, may be less than 10K/W, there in satisfying equation (5.17). See problem 5.6.

♣

If the junction operating temperature is unknown but can be assumed greater than 25°C, from equation (5.49), the total power loss can be expressed as

$$P_d = P_o + I^2_{d(\text{rms})} R_{ds(\text{on})}(25°\text{C})\{1 + \alpha\,(T_j - 25°\text{C})\} \qquad \text{(W)} \qquad (5.55)$$

where P_o represents all losses other than the conduction loss, and is assumed temperature independent. The temperature coefficient α for $R_{ds\ (\text{on})}(25°\text{C})$ is positive, typically 1 per cent/K as indicated in figure 3.13. The usual thermal equality holds, that is

$$T_j = T_a + R_{\theta j\text{-a}}\, P_d \qquad \text{(K)} \qquad (5.56)$$

Combining equations (5.55) and (5.56) by eliminating T_j yields

$$P_d = \frac{P_o + I^2_{d(\text{rms})} R_{ds(\text{on})}(25°\text{C})\{1 + \alpha\,(T_a - 25°\text{C})\}}{1 - I^2_{d(\text{rms})} R_{ds(\text{on})}(25°\text{C})\alpha\, R_{\theta j\text{-a}}} \qquad \text{(W)} \qquad (5.57)$$

The denominator yields an asymptotic maximum drain current of

$$I_{d(\text{rms})} = \frac{1}{\sqrt{R_{ds(\text{on})}(25°\text{C})\ \alpha\ R_{\theta j\text{-c}}}} \quad \text{(A)} \qquad (5.58)$$

at which current thermal runaway would result. In practice, insufficient gate voltage is available and the device would leave the constant-resistance region and enter the constant-current region, where the above analysis is invalid.

Example 5.9: ***Heat-sink design for a MOSFET - repetitive operation at high duty cycle***

A power MOSFET switches 5 A rms at 10 kHz with a maximum on-state duty cycle of 90 per cent. The junction-to-case thermal resistance is 0.7 K/W, the maximum ambient temperature 35°C, and on-state resistance at 25°C is 1 Ohm. If the heat-sink arrangement yields an effective case-to-ambient thermal resistance of 1.3 K/W and α = 0.01 /K, what is the junction operating temperature?

Solution

Since the switching frequency and duty cycle are both relatively high, the thermal resistance concept based on average junction power dissipation is valid.

Assuming zero losses other than conduction losses, then P_o = 0. Equations (5.55) and (5.56) rearranged to eliminate P_d yielding

$$T_j = \frac{T_a + R_{\theta j\text{-a}} I^2_{d(\text{rms})} R_{ds(\text{on})}(25°\text{C})\{1-25\alpha\}}{1-\alpha R_{\theta j\text{-a}} I^2_{d(\text{rms})} R_{ds(\text{on})}(25°\text{C})} \quad \text{(W)} \qquad (5.59)$$

Assuming typical α = 0.01/K and $R_{\theta\,\text{j-a}} = R_{\theta\,\text{j-c}} + R_{\theta\,\text{c-a}}$

$$T_j = \frac{35°\text{C} + 2\times 5^2 \times 1\Omega \times (1\text{-}25\times 0.01)}{1 - 0.01\times 2\times 5^2 \times 1\Omega} = 145°\text{C}$$

♣

Example 5.10: ***Two thermal elements on a common heatsink***

A dc chopper has a MOSFET switch that dissipates 40W and a load freewheel diode that dissipates 20W. Each power device is mounted on a common heatsink. The MOSFET has a junction-to-case thermal resistance of 0.7K/W and a case-to-heatsink thermal resistance of 0.5K/W. The diode has a junction-to-case thermal resistance of 0.8K/W and a case-to-heatsink thermal resistance of 0.6K/W.

i. Determine the maximum heatsink thermal resistance that maintains both junction temperatures below 90°C in a 30°C ambient.
ii. Semiconductor lifetime approximately doubles for every 10°C decrease in junction temperature. If the heatsink in the previous case is fan cooled, estimate the lifetime improvement if the heatsink thermal impedance is halved with fan cooling.
iii. If the load current is constant (25A) and the switch and diode on-state voltages are the same, determine the chopper on-time duty cycle and device instantaneous losses assuming no switching losses (only on-state losses).

Solution

i. Applying Kirchhoff's voltage law to each loop of the equivalent thermal circuit shown gives:

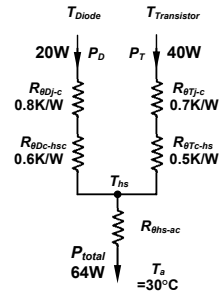
$$T_{Dj} - T_{hs} = 20\text{W} \times (0.8\text{K/W} + 0.6\text{K/W}) = 28°\text{C}$$
$$T_{Tj} - T_{hs} = 40\text{W} \times (0.7\text{K/W} + 0.5\text{K/W}) = 48°\text{C}$$

Since both semiconductor devices are mounted on the same heatsink, T_{hs} is the same in each case, the MOSFET virtual junction will operate 20°C hotter than the diode junction. Therefore the MOSFET junction temperature should not exceed 90°C, that is

$$90°\text{C} - T_{hs} = 40\text{W} \times (0.7\text{K/W} + 0.5\text{K/W}) = 48°\text{C}$$

giving a heat sink surface temperature of 90°C - 48°C = 42°C and a diode junction temperature of 42°C + 28°C = 70°C. The heatsink thermal resistance requirement is

$$T_{hs} - T_a = 42°\text{C} - 30°\text{C} = R_{\theta hs-a} \times (40\text{W} + 20\text{W})$$
$$R_{\theta hs-a} = 42°\text{C} - 30°\text{C} \,/\, 40\text{W} + 20\text{W} = 12°\text{C} \,/\, 60\text{W} = 0.2\,\text{K/W}$$



ii. Assume device losses are not affected by temperature and the heatsink thermal resistance is decreased to ½×0.2 = 0.1 K/W, then

$$T_{hs} - T_a = T_{hs} - 30°\text{C} = 0.1\text{K/W} \times (40\text{W} + 20\text{W})$$
$$T_{hs} = 0.1\text{K/W} \times (40\text{W} + 20\text{W}) + 30°\text{C} = 36°\text{C}$$

The device junction temperatures are given by

$$T_{Dj} - 36°\text{C} = 20\text{W} \times (0.8\text{K/W} + 0.6\text{K/W}) \text{ that is } T_{Dj} = 64°\text{C}$$
$$T_{Tj} - 36°\text{C} = 40\text{W} \times (0.7\text{K/W} + 0.5\text{K/W}) \text{ that is } T_{Tj} = 84°\text{C}$$

The junction temperature of each device has decreased by about 6°C, so although the lifetime will have increased, lifetime improvement is not doubled. Device package thermal properties are more dominant than the heatsink in determining junction temperatures.

iii. If the on-state duty cycle is δ and the instantaneous device losses are P (and the same since on-state voltage is the same for both devices and the current is constant hence the same when each device is conducting) then

mosfet $\quad \delta P = 40\text{W}$
diode $\quad (1-\delta)P = 20\text{W}$

Summing these two equations gives an instantaneous loss of P = 60W, whence a switch on-state duty cycle of $\delta = \frac{2}{3}$, that is the switch conducts for $66\frac{2}{3}$% of the cycle period. The diode on-state voltage is therefore 60W/30A = 2.0V and the MOSFET on-state resistance is 60W/30A^2 = 67mΩ.

♣

Example 5.11: ***Six thermal elements (on a common substrate) in a common package***

A three-phase full-wave diode rectifier package consists of six-diode die within a single module. The junction-to-case thermal resistance of each die is 0.24K/W. The module is mounted on a heatsink with a module-to-heatsink contact thermal resistance of 0.2K/W and a heatsink-to-ambient thermal resistance of 0.1K/W. The maximum ambient temperature is 30°C and the highly inductive load current is constant at 100A. If the diode on-state voltage is 1V, determine

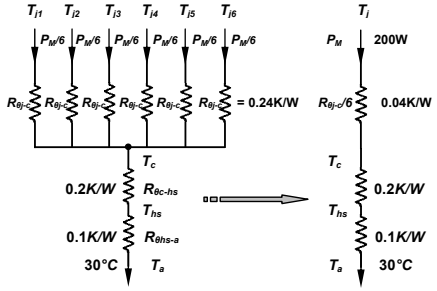
i. the diode junction temperature
ii. the current to double the rectifier lifetime (decrease junction temperature by 10°C)
iii. the heatsink to double the rectifier bridge lifetime (at 100A).

Solution

i. During rectification, two diodes always conduct therefore total module conduction losses are

$$P_M = 2 \times I_o \times V_{Don} = 2 \times 100\text{A} \times 1\text{V} = 200\text{W}$$

The figure shows how the six thermal paths can be reduced to the simplified equivalent thermal model on the right.



Applying Kirchhoff's voltage law

$$T_j - T_a = P_M \times (\tfrac{1}{6} R_{\theta j-c} + R_{\theta c-hs} + R_{\theta hs-a})$$
$$T_j - 30°\text{C} = 200\text{W} \times (\tfrac{1}{6} \times 0.24\text{K/W} + 0.2\text{K/W} + 0.1\text{K/W})$$
$$\Rightarrow \quad T_j = 98°\text{C}$$

ii. If the current is to reduce so as to decrease the diode junction temperature by 10°C then

$$T_j - T_a = P_M \times \left(\tfrac{1}{6} R_{\theta j-c} + R_{\theta c-hs} + R_{\theta hs-a}\right)$$

$$88°C - 30°C = P_M \times \left(\tfrac{1}{6} \times 0.24K/W + 0.2K/W + 0.1K/W\right) \Rightarrow P_M = 170.6W$$

Assuming the diode on-state voltage drop is independent of current, that is remains 1V then

$$P_M = 2 \times I_o \times V_{Don}$$

$$170.6W = 2 \times I_o \times 1V \Rightarrow I_o = 85.3A$$

iii. When the junction temperature is reduced by 10°C to 88°C by decreasing the heatsink thermal resistance, and the constant load current is maintained at 100A

$$T_j - T_a = P_M \times \left(\tfrac{1}{6} R_{\theta j-c} + R_{\theta c-hs} + R_{\theta hs-a}\right)$$

$$88°C - 30°C = 200W \times \left(0.04K/W + 0.2K/W + R_{\theta hs-a}\right) \Rightarrow R_{\theta hs-a} = 0.5K/W$$

♣

5.8 High-performance cooling for power electronics

In many instances, standard finned aluminium heat sinks, even with fan assistance, cannot achieve the required cooling performance due to physical limitations in heat transfer capabilities, namely the limited thermal conductivity of air for convection and copper and aluminium for conduction.
Figure 5.19 shows a comparison of various cooling techniques as a function of the attainable heat transfer in terms of the heat transfer coefficient, *h*. For example, to accommodate a heat flux of 100W/cm^2 at a temperature difference of 50K requires an effective heat transfer coefficient (including a possible area enlarging factor) of 20kW/m^{2}K (*h* = 100W/cm^2×50K). From Figure 5.19 it can be concluded that liquid cooling can play an important role in thermal management.

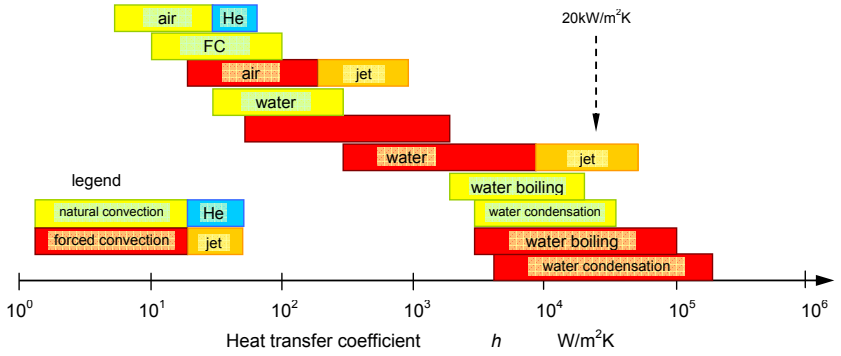


Figure 5.19. *Heat transfer coefficient h attainable with natural convection, single-phase liquid forced convection, and boiling for different coolants. (See figure 5.50)*

5.9 Conduction and heat spreading

In all power electronic cooling applications, heat from the device sources must be transmit via thermal conduction to the surfaces exposed to the cooling fluid before it can be rejected to the coolant. As shown in Figure 5.20, heat must be conducted from the die to the $A\ell_2 0_3$ substrate and Cu base plate to the heat sink before it can be rejected to the flowing air. A thermal interface material (TIM) may be used to facilitate thermal conduction from the die to the base plate and from the base plate to the heat sink. In power electronics, heat spreaders (heat sinks without any cooling fins and less bulky structures) in the form of a flat plate with good thermal conductivity may be interposed between the die mounted substrate and heatsink, to facilitate spreading of the heat from its small source. Vapour chambers are also used to spread heat from a concentrated die or module heat source to a larger heat sink.

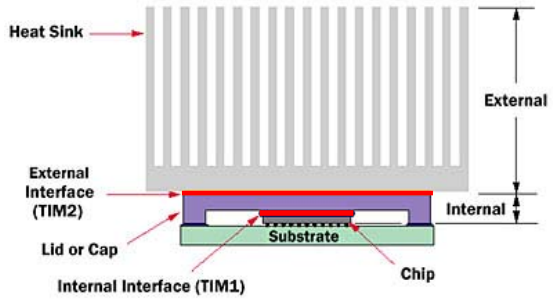


Figure 5.20. *Power semiconductor package with thermal conduction path to heat sink via TIMs.*

For high-power applications, the interface thermal resistance becomes an important constraint. Direct soldering (for example, reflow soldering) is often difficult, certainly when copper is used because of the large co-efficient of thermal expansion CTE mismatch between Cu and $A\ell_2 0_3$ and in turn, Si. Diamond-filled greases have an effective thermal conductivity of over 20W/mK. Also possible is a nanostructured foil, which utilizes a fast exothermic reaction to create a soldered connection virtually at room temperature.
Heat spreading is an effective method of mitigating the need for complicated high-heat flux cooling options. To be effective the benefits of decreasing the heat flux density by increasing the area should outweigh the penalty of adding another thermal layer through which the heat must conduct. Other than a traditional copper heat spreading base plate, the option is to use advanced heat spreading materials such as carbonaceous materials, metal-matrix composites, ceramic matrix composites (for example, diamond-particle-reinforced silicon carbide), or ScD (Skeleton cemented Diamond), all with higher thermal conductivities than copper, much lighter, and tuneable CTEs.
By applying heat spreaders, cooling methods such as loop heat pipes and low-flow liquid cooling may be augmented to accommodate higher heat flux applications. Figure 5.21a shows heat spreading results for a 300W heat source of 2cm^2 area as a function of thermal conductivity λ, thickness *t*, and cooling boundary condition (that is, heat transfer coefficient *h*). Heat spreading is a complex phenomenon because the conduction and convection effects cannot be separated and the two effects compete: increasing the thickness increases the through-plane thermal resistance but decreases the in-plane thermal resistance. For example, comparing the two upper curves with the two lower curves, their order is changed. The figure also show that heat spreaders can be used to decrease the required fluid-side heat transfer coefficient to manageable values, below 5kW/m^{2}K, which can be realized with hydrofluoroether (HFE) cooling fluids. For example, using an 8×8=64cm^2 heat spreader of an advanced composite with a thermal conductivity λ of 800W/mK and a thickness *t* of 4mm results in a temperature rise of 40°C with a heat transfer coefficient *h* of 2500W/m^{2}K.
For a single heat energy source, minimal thermal gain results from a Cu base plate thickness in excess of 6mm. In the case of power IGBT modules, this boundary is complicated by the fact that many heat sources, die, are adjacently bonded to a given copper coated ceramic substrate, which is then bonded to a Cu spreader base plate.
The relative heat spreading resistance for varying spreader thicknesses in aluminium, copper, silver, and Cusil (72%Ag 28%Cu) is shown in figure 5.21b. Aluminium is extensively used (also for air-cooled heat sinks) even though its relative thermal conductivity is the poorest of the materials shown; because of its:

- lowest cost
- ease of fabrication
- performance adequacy

5.10 Heat-sinks

Heat sinks are devices that enhance heat dissipation from a hot surface, usually the case of a heat-generating component, to a cooler ambient, usually air. Air is assumed to be the cooling fluid. In most situations, heat transfer across the interface between the solid surface and the coolant air is the least efficient within the system, and the solid-air interface represents the greatest barrier for heat dissipation. A heat sink lowers this barrier by increasing the surface area that is in direct contact with the coolant. This allows more heat to be dissipated and/or lowers the device operating temperature. The primary purpose of a heat sink is to maintain the device junction temperature below the maximum allowable temperature specified by the device manufacturer.

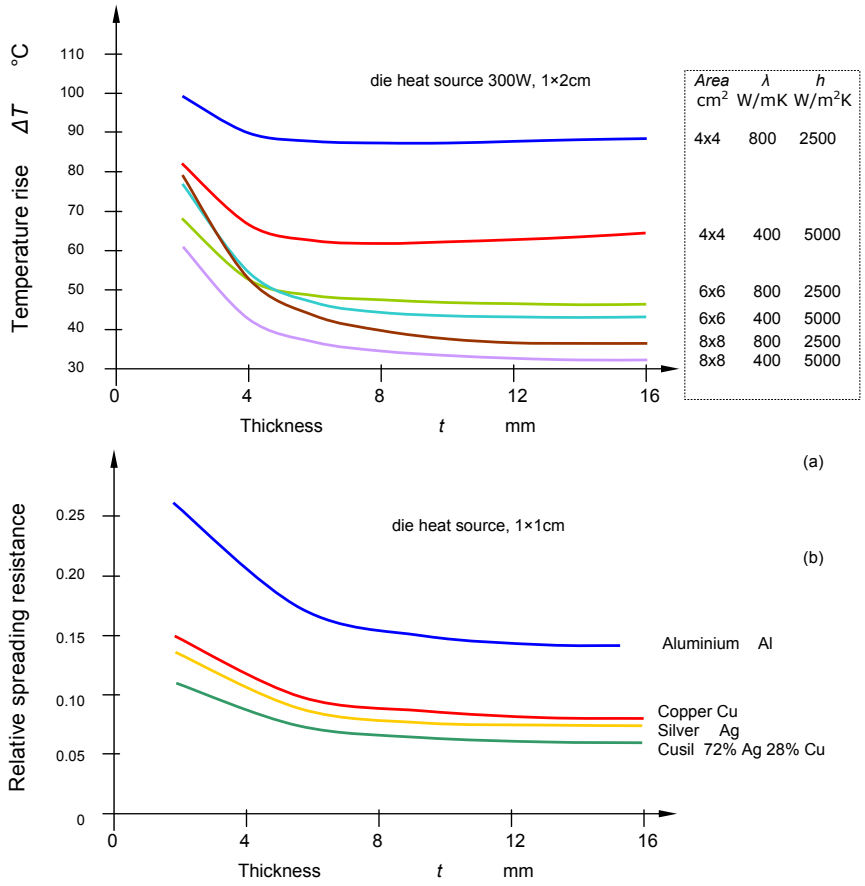


Figure 5.21. *Example of effect of thickness of heat spreader for various: (a) heat source areas, material thermal conductivities, and heat transfer coefficients and (b) material and resultant relative heat spreading resistance.*

5.10.1 Required heat-sink thermal resistance

To begin the heat sink selection, the heat sink thermal resistance required to satisfy the thermal criteria of the component is determined. By rearranging equation (5.27) into terms of the ambient temperature T_a, the heat sink resistance is obtained as

$$R_{\theta s-a} = \frac{T_j - T_a}{P_d} - R_{\theta j-c} - R_{\theta c-s} \qquad (5.60)$$

where T_j, P_d and $R_{\theta j\text{-}c}$ are provided by the device manufacturer, and T_a and $R_{\theta c\text{-}s}$ are user-defined parameters.
The ambient air temperature T_a for cooling electronic equipment depends on the operating environment in which the component is used. Typically, it ranges from 35 to 45°C, if the external air is used, and from 50 to 60°C, if the component is enclosed or in the wake of another heat-generating component.

5.10.2 Heat-sink selection

In selecting an appropriate heat sink that meets the required thermal criteria, it is necessary to examine various parameters that affect not only the heat sink performance itself, but also the overall performance of the system. The choice of a particular type of heat sink depends largely on the thermal budget allowed for the heat sink and external conditions surrounding the heat sink. There is not a single thermal resistance value assigned to a given heat sink, since the thermal resistance varies with external cooling conditions.
When selecting a heat sink, it is necessary to classify the airflow as natural, low flow mixed, or high flow forced convection. Natural convection occurs when there is no externally induced flow and heat transfer relies solely on the free buoyant flow of air surrounding the heat sink. Forced convection occurs when the flow of air is induced by mechanical means, usually a fan or blower. There is no clear distinction on the flow velocity that separates the mixed and forced flow regimes. Generally the effect of buoyant force on the overall heat transfer diminishes to a negligible level (under 5%) when the induced airflow velocity excess 1 to 2m/s.
Next, the required volume of a heat sink is determined. Table 15.12 shows approximate ranges of volumetric thermal resistance of a typical heat sink under different flow conditions.

Table 15.12: Range of volumetric thermal resistance

flow condition m/s	volumetric resistance cm^3 °C/W
natural convection	500-800
1.0	150-250
2.5	80-150
5.0	50-80

The heat sink volume for a given low flow condition is obtained by dividing the volumetric thermal resistance by the required thermal resistance. Table 5.12 is used only as a guide for estimation purposes at the beginning of the selection process. The actual resistance values may vary outside the above range depending on additional parameters, such as actual dimensions of the heat sink, type of the heat sink, flow configuration, orientation, surface finish, altitude, etc. The smaller values shown correspond to a heat sink volume from approximately 100 to 200cm^3 up to about 1000cm^3.

The tabulated ranges in Table 5.12 assume that the design has been optimized for a given flow condition. Although there are many parameters to be considered in optimizing a heat sink, one of the most critical parameters is the fin density. In a planar fin heat sink, optimum fin spacing is mainly related to two parameters: flow velocity and fin length H in the direction of the flow, as shown in Table 5.5.
It is beneficial to increase the width of a heat sink rather than its length. Also, the effect of radiation heat transfer is important in natural convection, as it can be responsible of up to 30% of the total heat dissipation. As the ambient temperature rises the heatsink temperature increases for a constant thermal power loading, and the heatsink thermal resistance decreases due to the increased significance of thermal radiation (T^4 dependence in equation (5.1)) in the heat removal process. As the ambient temperature decreases the heatsink thermal resistance decreases only slightly.
Unless the component is facing a hotter surface nearby, it is imperative to have the heat sink surfaces suitably painted or anodized to enhance radiation.

5.10.3 Heat sink types

Heat sinks separate into three broad categories

- Plate-fin – suitable for general straight airflow
- Pin-fin – suitable for omi-directional airflow
- Foam-fin – suitable for ducted airflow with high pressure drop

Heat sinks can be classified in terms of manufacturing methods and their final form shapes. Conventional heat sink manufacturing methods involved extruding and die-casting.

The most common types of air-cooled heat sinks are summarised in Table 5.13 and include:

Extruded fins:
Extrusion is a process in which a solid block is converted into a continuous length of uniform cross-section by forcing it to flow under high pressure through a die orifice, which is so shaped, as to impart the required form to the product. Typically, billets of aluminium are placed within a strong walled enclosure and are caused to extrude through the die under a high pressure exerted by a ram, actuated hydraulically or mechanically. Extrusion is the most widely used method for heat sink manufacture.

This process allows the formation of elaborate two-dimensional shapes capable of dissipating large heat loads. They may be cut, machined, and coated. A cross-cutting will produce omni-directional, rectangular pin fin heat sinks, and incorporating serrated fins improves the performance by approximately 10 to 20%, but with a slower extrusion rate. Extrusion limits, such as the fin height-to-gap fin thickness, *L/s*, usually dictate the flexibility in design options (see figure 5.4). Typical fin height-to-gap aspect ratio *L/s* of up to 6 and a minimum fin thickness *t* of 1.3mm, are attainable with a standard extrusion. A 10 to 1 aspect ratio *W/L* and a fin thickness *t* of 20mm can be achieved with special die design features. However, as the aspect ratio increases, the extrusion tolerance is compromised.

Casted fins:
In the die-casting method, molten metal is forced under pressure into metal dies or moulds to produce accurately dimensioned parts. It is the fastest of all casting processes and is often employed where rapidity and economy in production are essential. The thermal conductivity of cast heat sinks may be worsened by porosity caused by gases evolving during solidification.

Sand, lost core, and die casting processes are available with or without vacuum assistance, in aluminium or copper/bronze. This technology is used in high-density pin fin heat sinks which provide maximum performance when using impingement cooling.

***Modified die-casted fins*:**
The modified die-casting process involves the extension of basic die-casting principles, whereby the base of the heat sink is die-cast around a fixtured array of extremely thin stamped fins. The fins are separated by spacers, which prevent the die-cast material from flowing into the fin-to-fin spacing. Aluminium is the most commonly used material for this technique. The absence of a so-called interface between the fins and the base eliminates the impact of an interface resistance. This process allows much higher aspect ratios while fulfilling requirements of small inter-fin spacing. These heat sinks are usually combined with a heat pipe to provide a thermal solution.

***Bonded/fabricated fins*:**
Bonded heat sinks are often built-up extrusions, typically manufactured by assembling extruded plates into slots on an extruded or machined heat sink base, and held in place by an interface, usually a two part thermosetting thermally conductive epoxy or a solder. However, the bonding agent presents a thermal barrier. These heat sinks are often costlier to manufacture, and the base typically requires special machining. Process limitations are usually related to the strength of the bonding agent and dimensional constraints for the slot in the heat sink base. Hybrid heat sinks utilizing different materials for the fins and the base are possible. Bonded fin arrays are most commonly rectangular plate fin arrays. Most air-cooled heat sinks are convection limited, and the overall thermal performance of an air-cooled heat sink can be improved if more surface area can be exposed to the air stream. These high performance heat sinks utilize thermally conductive aluminium-filled epoxy to bond planar fins onto a grooved extrusion base plate. This process allows for a much greater fin height-to-gap aspect ratio *L/S* of 20 to 40, greatly increasing the cooling capacity without increasing volume requirements.

Forged/stamped fins:
In *forged heat sinks*, the fin arrays are formed by forcing raw material into a moulding die by a punch. Common problems in forging are the choking of material in the moulding die cavity, which leads to fins of uneven height. Aluminium and magnesium alloys are readily forged, and an important economic advantage is a typically low rejection rate for the process.
Some of the attractive benefits of forging include high strength, superior surface finish, structural rigidity, close tolerance capabilities, continuity of shape, and high uniformity of material.

Copper or aluminium sheet metals are *stamped* into desired shapes. They are used in traditional air-cooling of electronic components and offer a low cost solution to low density thermal problems. They are suitable for high volume production, because advanced tooling with high-speed stamping lowers costs. Additional labour-saving options, such as taps, clips, and interface materials, can be factory applied to help to reduce the assembly costs.

***Folded/convoluted fins*:**
Folded heat sinks are built-up sheet metal, manufactured by folding sheet metal into a serpentine fin array. The folded metal sheets are attached to the base of the heat sink by soldering or brazing, which results in additional thermal resistance at that interface. However, this contact resistance between the folded fins and the heat sink base is smaller, due to the fact that the 'bends' of folded fins are typically flattened while bonding or brazing, thus increasing the contact surface area. Difficulty in achieving smaller fin pitches required to construct dense arrays is a common issue. Similar to the bonding process, this manufacturing method allows flexibility in designing hybrid heat sinks made up of a combination of different materials.

Alternatively, corrugated sheet metal of either aluminium or copper increases surface area, hence, the volumetric performance. The heat sink is then attached either to a base plate or directly to the heating surface via epoxying or brazing. It is not suitable for high profile heat sinks because of the availability and fin efficiency.

***Skived fins*:**
In the skiving process, fins are machined using special tooling, whereby precisely sliced layers from an extruded metal block are bent at the base of the slice to form slender curved fins. Since the fins and base are an integral unit, the interface resistance found in folded and bonded heat sinks is absent. Aluminium 6063 is the preferred material because of its superior machinability and strength, but copper arrays are also available. The depth of cut determines the fin thickness and can result in extremely thin fin structures, yielding light and competitive heat sink designs.

***Machined fins*:**
Heat sinks are machined out of a metal block by material removal to create the inter-fin spaces. Most commonly, they are manufactured by gang saw cutting on a computer numerical control machine. The gang saw consists of multiple saw cutters on an arbour with precise spacing, which depends on the heat sink geometry to be machined. Fins damaged and distorted during processing require extensive secondary operations. Material is also consumed in an unproductive manner by the generation of scrap metal.

Swaged fins:
Individual fins are placed in a pre-grooved base, and then roller swages the sides of the fins to maintain them in place.

Table 5.13: Feature of different types of heatsinks

Heatsink fin type	Applications	Thermal resistance	Advantages	Disadvantages
extruded	most applications	varies	versatile	limited size
die-casted	low power	high	expensive	low thermal conductivity, expensive die charge
bonded	large applications	high	close tolerances	expensive
single-fin fabricated assembly	all application	very low	light weight and low profile with high degree of flow	expensive
forged	most applications	moderate	inexpensive	limited in design and flow management
stamped	low power	high	inexpensive	low performance
convoluted (folded) fin	ducted air	high at low flows low at high flows	high heat flux density	expensive, needs ducting
skived	most applications	moderate	close tolerances	thick base, high weight, orientation sensitive
machined	prototypes	design dependant	quickly produced	high aspect ratio fins difficult to machine – inconsistent fin geometry
swaged	high power	medium	suitable for power devices	heavy and bulky, limited availability for flow management

5.10.4 Heatsink fin geometry

Figure 5.22, showing the nomenclature of the array geometry, including the fin height, *L*, fin thickness, *t*, inter-fin spacing, *s*, width of base, *W*, and the length of the heat sink base, *H*.

Figure 5.23 shows the results of thermal resistance calculations for an aluminium heat sink (0.1x0.1m heat sink base; fin height, *L*=0.05m), in forced air convection for flow conditions of 20Pa and 0.01m^3/s. The thermal resistance is defined as the ratio of the excess temperature difference in °C to the heat dissipation rate in W, where the excess temperature difference is between the bottom surface of the heat sink base and the incoming air at the heat sink inlet. It is assumed that such a flow condition is defined by a pressure drop and flow rate represented a single point on a fan characteristic curve.

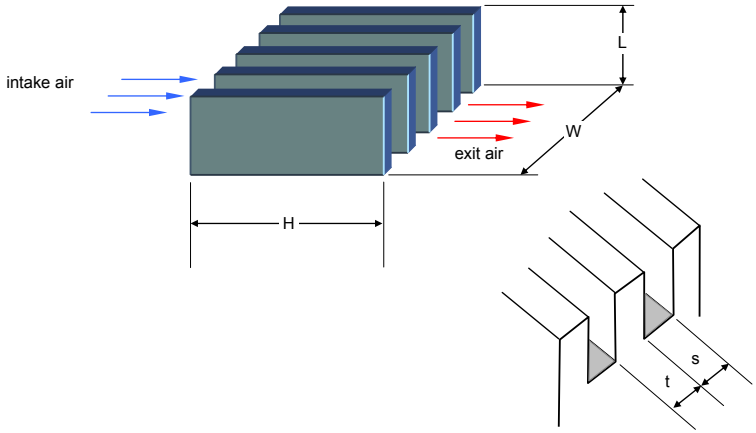


Figure 5.22. *Geometry of the plate fin heat sink analyzed.*

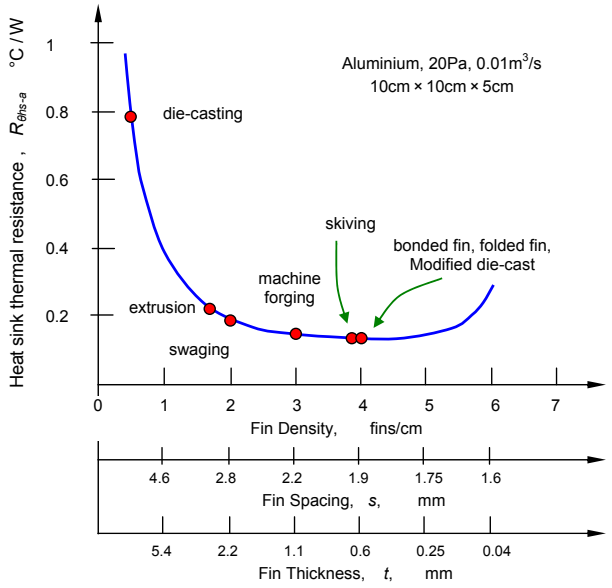


Figure 5.23. *Typical thermal characteristic plot for a heat sink: Thermal Resistance versus Fin Density, 0.1mx0.1m, 0.05m, Aluminium, 20Pa, 0.01m³/s.*

For each value of fin density, there is a corresponding fin-to-fin spacing, *S*, and fin thickness, *t*, which meets the pressure drop requirement at the specified flow rate. Increasing fin number decreases both the fin thickness and the spacing. For a given thermal operating condition, described by pressure drop and volumetric flow rate, designs with very small and very large fin densities yield high values of thermal resistance, the former due to limited surface area and the latter caused by the highly inefficient thin fins. An intermediate geometry, which minimizes the thermal resistance, can thus be identified, where the thermal optimization illustrated in Figure 5.23 uses a heat sink volume as the driving thermal constraint and not an actual application specification.

Table 5.14: Process capabilities for different manufacturing techniques

	Conventional Processes		Modern Processes					
Parameter (see figure 5.4)	Extruded	Die-Casting	Bonding	Folding	Modified Die-Casting	Forging	Skiving	Machining
min *t*, mm	1	1.75	0.75	0.25	0.2	0.4	0.3	0.5
max *L/s*	8:1	6:1	60:1	40:1	---------	50:1	25:1	50
min *s*, mm	6.6	8.3	0.8	1.25	0.2	1	2	1
Material	Aℓ	Aℓ, Zn-Alloy	Aℓ, Cu, Mg	Aℓ, Cu	Aℓ, Zn-Alloy	Aℓ	Aℓ	Aℓ, Cu, Mg

For the conditions illustrated in Figure 5.23, a minimum value of 0.135°C/W occurs at 4 fins/cm, yielding a design that will dissipate 186W at a 25°C excess temperature from the base to the inlet air. The fin thickness and spacing for this aluminium structure are 0.57mm and 1.93mm, respectively with a heat sink mass of 0.308kg. In Table 5.14, not all the processes are capable of creating a heat sink with such dimensions with a gap aspect ratio of 26 and a fin aspect ratio of 86.

Figure 5.24 shows the typical range of cost functions for different types of heat sinks in terms of required thermal resistance R_θ.

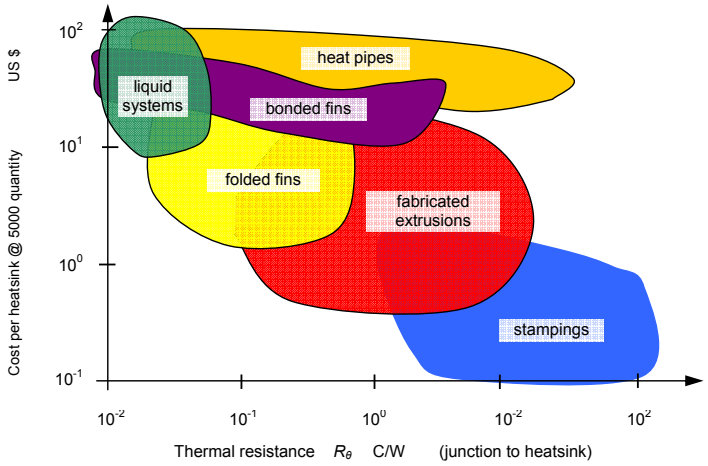


Figure 5.24. *Cost versus required thermal resistance.*

The performance of different heat sink types varies dramatically with the airflow across the surface area of the heat sink. To quantify the effectiveness of different types of heat sinks, the volumetric heat transfer efficiency is defined as

$$\eta_v = \frac{P_D}{m_f c_p \Delta T_{sa}} \tag{5.61}$$

where m_f is the mass flow rate through the heat sink, c_p is the heat capacity of the fluid, and ΔT_{sa} is the average temperature difference between the heat sink and the ambient air. The heat transfer efficiencies for a wide range of heat sink configurations, and their ranges are listed in Table 5.15. The improved thermal performance is associated with additional costs in either material or manufacturing, or both.

Table 5.15: Range of heat transfer efficiencies

Heat sink type	η_v range %
Stamping & flat plates	10-18
Finned extrusions	15-22
Impingement flow Fan heat sinks	25-32
Fully ducted extrusions	45-58
Ducted pin fin, Bonded and folded fins	78-90

5.10.5 Thermal performance graph

Typical heat sink performance graphs are shown in figure 5.25. It is assumed that the device to be cooled is correctly mounted, and the heat sink is in its normal mounting orientation with respect to the direction of airflow. The solid plot is the natural convection curve of heat sink temperature rise, ΔT_{sa}, versus P_D, which assumes that the heat sink is appropriately painted or anodized black. The dashed curve is the forced convection curve of thermal resistance versus air velocity. In forced convection, ΔT_{sa} is linearly proportional to P_D, hence $R_{\theta sa}$ is independent of P_D and becomes a function only of the flow velocity. However, the natural convection phenomenon is non-linear, making it necessary to present ΔT_{sa} as a function of P_D.
The performance graphs can be used to identify the heat sink and, for forced convection applications, to determine the minimum flow velocity that satisfy the thermal requirements. For example, if the required thermal resistance in a force convection application is 8°C/W, the thermal resistance versus flow velocity curve indicates a velocity of at least 2.4m/s (480lfm). For natural convection applications, the required thermal resistance R_{sa} can be multiplied by P_D to yield the maximum allowable ΔT_{sa}. The temperature rise of a chosen heat sink must be equal to or less than the maximum allowable ΔT_{sa} at the same P_D.

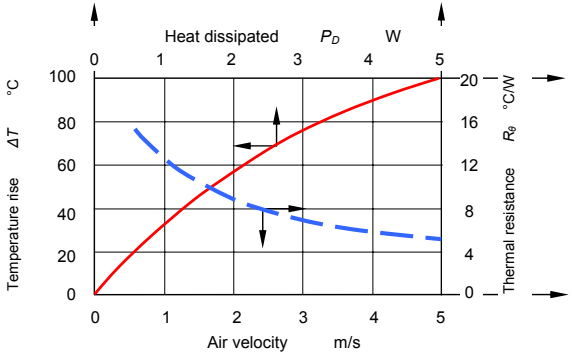


Figure 5.25. *Typical finned heatsink thermal performance graphs.*

The natural convection curves assume an optimal orientation of the heat sink with respect to the gravity. Also, the flow velocity in the forced convection graph represents the approach flow velocity without accounting for the effect of flow bypass. Flow bypass reduces the performance of a heat sink by as much as 50% for the same upstream flow velocity.
When a device is substantially smaller than the base plate of a heat sink, the thermal spreading resistance needs to be considered in the selection process. Performance graphs generally assume that the heat is evenly distributed over the entire base area of the heat sink, and therefore, do not account for the additional temperature rise caused by a concentrated heat source. This spreading resistance could typically be 5 to 30% of the total heat sink resistance.
Metal heat sinks may act as electromagnetic radiators, even when earthed. When EMC is an issue, thermal conducting plastic resins may be a viable alternative to a metal heat sink. Thermal conductivities from 5 up to 50W/mK (similar to stainless steel 15W/mK and ceramic aluminium oxide 25W/mK) are a result of additives. Such plastic heat sinks are 50% lighter than aluminium equivalents and advantageously are available in electrically insulative and electrical conductive grades.

5.11 Heatsink cooling enhancements

Other thanr natural convection cooling of heatsinks, other possible active thermal management technologies for power electronics applications include:

- Heatsink air-cooling with fans and blowers
- Enhanced air-cooling
  - Piezo fans
  - 'Synthetic' jet cooling
  - 'Nanolightning'
- Indirect liquid cooling
  - Heat pipes
  - Cold plates
- Direct liquid cooling
  - Immersion cooling
  - Liquid jet impingement
  - Spray cooling
- Microchannels and minichannels
- Electrohydrodynamic and electrowetting cooling
- Liquid metal cooling
- Solid-State cooling
  - Thermoelectric
  - Superlattice and heterostructure cooling
  - Thermionic and thermotunnelling cooling
- Phase change materials and heat accumulators

5.12 Heatsink fan and blower air-cooling

Fans are low-pressure air pumps that utilize power from an electric motor to output a volumetric flow of air at a given pressure. A propeller converts torque from the motor to increase static pressure across the fan rotor and to increase the kinetic energy of the air molecules. Electronic cooling fans draw or flow air from outside an electronic enclosure into the electronics area, expel heated air from inside the enclosure, or move air across a heat sink or electronic device to accelerate the removal of heat energy from the device.
The fan motors are typically permanent split capacitor ac induction or brushless dc.
Air moving devices are generally either a type of axial (and/or propeller) fan, figure 5.26a, or a centrifugal blower, figure 5.26b. The main difference between fans and blowers is their flow and pressure characteristics. *Axial fans* take and deliver air in an overall direction that is parallel to the fan blade axis and can be designed to deliver a high flow rate, but tend to work against low pressure. *Radial (centrifugal) blowers* tend to deliver air in a direction that is perpendicular to the blower axis at a relatively low flow rate, but against high pressure.

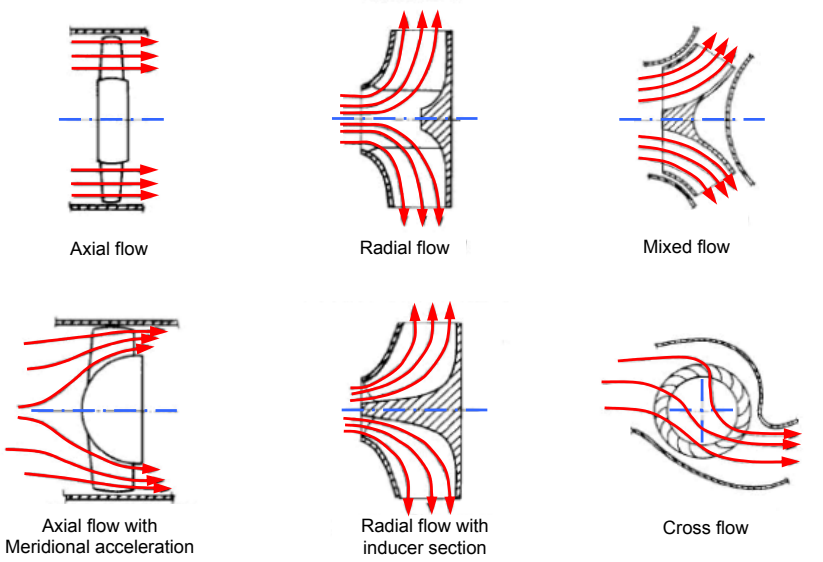
The most common fans are propeller, tube-axial and vane-axial styles.

- Propeller fans are the simplest type of fan, consisting of a motor and propeller. One problem with propeller fans is that tip vortices are produced by the pressure differential across the airfoil section. The required pressure is low.
- A tube-axial fan (the most common type in electronic cooling systems) is similar to a propeller fan, but also has a Venturi around the propeller to reduce the vortices. It develops high pressure but has a low efficiency, with peak efficiency generally occurring just before the stall dip.
- The vane-axial fan has vanes that trail behind the propeller in the airflow to straighten the swirling flow created as the air is accelerated.

Impeller types, known as flat packs, have a small aspect ratio, with good flow rate and pressure drop.

Centrifugal blowers may have a forward curved wheel, a backward curved wheel, or be of the squirrel cage variety. They tend to be quiet: noise decreases with increased number of blades and have excellent pressure drop characteristics.

Mixed flow fans combine the characteristics of both the axial fan the radial blower. The air flows in both axial and radial directions relative to the shaft. Mixed flow fans develop higher pressures than axial fans. In a *cross flow fan* the airflows in an inward direction and then in an outward radial direction.



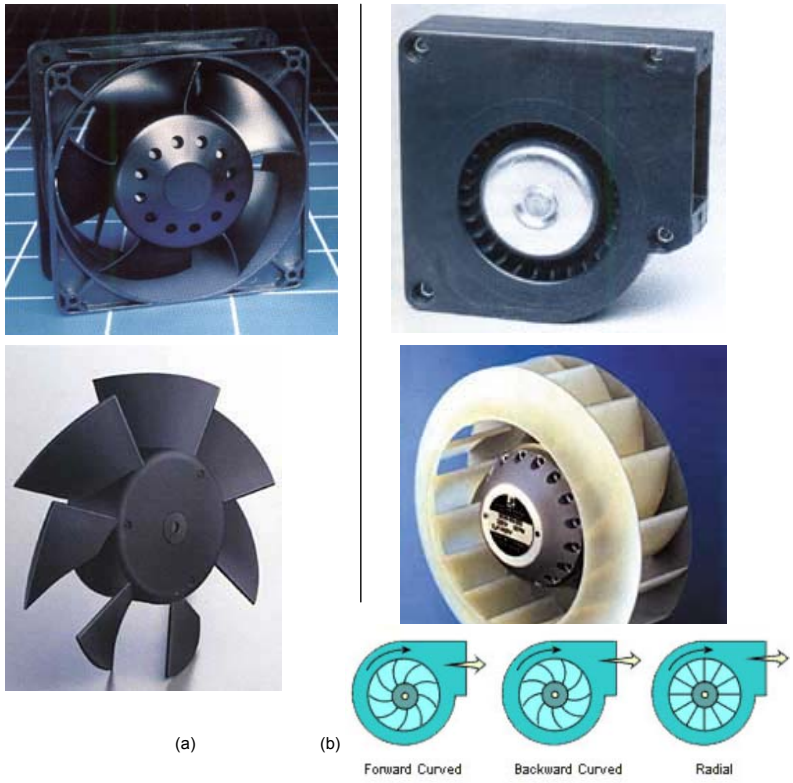


Figure 5.26. *Typical (a) impeller axial fan and (b) radial blower with backward curved blades.*

The pressure increase and the flow rate are referred appropriately to the rotational velocity ω at the perimeter, the outer diameter D_o of the impeller, and the density ρ_f of the fluid medium resulting in the dimensionless parameters 'pressure figure' Φ_f and 'volume figure', θ_f.

$$\phi_f = \frac{\Delta p}{1/2 \rho_f \omega^2} \tag{5.62}$$

$$\theta_f = \frac{V}{1/4\, \pi D_o^2 \omega} \tag{5.63}$$

The parameters allow comparison of different designs, dimensions and speed, with one another. Figure 5.27 illustrates this comparison for typical characteristic curves of the various designs, making the advantages apparent, in particularly:

- Radial fans are for a large increase in pressure and low flow rates
- Mixed-flow fans are for medium pressure and medium flow rates
- Axial fans are for high volume flow rates and low increase in pressure

The fan curve

The aerodynamic aspects of a fan are exhibited in a fan curve such as is shown in figure 5.28. Healthy aerodynamic flow is on the x-axis and rotates anticlockwise through to aerodynamic stall. A stalled fan continues to deliver air, but at an increased static pressure and a decreased volumetric flow rate, and also at the cost of an increase in noise. If noise is not a consideration, the fan can be utilized in this condition.

The fan performance curve can be understood in terms of energy. At the shut-off or no-flow point, A, the fan is in a condition of the maximum potential energy, hence produces the maximum possible pressure. At free delivery, point D, the fan is in the condition of maximum kinetic energy. Although neither of these extreme conditions are likely to occur in practice, they are useful parameters in comparing fans. The fan stall region, B, is unstable, and should be avoided. The region C to D is the stable low-pressure area of the fan, where a preferred operating point closer to D improves efficiency and compensates for filter clogging.

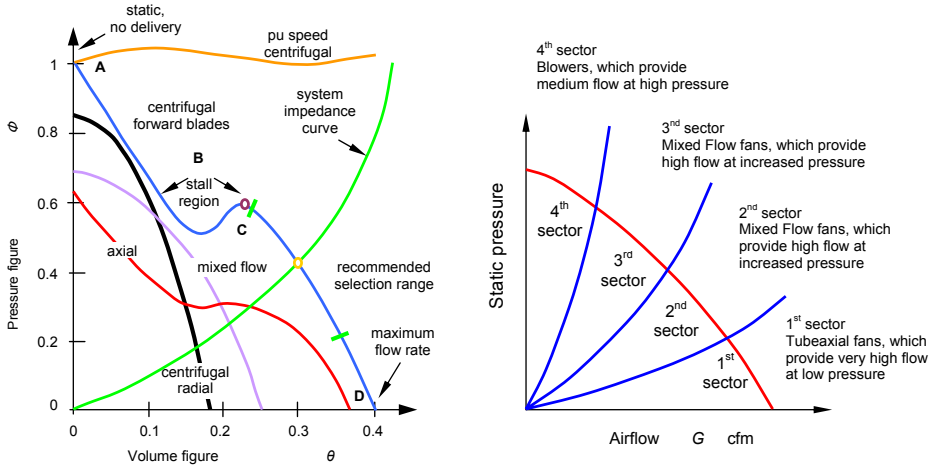


Figure 5.27. *Fan curves: (a) comparison of normalized curves for various fan designs and (b) typical fan characteristic regions.*

The governing principle in fan selection is that any given fan can only deliver one flow at one pressure in a particular system. This 'operating point' is determined by the intersection of the fan static pressure curve and the system pressure curve. Figure 5.28 illustrates the operating points of both high and low resistance systems. It is better to select a fan that gives an operating point toward the high flow, low-pressure end of the performance curve to maintain propeller efficiency and to avoid propeller stall. Each particular power electronic system should be analyzed for possible reduction in the overall resistance to airflow. Other considerations, such as available space and power, noise, reliability, and operating environment are other deciding factors.

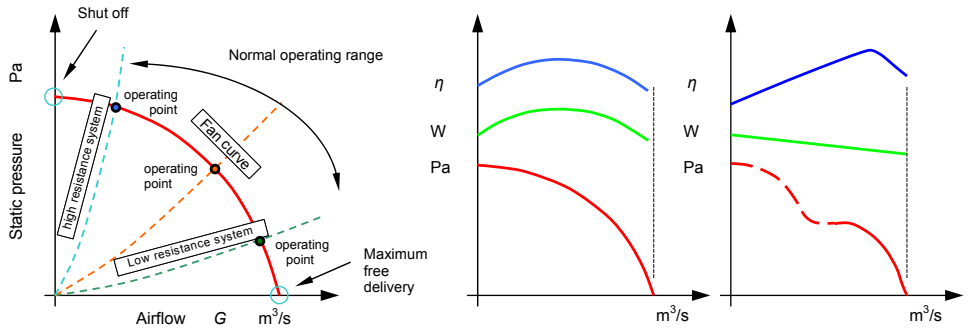


Figure 5.28. *Fan curve: (a) fan (backward curved centrifugal) system interaction; (b) input power and efficiency (backward curved centrifugal) and (c) input power and efficiency (axial).*

Acoustic noise

Sound is propagated in air by pressure waves. The effective value of pressure change is expressed relatively as sound pressure level, *SPL*, in decibels, dB. The so-called A weighting curve is commonly used and the sound pressure level obtained is expressed in dB(A).

$$SPL = 20\log\frac{P}{P_{Ref}} \tag{5.64}$$

where P is pressure and
P_{ref} is the reference pressure

Since sound pressure level varies with distance and direction to a device, it is not suitable as a fan comparison basis. By contrast, the sound power level, *PWL*, determined from sound pressure measurements, comprises all sound emissions: it is unaffected by distance to the fan noise source.

$$PWL = 10\log\frac{W}{W_{Ref}} \tag{5.65}$$

where W is the acoustic power of the source and
W_{ref} is the acoustic reference power

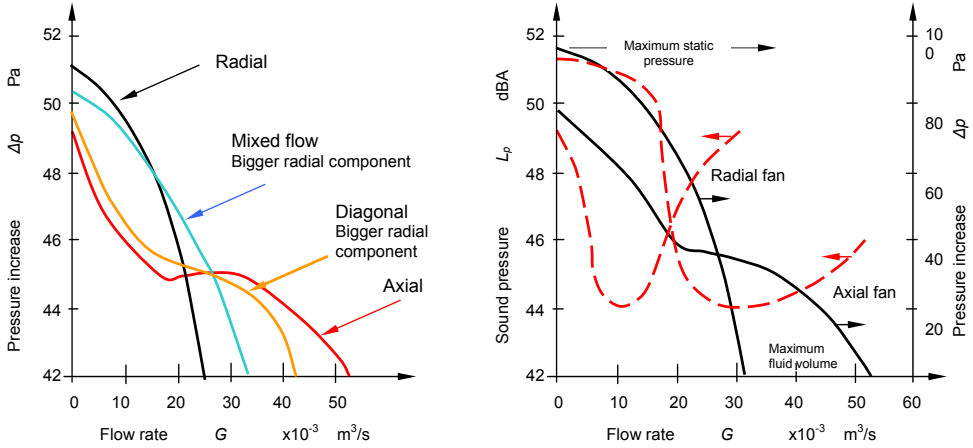


Figure 5.29. *Sound pressure level characteristics at 1m; comparison between axial and radial fans.*

The sound radiation of a fan changes with its operating state, so sound power level is only conditionally indicative for applications in which the fan does not operate under optimum conditions. A 'characteristic acoustical curve' of the fan results when the sound pressure level is measured in relation to the pressure increase or flow rate. Figure 5.29 depicts this characteristic for axial and radial fans, wherein the sound pressure level measured at a distance of 1m from the intake side of the fan is depicted as a function of flow rate.

For axial designs, a sharp increase in noise is particularly noticeable when the flow rate is excessively restricted. The axial fan enters an operating range in which the airflow no longer follows the contour of the impeller hub, resulting in additional noise.
For a specific operating point, an ideal blade geometry achieves the highest aerodynamic efficiency, which coincides with minimal noise generation. For this reason, the lowest noise generation can only be achieved for the given point of operation.

Sources of fan noise include:

- *Vortex shedding* - This is a broad-band noise source generated by air separation from the blade surface and trailing edge. It can be controlled by the blade profile design, proper pitch angle, and notched or serrated trailing blade edges.
- *Turbulence* - Turbulence is created in the airflow stream itself. It contributes to broad band noise. Inlet and outlet disturbances, sharp edges and bends will cause increased turbulence and noise.
- *Speed* - Speed is a major contributor to fan noise and its effect can be seen through the fan laws in Table 5.16.
- *Fan load* - Noise varies as the system load varies. This variation is unpredictable and fan dependent. However, fans are generally quieter when operated near their peak efficiency.
- *Structure vibration* - This can be caused by the components and mechanism within the fan, such as residual unbalance, bearings, rotor to stator eccentricity and motor mounting. Motor mounting noise is difficult to define. Cooling fans are motors and should be treated as such when mounted.

The following points will aid in the minimization of fan noise.

- *System impedance* - This should be reduced so that the least noise for the greatest airflow is obtained. The inlet and outlet ports of a cabinet can make up to between 60 and 80% of the total system impedance, which is too high for a low-noise result. Also, if a large part of the fan's flow potential is used up by the impedance of the inlet and outlet, a larger, faster and noisier fan will be required to provide the necessary cooling.
- *Flow disturbance* - Obstructions to the airflow must be avoided, especially in the critical inlet and outlet areas. When turbulent air enters the fan, noise is generated, usually in discrete tone form, adding up to 10 dB. Obstructions placed near the fan intake raise the noise level more than obstruction on the exhaust side of the fan.
- *Fan speed and size* - Most fans have several low speed versions. These should be assessed and used if possible. Various fan sizes should also be explored; quite often a larger, slower fan will be quieter than a smaller, faster fan delivering the same airflow.
- *Temperature rise* - Airflow is inversely proportional to allowable temperature rise within the system. Therefore, the ΔT limit placed on the equipment will dictate the required flow, and therefore, noise. If the temperature limit can be relaxed slightly, a noise reduction results.
- *Vibration isolation* - Fan isolation from the cabinet will avoid vibration transmission. Because fans operate at a low frequency, and are light in weight, vibration isolators must be soft and flexible. Since noise transmission is dependent on the system, experimentation is the best approach to identify quiet system/fan interaction. In systems that require 20 CFM or less, noise radiated by the cabinet is the predominant noise, and isolation of the fan is the only practical solution. Mount the fan on an interior surface of the enclosure rather than on an exterior surface. Use structural reinforcements to control enclosure resonant frequencies.

5.12.1 Fan selection

Estimate the required airflow

Before selecting a fan, it is necessary to estimate accurately the heat to be dissipated, because the overall system air temperature differential above the inlet ambient is directly proportional to the heat dissipated. Then the amount of required cooling air can be estimated.
The basic heat transfer equation is:

$$P_D = m_f \times c_p \times \Delta T \tag{5.66}$$

where P_D = amount of heat dissipated in the enclosure and transferred to cooling system, W
c_p = specific heat of air, 1.021 kJ/kg.K
m_f = mass flow rate of air through enclosure, kg/s
ΔT = desired air temperature differential (enclosure inlet to discharge ambient outside air), K

The relationship between mass flow rate and volumetric flow rate is

$$m_f = \rho \times G \tag{5.67}$$

where G = volumetric flow rate, m^3/s [1 cfm = 1.7m³/h]
ρ_ℓ = air density, kg/m^3
From equations (5.66) and (5.67), the required volumetric flow rate is calculated as

$$G = \frac{P}{\rho_\ell \times c_p \times \Delta T} \qquad (5.68)$$

This equation yields an estimate of the airflow needed to dissipate a given amount of heat at sea level. Note that it is the mass flow rate of air, not its volumetric flow rate, that governs the amount of cooling.

Estimate the actual system airflow
The actual operating airflow is determined by the intersection of the fan curve and the system resistance curve, as shown in figure 5.27b. There are three options for estimating the operating point:

- experimental measurement using a thermal/mechanical mock-up of the system,
- calculation of the operating point using airflow network methods, or
- calculation of the system airflow using computational fluid dynamics software.

The experimental procedure can be used to measure the total airflow for specific fans or several pressure-airflow data pairs can be measured to develop a complete system resistance curve. The latter experimental method requires superimpose of the selected fan pressure versus airflow curve and system resistance curve to obtain the operating airflow.
The airflow network procedure provides adequate results when the geometry is simple and the flow path within the enclosure is known or an estimate can be made.

Determining System Impedance
To estimate the system airflow, all enclosures are characterized by a system resistance curve of the type shown in Figure 5.30a. System resistance curves are expressed as a non-linear expression of pressure versus airflow:

$$\Delta P = k_{exp} \times \rho_\ell \times G^{n_q} \qquad (5.69)$$

where ΔP = system static pressure loss [1Pascal = 1N/m², and velocity pressure $P_v = \frac{1}{2}\rho_\ell v^2$]
k_{exp} = a load factor specific to the system (determined experimentally)
ρ_ℓ = density of fluid, air
G = airflow rate
n_q = airflow quality constant, which varies between 1 and 2 depending on whether the flow is completely laminar ($n_q = 1$) or completely turbulent ($n_q = 2$)

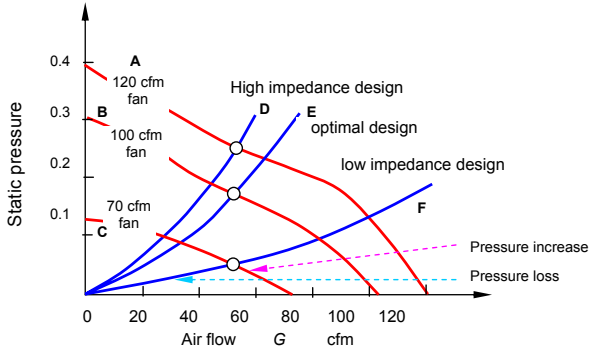


Figure 5.30. *Fan characteristics of three different fans.*

System flow
Once the volume of air and the static pressure of the system to be cooled are known, their intersection specifies the fan, specifically its airflow. To find the most effective fan for a system the typical airflow curve is divided into four sectors, as in Figure 5.27b. In general, each type of air mover will be best suited to one area, from high flow/low pressure to high pressure/low flow, as shown. Any given fan can only deliver one flow at one pressure in a given system.
Figure 5.27a shows a typical fan pressure versus flow curve along with what is considered the normal operating range of the fan. The fan, in any given system, can only deliver as much air as the system will pass for a given pressure. If the estimated value of the actual airflow is significantly less than the required value, before increasing the number of fans in a systems, or attempting to increase the air

volume using a larger fan, the system should be analyzed for possible reduction in the overall resistance to airflow. Other considerations, such as available space and power, noise, reliability and operating environment should also be considered.
If this should fail to provide a solution, a different fan or perhaps even multiple fans should be considered. The consideration of multiple fans is a more complex. An additional fan doubles the cost, doubles the noise, doubles the heat generated by fans, and may provide only a minimal improvement to the cooling, but redundant fans may increase system reliability.

Impact of varying system impedance
To demonstrate the impact of system resistance on fan performance, figure 5.30 shows three typical fans, where A is a 120 cfm fan, B is a 100 cfm fan and C is a 70cfm fan. Line D represents a system impedance within a given designed system. If 50 cfm of air are needed, fan A will meet the need. However, fan A is a high performance, higher noise fan that draws more power and is more costly. If the system impedance could be improved to curve E, then fan B would meet the 50 cfm requirement, with reduced cost, noise and power draw. If the system impedance could be optimized to where curve F were representative, then fan C meets the airflow requirement, at a dramatically lower power, noise and cost level. This is a well-designed system from a forced convection cooling viewpoint, noting that a given fan can only deliver a single airflow into a given system impedance.

Multiple fans –series and parallel operation
Combining fans in series or parallel can achieve the desired airflow without greatly increasing the system enclosure size or fan diameter.

- Parallel operation is two or more fans blowing together side by side. The performance of two fans in parallel will increase the volume flow rate (by ΔG and double at maximum delivery). The best results for parallel fans are achieved in systems with low resistance. A fan curve simulating multiple, identical fans in parallel may be constructed by scaling the fan curve airflow axis data in direct proportion to the number of fans. As figure 5.31 shows, when a system curve is overlaid on the parallel performance curves, the higher the system resistance, the less increase in flow results with parallel fan operation. Thus, this type of application should only be used when the fans can operate in a low impedance near free delivery.
- In series operation, the fans are stacked one upon the other, resulting in an increase of static pressure, ΔP, doubling at shut-off, but less elsewhere, as seen in figure 5.31. The best results for series fans are achieved in systems with high resistance. A fan curve simulating multiple, identical fans in series, may be constructed by scaling the fan curve pressure axis data in direct proportion to the number of fans.

In both series and parallel fan operation, certain areas of the combined performance curve will be unstable and should be avoided. This instability is unpredictable and is a function of the fan and motor construction and the operating point.

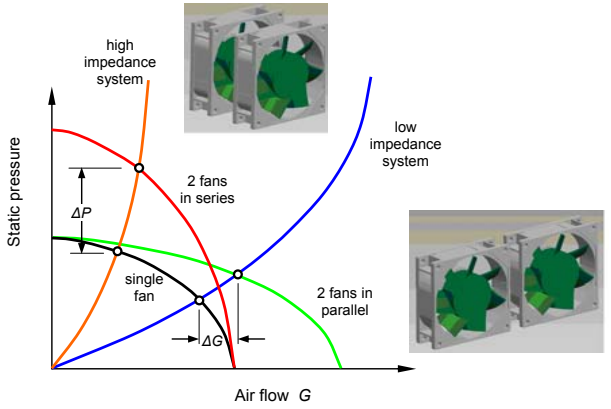


Figure 5.31. *The effects of multiple fans (series versus parallel fan operation) on system pressure and flow rate.*

5.12.2 The fan laws

It may be necessary to determine the output of a given fan under other operating conditions of speed or fluid density, or to convert the known performance of an air mover of one size to that of another geometrically similar unit of a different size. The fan laws permit this and geometrically similar fans can be characterized by the following five equations:

$$\begin{aligned} \text{Volumetric Flow-rate: } & G = K_q N D^3 \\ \text{Mass Flow Rate: } & m_f = K_m \rho_\ell N D^3 \\ \text{Pressure: } & P = K_p \rho_\ell N^2 D^2 \\ \text{Power: } & HP = K_{HP} \rho_\ell N^3 D^5 \\ \text{sound: } & L_w = 3.71 \log_{10} V_t + 0.96 \log_{10} Q - 10.8 \quad \text{(imperial units)} \end{aligned} \tag{5.70}$$

where: K_q, K_m, K_p, K_{HP} = constants for geometrically and dynamically similar operation
G = volumetric flow rate, m^3/s
m_f = mass flow rate, kg/s
N = fan impeller speed, rps
D = fan diameter, m
HP = impeller input power to rotate
ρ_ℓ = air density, kg/m^3
L_w = sound pressure level, dB
V_t = tip speed of impeller

These five fan laws apply where the fan airflow rate and pressure are independent of Reynolds's number, *Re*, specifically when

$$Re = \frac{\pi \rho_\ell N D^2}{C_R \nu} > 2 \times 10^6 \tag{5.71}$$

where C_R = correction factor
ν = absolute viscosity, Ns/m^2

From these relationships, it is possible to calculate a fan performance at a different condition. Table 5.16 is a summary of the fan law equations in a dimensionless form, useful for fan analysis.

Table 5.16: Basic fan laws, scaling

Variable	Constants	Fan Laws
Speed (N)	Diameter (D) Density (ρ_ℓ)	$G_2 = G_1 (N_2 / N_1)$ $P_2 = P_1 (N_2 / N_1)^2$ $HP_2 = HP_1 (N_2 / N_1)^3$
Diameter (D)	Speed (N) Density (ρ_ℓ)	$G_2 = G_1 (D_2 / D_1)^3$ $P_2 = P_1 (D_2 / D_1)^2$ $HP_2 = HP_1 (D_2 / D_1)^5$
Density (ρ_ℓ)	Diameter (D) Speed (N) Volumetric Flow Rate (G)	$P_2 = P_1 (p_2 / p_1)^2$ $HP_2 = HP_1 (p_2 / p_1)^5$
Density (ρ_ℓ)	Diameter (D) Mass Flow Rate (m)	$G_2 = G_1 (p_2 / p_1)$ $P_2 = P_1 (p_2 / p_1)$ $N_2 = N_1 (p_2 / p_1)$ $HP_2 = HP_1 (N_2 / N_1)^2$

$$\text{sound: } L_{w2} = L_{w1} + 55 \log_{10}\left(\frac{N_2}{N_1}\right) + 55 \log_{10}\left(\frac{D_2}{D_1}\right) + 20 \log_{10}\left(\frac{\rho_2}{\rho_1}\right)$$

Example 5.12: ***Fan laws***

A chassis uses a single 120mm fan for cooling. The maximum acceptable temperature rise in the enclosure is 15C, when it dissipates 800W. A redesign results in the power dissipation increasing to 1200W. At 800W dissipation, the 120mm fan produces a $3m^3/s$ flow rate at 3000rpm using 8W of power. What are the fan requirements at 1200W enclosure dissipation?

Solution

The new flow rate, from equation (5.68), is

$$G = \frac{P}{\rho \times c_p \times \Delta T} = \frac{0.05 \times P}{\Delta T} = \frac{0.05 \times 1200\text{W}}{15\text{K}} = 4\text{m}^3/\text{s}$$

This is, a flow increase of 1 m^3/s.
The volumetric flow rate G is given by equation (5.70)

$$G = K_q N D^3$$

That is

$$rpm_2 = \frac{G_2}{G_1} \times rpm_1$$

$$= \frac{4\text{m}^3/\text{s}}{3\text{m}^3/\text{s}} \times 3{,}000\text{rpm} = 4{,}000\text{rpm}$$

The fan power requirements are given by

$$HP = K_{HP} \rho N^3 D^5$$

That is

$$Power_2 = Power_1 \times \left(\frac{rpm_2}{rpm_1}\right)^3$$

$$= 8\text{W} \times \left(\frac{4{,}000}{3{,}000}\right)^3 = 19.0\text{W}$$

The pressure increase produced is given by

$$P = K_p \rho N^2 D^2$$

That is

$$\text{Pressure}_2 = \text{Pressure}_1 \times \left(\frac{rpm_2}{rpm_1}\right)^2$$

$$= \text{Pressure}_1 \times \left(\frac{4{,}000}{3{,}000}\right)^2 = \text{Pressure}_1 \times 1.78$$

The pressure is increased by 77.8%. The expected noise increase is $55 \times \log_{10}(4000/3000) = 6.87$dBA.

♣

Density (altitude) effects on fan performance

Since a radial fan is a constant volume machine, it will move the same volume flow of air *G* independent of the density of the air, as shown in figure 5.32. A fan is not a constant mass flow machine, therefore mass flow *m* changes as the density changes. This is important when equipment must operate at various altitudes. The mass flow *m* is directly proportional to density change $\Delta\rho_\ell$, while the volume flow *G* remains constant. As air density decreased, mass flow decreases and the effective cooling diminishes proportionately. Therefore, equivalent mass flow is needed for equivalent cooling, or the volume flow required at altitude (low-density air) will be greater than that required at sea level to obtain the equivalent heat dissipation.

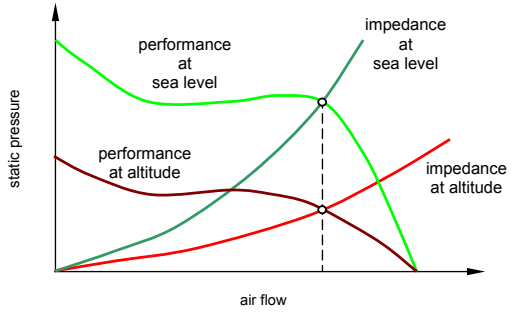


Figure 5.32. *Density effects on fan performance.*

G is volumetric flow rate; a measurement of volume over time. It pertains to no particular gas or gas density. G is strictly a rate of volume measurement. But within that volume of gas, and in this case - air, the quality of the air and its ability to transfer heat can be calculated. Every molecule of air has a mass, and this mass has the ability to absorb or emit energy; also known as transferring heat. The number of molecules for a given volume gives the density of the air (mass/volume). If more molecules of air are packed into a given volume, increasing the density, mass per volume increases and the ability to transfer heat increases; and vice versa.
At sea level, the density of air is 1.19kg/m^3, as seen Table 5.17. This value is created by all of the other molecules in the atmosphere weighing down on the molecules at sea level. As the elevation increases, there are fewer higher up molecules weighing down and the density of the air decreases. At non-sea level altitudes, the heat transfer equation is recalculated using the appropriate density for the altitude that the fan is operating at. In equation (5.68), the specific heat, c_p, which is mass dependant, is a constant for a given molecule, viz. 1.021kJ/kg.K for air.

Table 5.17: Air density ρ_ℓ change with altitude, z

Altitude Z m	ft	Density ρ_ℓ kg/m^3
Sea Level		1.19
1,524	5,000	1.06
3,049	10,000	0.904
4,573	15,000	0.771
6,098	20,000	0.652
7,622	25,000	0.549
9,146	30,000	0.458
10,671	35,000	0.379

Enclosure cooling

In addition to selecting a fan, it is important to consider fan placement in the enclosure. Pressurizing (as opposed the evacuating) the enclosure is the preferred cooling method, since incoming air can be readily filtered. In addition, a pressurized enclosure will prevent dust entering through cracks or crevices. The fan is also transferring cooler, denser air, and therefore has a slightly higher-pressure capability, which may be a slight advantage for low heat dissipating systems. An important feature of a pressurized system is that the fan life and reliability are increased due to the fan ambient temperature being lower. The disadvantage of pressurization is that heat generated by the fan is dissipated into the enclosure.
When locating the fan or fans, the enclosure layout illustration in Figure 5.33 highlight some desirable cooling aspects. The airflow path will always take the path of least resistance, so use baffles to eliminate recirculation of the same air and to direct the airflow. Importantly, air is forced in at a lower level than the outlet vents in order to benefit from the chimney effect, where less dense air rises.

- Locate components with highest heat dissipation near the enclosure air exits.
- Size the enclosure air inlet and exit vents at least as large as the Venturi opening of the fan used.
- Allow enough free area for air to pass with a velocity less than 7m/s.
- Avoid hot spots by spot cooling with a small fan.
- Locate components with the most critical temperature sensitivity nearest to inlet air to provide the coolest airflow.
- Blow air into cabinet to keep dust out, that is, pressurize the cabinet.
- Use the largest filter possible, in order to:
  - increase dust capacity
  - reduce pressure drop.

5.12.3 Estimating fan life

Convection air-cooling is the most commonly used method of cooling power electronics. In order to deliver air-cooled equipment with higher reliability, life expectancy of the air moving devices must be considered.

Definition of Fan Failure
Fan parametric failures typically include excessive vibration, noise (+ 3dBA), rubbing or hitting of the propeller, reduction in rotational speed (< 0.8×N_{nom}), increased running current (> 1.2×I_{nom}), etc. Non-functional failure include locked rotor and failure to start,
Increased noise is a result of a bearing failure, which is usually caused by a loss of lubricant, which leads to bearing wear.
The capacitor may fail in ac fans and the electronics may contribute to early failures in dc fans. Failure criteria in fan life tests can also include a change in coast-down time or start time to reach full speed. Problems with winding insulation breakdown or similar, are classified as workmanship problems or an out-of-control manufacturing process.

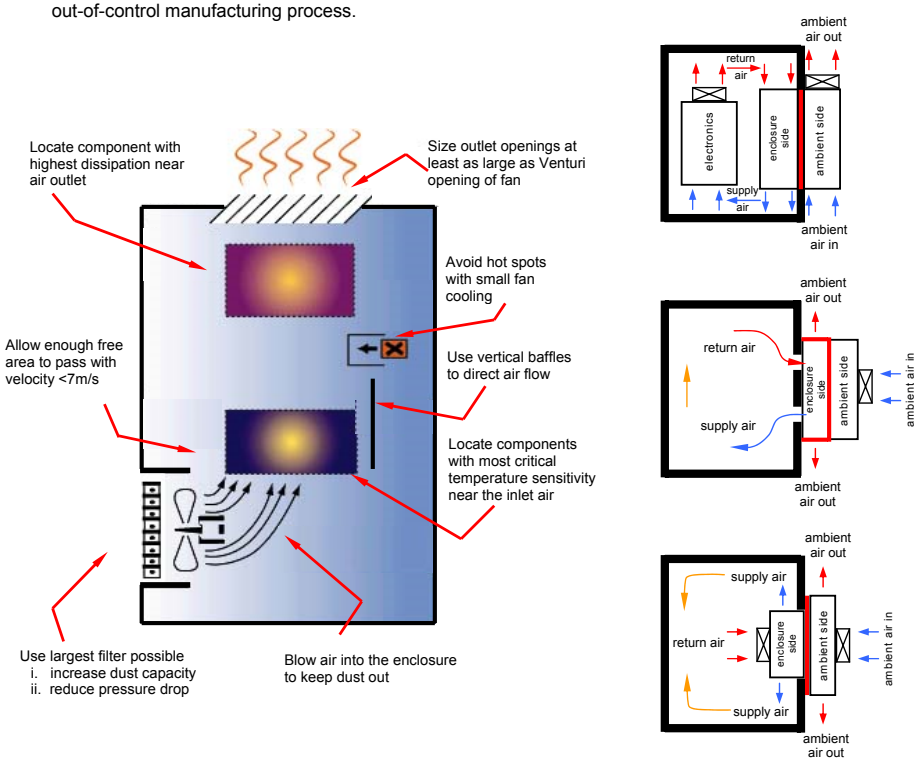


Figure 5.33. *Cabinet Cooling, external mount versus through-mount, with vertical orientation.*

Reliability concepts
Experimentation and model fitting have shown that the Weibull distribution provides a good fit to fan life data, because it accurately represents wear-out phenomena. For the Weibull distribution, the cumulative distribution function, a function of age t, is given by

$$F(t) = 1 - e^{-\left(\frac{t}{\alpha}\right)^{\beta}} \qquad (5.72)$$

where α_s is the characteristic life (for example, 9780 hours) and
$ß_s$ is the shape parameter (for example, 4.9).
Shape parameters for Weibull models fit to fan life are generally greater than 1, which means that a fan's failure tendency increases with age (wear-out). The reliability function is $1 - F(t)$, which at any age t represents the proportion of survivors from the original population. The Weibull hazard rate (also known as the failure rate or hazard function) is given by

$$H(t) = \frac{\beta_s}{\alpha_s} \times \left(\frac{t}{\alpha_s}\right)^{\beta_s - 1} \qquad (5.73)$$

Two metrics of fan reliability are the L_2 life and L_{10} life, which are the second and tenth percentiles under some assumed fan life distribution, such as the Weibull distribution. L_{10} refers to the time at which 90% of a large population of fans continue to operate. Since $F(t) = 0.1$ at L_{10} and 0.02 at L_2 in equation (5.72):

$$L_{10} = \alpha_s \times 0.10536^{1/\beta_s} \qquad L_2 = \alpha_s \times 0.02020^{1/\beta_s} \tag{5.74}$$

For example, given α_s = 100k POH (power on hours) and $\beta_s = 1.5$, $L_2 = 7,418$ hours represents the age at which 98% of the population is expected to still be operating. The advantage of specifying an L_2 life in place of L_{10} life, is that the desired early life failure distribution is more tightly specified.
Sometimes *the mean time to failure* (*mttf*) is also quote. For the Weibull distribution,

$$mttf = \alpha \times \Gamma\left(1 + \frac{1}{\beta}\right) \tag{5.75}$$

where Γ denotes the Gamma function.
The *mttf* is often confused with the *mean time between failures* (*mtbf*). The *mtbf* should only be used in a repairable system setting. If a system uses ten fans, and any failed fan is promptly replaced, then the *mtbf* may be used to understand the system's maintenance needs and service cost. But since the underlying hazard rate of the fans is not constant, computing the *mtbf* of a multiple-fan system is quite difficult. Instead, system reliability studies use a one-number hazard rate for the individual fans, in which case the average hazard rate may be appropriate.

Fan life estimation
The life of most fans is limited by the bearings. Electronics, even in dc fans, play a secondary role.
When temperatures ranged from 25-60°C, ball bearing fans on average outlasted sleeve-bearing fans by 50%. When temperatures exceeded 70°C, ball bearing fans ran for 45,000 hours, while sleeve-bearing fans became inoperable. Yet, when the ambient temperatures are relatively low, sleeve bearing fans lasted as long as ball bearing fans. Therefore, if an application generates high levels of heat, a ball bearing fan is used. If the equipment generates low heat intensities, or if the equipment has a short life span, a sleeve-bearing fan can be used.
Bearing life is generally limited by the grease life, which is primarily a function of temperature. Grease life is affected by the type of grease, percentage of grease fill, operating environment, load, and bearing design. The *Booser grease life* equation is based on grease life tests on electric motor bearings, and is valid for rolling-element bearings. The equation for the bearing grease life in the application is

$$\log L_{10} = -2.6 + \frac{k_T}{T_{brg}} - 0.301 \times S_{½} \tag{5.76}$$

where

$$S_{½} = S_G + S_N + S_P \tag{5.77}$$

$$S_N = 0.86 \times \frac{D_H \times N}{DN_L}$$

where

$$S_P = 2.95 \times \frac{D_H \times N \times P}{C_r^2}$$

P	equivalent dynamic bearing load, kg
N	speed, rpm
C_r	basic dynamic load Capacity, kg
D_H	bore diameter, mm
DN_L	speed limit, rpm-mm
$S_{½}$	half-life subtraction factor; for $S_{½} = 1$, the life falls 50%
S_G	grease half-life subtraction factor, typically 0 for many greases
S_N	speed half-life subtraction factor
S_P	load half-life subtraction factor
k_T	grease temperature factor = 2450 for acceleration factor of 1.5 for each 10°C
T_{brg}	bearing temperature, K

This equation, however, does not account for the effect of grease quantity and may not cover all available greases, particularly modern synthetic oils. For these new greases and depending upon the operating conditions, the results from the Booser equation may be conservative. Therefore, unless adjustment factors are available for a certain fan type, it is better to use the Booser equation to obtain a qualitative comparison of two fan designs rather than an absolute life estimate.

Example 5.13: ***Fan lifetime***

Calculate fan tenth percentile lifetime, L_{10}, using the following information fan data sheet information.
P = 960g, C_r = 57kg,
D = 3 mm
N = 2200 rpm, DN_L = 270,000 rpm-mm,
T_{brg} = 42°C when T_{amb} = 25°C

Solution

The half-life subtraction factor, from equation (5.77), is calculated as

$$S_N = 0.86 \times \frac{D \times N}{DN_L} = 0.86 \times \frac{3\text{mm} \times 2200\text{rpm}}{270{,}000\text{rpm-mm}} = 0.021$$

$$S_P = 2.95 \times \frac{D \times N \times P}{C_r^2} = 0.28 \times \frac{3\text{mm} \times 2200\text{rpm} \times 0.96\text{kg}}{57\text{kg}^2} = 0.540$$

thus
$S = S_G + S_N + S_P = 0 + 0.021 + 0.540 = 0.561$
$k_T = 2450/1.5^{(25\text{-}25)/10} = 2450/1.5^0 = 2450$, that is, assume no temperature derating.
From equation (5.76)

$$\log L_{10} = -2.6 + \frac{2450}{273.2°\text{C} + 42°\text{C}} - 0.301 \times 0.561$$
$$= -2.6 + 7.78 - 0.169 = 5.01$$

The resulting life estimate is L_{10} = 102,000 hours.

♣

In situations where fan reliability is critical, limit the bearing temperature rise to 10°C, particularly when a single fan failure results in a system shutdown. The Booser life estimate can also be significantly affected by the bearing load and the bearing size. Installing a fan with the shaft mounted vertically will result in a lower bearing load and a longer fan life. Using a larger bearing will also yield a longer fan life.

Fan life experiments
Because of economic and time constraints, a zero failure test strategy and/or accelerated testing techniques are adopted. A zero failure test strategy may be used to estimate the test time required to verify a life expectancy criterion such as a minimum L_{10} life. The precision of this approach depends on the accuracy of the shape parameter assumption.

Example 5.14: ***Fan testing***

How long should a sample size of 30 fans be tested to determine with 90% confidence that L_{10} is greater than or equal to 80,000 hours, at 30°C?

Solution

Assuming a Weibull distribution, each of n fans should be tested t_1 hours, with

$$t_1 = \alpha \left[\chi^2_{2;\frac{C}{2n}}\right]^{\frac{1}{\beta}}$$

where χ^2_2 is the *C-th* percentile of the Chi Square distribution with two degrees of freedom; C is determined by the desired confidence level. From a Chi Square table, $\chi^2_{2;\,0.90} = 4.60$.
Assuming $\beta = 2$, solving for α in equation (5.74) gives

$$\alpha = L_{10}\,(0.10536)^{-½} = 246{,}460\,\text{hours}$$

Substituting α, with $n = 30$ into the Weibull distribution equation, gives t_1 = 68,280 hours of test time for each fan. If all 30 fans operate t_1 hours, at 30°C, without failure, then it can be asserted with 90% confidence that L_{10} is at least 80,000 hours.

♣

Accelerated Life Testing
Since life test durations are lengthy, even when a zero failure test strategy is used, accelerated testing techniques are essential to complete component evaluation within a reasonable time and cost.
The first acceleration factor is on/off cycles. These cycles stress the fan by accelerating the bearing from zero speed to normal speed. An on/off cycle every 8 hours would be representative of a personal computer application. Even if this degree of stress is not appropriate, some on/off cycles are required to detect fan problems such as failure to start, changes in rotational speed, coast down time or start time, and increased noise.
Elevated temperature is generally the primary acceleration factor. The range of acceleration factors typically used in fan reliability calculations is 1.3 to 2 per 10°C increase. For fan failures caused by lubricant breakdown, it is reasonable to use the acceleration factor of 1.5 per 10°C increase as in Booser's equation. For example, to extrapolate the results of a life test run at 80°C down to 40°C, use a decreasing acceleration factor of $1.5^{(80\text{-}40)/10} = 5.1$.

At the accelerated life test temperature, there should not be a significant change in grease structure. The performance of the grease is degraded mainly due to evaporation loss and oxidation. Accelerated life testing should therefore be conducted at air temperatures below 85°C.
Although Booser's nominal temperature acceleration factor applies specifically at a bearing temperature of 100°C, no model exists for a room temperature of 25°C.
A reliable fan will maintain system cooling and protect against system meltdown. A quality axial fan can fulfil these requirements. Debate exists over which bearing to use in the axial fan. The bearing type is a crucial factor in determining an axial fan's reliability. Table 5.18 outlines the relatives feature of the two bearing types when used in axial fans.

Table 5.18: Fan ball and sleeve bearing comparison

Criteria	Ball Bearing	Sleeve Bearing
Fan Longevity	Longer life	Shorter life
Heat Endurance	Higher	Lower
Fan Mounting Options	Vertical, shaft centre line parallel, perpendicular	Vertical
Noise Emission	Quieter at high speeds	Quieter at low speeds in early life
Parts	Precision	Non-precision
Lubricant	Less evaporation	More evaporation
Contact	Point	Line
Cost	More expensive per unit	Less expensive per unit

In summary, if the system has a short life span, or will not generate high levels of heat, a sleeve-bearing fan can be used. However, if the application is a densely packed or a compact electronic system, a ball bearing fan will endure hotter temperatures, have a greater life span, and ultimately provide a greater long-term investment.

5.13 Enhanced air cooling

With ever increasing gravitational and volumetric power density demands, traditional air-cooling techniques have reached their limit for cooling of high-power applications. With standard fans and blowers, a maximum heat transfer coefficient of h = 150W/m^{2}K can be reached with acceptable noise levels, which is about 1W/cm^2 for a 60°C temperature difference. Using 'macrojet' impingement, theoretically 900W/m^{2}K may be reached, but with unacceptable noise levels. Dedicated non-standard fans - heat sink combinations for cooling have a maximum of about 50W/cm^2. Advanced methods to extend the useful range of air-cooling are *piezo fans*, *'synthetic' jet cooling*, and *'nanolightning'*.

Piezo fans:
Piezoelectric fans are low power, small, relatively low noise, solid-state devices that are viable thermal management solutions for a variety of portable power electronics applications in laptop computers and cellular phones. Piezoelectric fans utilize piezoceramic patches bonded onto thin, low frequency flexible blades to drive the fan at its resonance frequency. The resonating low frequency blade creates a streaming airflow directed at the electronic components.

'Synthetic' jet cooling:
Due to the periodic pulsating flow nature, synthetic jets introduce a stronger entrainment than conventional-steady jets of the same Reynolds number and more vigorous mixing between the wall boundary layers and the rest of the flow. A synthetic jet draws cool air from ambient, impinges on the top hot surface and circulates the heated air back to the ambient through the edges of the plate. A radial counter air current flow is created in the gap between the plates with hot air dispersed along the top and ambient air entering along the bottom surface.

'Nanolightning':
'Nanolightning' increases the heat transfer coefficient with 'micro-scale ion-driven airflow' using high electric fields created by nanotubes. The ionized air molecules are moved by another electric field, thereby inducing secondary airflow. Cooling a heat flux level of 40W/cm^2 is possible.

5.14 Liquid coolants for power electronics cooling

Although air-cooling continues to be the most widely used method for cooling the electronic packages, significantly higher heat fluxes (W/m^2), due to the higher heat transfer coefficient are achievable with liquid cooling. Coolants are used in both single phase and two-phase applications.

- A *single-phase* cooling loop consists of a pump, a heat exchanger (cold plate/mini- or micro-channels), and a heat sink (radiator with a fan or a liquid-to-liquid heat exchanger with chilled water-cooling). The heat source in the power electronics system is attached to the heat exchanger. The fluid does not change state: water does not change to steam.
- Liquid coolants are also used in a *change of state* or *two-phase systems*, such as heat pipes, thermo-siphons, sub-cooled boiling, spray cooling, and direct immersion systems.

A heat flux of 2kW/cm^2 can be removed through boiling water, based on water molecules turning into vapour without influencing each other, using high velocities and high pressures. Available microcoolers can handle about 1kW/cm^2. Liquid cooling for power electronics applications is generally divided into the two main categories of *indirect* and *direct* liquid cooling.

- *Indirect liquid cooling* is one in which the liquid does not directly contact the components to be cooled.
- *Direct liquid cooling* brings the liquid coolant into direct contact with the components to be cooled.

The following sections discuss indirect liquid cooling in the form of heat pipes and cold plates and direct liquid cooling in the form of immersion cooling and jet impingement.

Liquid cooling can reduce the effective thermal resistance to as low as 0.01K/W. Both oil and water (which has 4 times the thermal capacity and 770 times the density of air) are used as the coolant and the heat-sink arrangement can either be immersed in the fluid (direct), or the fluid is pumped through a hollow heat sink (indirect). The fluid heat is dissipated remotely. Fluid boiling with direct cooling should be avoided, since the creation of air bubbles can cause local hot spots. Water has the advantage of low viscosity, so can be pumped faster than mineral oil. While oil may be inflammable, water corrodes thus requiring the use of de-ionised water with an oxide inhibitor, like antifreeze (ethylene glycol). Oil emersion has the added advantage of offering possibilities of increasing the breakdown and corona voltage levels, particularly with devices rated and operated at voltages above a few kilovolts.

5.14.1 Requirements of a liquid coolant

There are many requirements for a liquid coolant for power electronics applications and vary depending on the type of application. Some of the general requirements are:

- Good thermo-physical properties (high thermal conductivity and specific heat; low viscosity; high latent heat of evaporation for two-phase application)
- Low freezing point and burst point. The burst point (or solidification temperature is the temperature associated with expansion of a freezing coolant.)
- High atmospheric boiling point (or low vapour pressure at the operating temperature) for single phase system; a narrow desired boiling point for a two-phase system
- Good chemical and thermal stability for the life of the power electronics system
- High flash point and auto-ignition temperature (sometimes non-combustibility is a requirement)
- Non-corrosive to the construction materials (metals as well as polymers and other non-metals)
- No or minimal regulatory constraints (environmentally friendly, non-toxic, and possibly biodegradable)
- Economical

The best electronics coolant is an inexpensive and non-toxic liquid with excellent thermo-physical properties and a long service life. A high flash point and auto-ignition temperature are desired so that the fluid is less susceptible to ignition. Good thermo-physical properties are required to obtain the high heat transfer coefficients and low pumping power needed for the fluid to flow through a tube or a channel.
Electrical conductivity (not mentioned in the list) of a coolant becomes important if the fluid comes in direct contact with the power electronics (such as in direct immersion cooling), or if it leaks out of a cooling loop or is spilled during maintenance and comes in contact with the electrical circuits. In certain applications, a dielectric coolant is necessary, whereas in many other applications it is not a requirement because of the very remote chance of coolant leakage (or in case of a leak, the coolant does not come in contact with the power electronics).

The various liquid coolant chemistries are divided into *dielectric* and *non-dielectric fluids* and their properties: refer to Table 5.19 and Table 5.29. Water, deionised water, glycol/water solutions, and dielectric fluids such as fluorocarbons and polyalphaolefins PAO are the heat transfer fluids most

commonly used in high performance liquid cooling applications. The heat transfer fluid must be compatible with the fluid path, offers corrosion protection or minimal risk of corrosion, and meets the application's specific requirements.

5.14.2 Dielectric liquid coolants

While the food industry might be more likely to select propylene glycol PG over ethylene glycol EG for heat transfer, the power electronics, laser, and semiconductor industries might be more likely to choose dielectric fluids over water. A dielectric fluid is non-conductive and therefore preferred over water when working with sensitive electronics. Perfluorinated carbons, such as dielectric fluid Fluorinert, are non-flammable, non-explosive, and thermally stable over a wide range of operating temperatures. Although deionised water is also non-conductive, Fluorinert is less corrosive than deionised water and therefore may be a better choice for some applications. However, water has a thermal conductivity of approximately 0.59W/m°C, while Fluorinert FC-77 has a thermal conductivity of only about 0.063W/m°C. Fluorinert is also much more expensive than deionised water.

Table 5.19: Properties of different liquid coolant chemistries at 20°C

cooling chemistry	freezing point	flash point	viscosity	thermal conductivity	specific heat	density
			υ	λ	c_p	ρ_f
	°C	°C	$\times 10^{-1}$ kg.m^{-1}.s^{-1}	W.m^{-1}.s^{-1}	J.kg.K^{-1}	kg.m^{-3}
Aromatic (DEB)	< -80	57	1	0.14	1700	860
Silicate-ester (coolanol 25R)	< -50	> 175	9	0.132	1750	900
Aliphatic (PAO)	< -50	> 175	9	0.137	2150	770
Silicone (syltherm XLT)	< -110	46	1.4	0.11	1600	850
Fluorocarbon (FC-77)	< -100	none	1.1	0.06	1100	1800
EG/water 50:50 vol	-37.8	none	3.8	0.37	3285	1087
PG/water 50:50 vol	-35	none	6.4	0.36	3400	1062
Methanol/water 40:60 wt.	-40	29	2	0.4	3560	935
Ethanol/water 44:56 wt	-32	27	3	0.38	3500	927
Potassium Formate/water 40:60 wt	-35	none	2.2	0.53	3200	1250
Ga-In-Sn	-10	none	2.2	39	365	6363

Polyalphaolefin PAO is a synthetic hydrocarbon used frequently in military and aerospace applications for its dielectric properties and wide range of operating temperatures. PAO compatible recirculating chillers are available for cold plates and heat exchangers that use PAO as the heat transfer fluid. PAO has a thermal conductivity of 0.14 W/m°C. Although dielectric fluids provide low risk liquid cooling for electronics, they generally have a much lower thermal conductivity than water and most water-based solutions.

- ***Aromatics*:** Synthetic hydrocarbons of aromatic chemistry (that is, diethyl benzene DEB, dibenzyl toluene, diaryl alkyl, partially hydrogenated terphenyl) are common heating and cooling fluids used in a variety of applications. However, these compounds cannot be classified as non-toxic. Also, some of these fluids (namely, alkylated benzene) have strong odours, which can be irritating to the personnel handling them.
- ***Silicate-ester*:** This chemistry (that is, Coolanol 25R) was widely used as a dielectric coolant in airborne radar and missile systems. These fluids have caused significant and sometimes catastrophic problems due to their hygroscopic nature and subsequent formation of flammable alcohols and silica gel. Therefore, these fluids have been replaced by more stable and dielectric aliphatic chemistry (polyalphaolefins or PAO).
- ***Aliphatics*:** Aliphatic hydrocarbons of paraffinic and iso-paraffinic type (including mineral oils) are used in a variety of direct cooling of electronics parts as well as in cooling transformers. Many petroleum based aliphatic compounds meet the criteria for incidental food contact. These petroleum-based fluids do not form hazardous degradation by-products. Most of these fluids have a non-discernible odour and are non-toxic in case of contact with skin or ingestion. As mentioned before, aliphatic PAO-based fluids have replaced the silicate-ester fluids in a variety of military electronics (and avionics) cooling applications in the last decade.
- ***Silicones*:** Another class of coolant chemistry is dimethyl- and methyl phenyl-poly (siloxane) or commonly known as silicone oil. Since this is a synthetic polymeric compound, the molecular weight as well as the thermo-physical properties (freezing point and viscosity) can be adjusted by varying the chain length. Silicone fluids are used at temperatures down to -100°C and as high as 400°C. These fluids have excellent service life in closed systems in the absence of oxygen. Also, with essentially no odour, the non-toxic silicone fluids are known to be workplace friendly. However, low surface tension gives these fluids the tendency to leak around pipe-fittings, although the low surface tension improves the wetting property. Similar to the aliphatics, high molecular weight silicone oils are used in cooling transformers.
- ***Fluorocarbons*:** Fluorinated compounds such as perfluorocarbons (that is, FC-72, FC-77) hydrofluoroethers (HFE) and perfluorocarbon ethers (PFE) have certain unique properties and can be used in contact with the electronics. First, these fluids are non-combustible and non-toxic. Some fluorinated compounds have zero ozone depleting potential and other environmental properties. Second, some of these fluids have low freezing points and low viscosities at low temperatures. However, these fluids are expensive, have poor thermal properties, some have global warming potential (greenhouse effect), and, due to the extremely low surface tension, leaks can develop around fittings.

5.14.3 Non-dielectric liquid coolants

Non-dielectric liquid coolants are often used for cooling electronics because of their superior thermal properties, as compared with the dielectric coolants. Non-dielectric coolants are normally water-based solutions. Therefore, they possess a high specific heat and thermal conductivity. De-ionized water is an example of a widely used coolant. A discussion of non-dielectric coolant chemistries follows:

- ***Water:*** Water is a suitable choice for liquid cooling applications due to its high heat capacity and thermal conductivity. It is also compatible with copper, which is one of the best heat transfer materials to use in the fluid path. Water used for cooling comes from different sources. The benefit to using facility or tap water is that it is readily available and inexpensive. However, what is important to note about facility water or tap water is that it is likely to contain impurities, which cause corrosion in the liquid cooling loop and/or clog fluid channels. Therefore, using good quality water is recommended in order to minimize corrosion and optimize thermal performance.
  Water's ability to corrode metal can vary considerably depending on its chemical composition. Chloride, for example, is commonly found in tap water and can be corrosive. Facility or tap water should not be used in liquid cooling loops if it contains more than 25ppm of chloride. The levels of calcium and magnesium in the water also need to be considered, since calcium and magnesium form scale on metal surfaces and reduce the thermal performance of the components.
  If the facility water or tap water contains a large percent of minerals, salts, or other impurities, the water can be purchased filtered or deionised water. If the tap water is relatively pure and meets recommended limits, it is still recommended that a corrosion inhibitor be added for additional protection. Phosphate is an effective corrosion inhibitor for stainless steel and most aluminium components. It is also effective for pH control. One disadvantage of phosphate, however, is that it precipitates with calcium in hard water. For copper and brass, tolyltriazole is a common and highly effective corrosion inhibitor. For aluminium, organic acids such as 2-ethyl hexanoic or sebacic acid offer protection.
- ***Deionised Water:*** Tap water meets the needs of most liquid-cooling applications. However, deionised (DI) water has chemical and electrical properties that make it the optimal choice for cooling when the liquid circuit contains micro-channels or when sensitive electronics are involved. Deionised water is water that has an extremely low concentration of ions, including sodium, calcium, iron, copper, chloride, and bromide. The lack of ions in DI water eliminates the following problems.
  First, it eliminates mineral, salts and impurities that can cause corrosion or scale formation and block the coolant flow. This will degrade cooling efficiency and system operating performance.
  Second, it eliminates the risk of electrical arcing due to static charge build up from the circulating coolant. The arcing can damage sensitive control electronics in the equipment being cooled. Compared to tap water and most fluids, deionised water has a high resistivity, and as an excellent insulator, is used in the manufacturing of electrical components where parts must be electrically isolated. However, as water's resistivity increases, its corrosivity increases. Deionised water has a pH of approximately 7.0 but quickly become acidic when exposed to air. The carbon dioxide in air dissolves in the water, introducing ions and giving an acidic pH of around 5.0. Therefore, when using water that is virtually pure, it is

necessary to use a corrosion inhibitor. When using deionised water in a recirculating chiller, special high purity plumbing is needed. The fittings should be nickel-plated and the evaporators should be nickel-brazed. When using deionised water in cold plates or heat exchangers, stainless steel tubing is recommended.
The lack of ions makes this coolant unusually corrosive. Called the 'universal solvent', DI water is one of the most aggressive solvents known. In fact, to a varying degree, it will dissolve everything to which it is exposed. Therefore, all materials in the cooling loop must be corrosion-resistant. Copper and many other common materials are not compatible with DI water and will contaminate it.

- ***Ethylene Glycol* (EG):** Commonly used as antifreeze in automotive engine cooling, EG also has found use in many industrial cooling applications, at lower temperatures. Ethylene glycol is colourless and practically odourless and is completely miscible with water. Ethylene glycol has desirable thermal properties, including a high boiling point, low freezing point, stability over a wide range of temperatures, and high specific heat and thermal conductivity. It also has a low viscosity and, therefore, reduced pumping requirements. When properly inhibited, it has a relatively low corrosivity. However, this coolant is classified as toxic and should be handled and disposed of with care. Typically, water with low chloride and sulphate ion concentration (<25ppm) is recommended. Also, a monitoring schedule should be maintained to assure that inhibitor depletion is avoided and the pH of the solution is consistent. Once the inhibitor has been depleted, the old glycol should be removed from the system and a new charge installed. Even though EG's thermal conductivity is not as high as water's, EG provides freeze protection that can be beneficial during use or during shipping. Ethylene glycol is the chemical used in automotive antifreeze, but should not be used in a cooling system or heat exchanger because it contains silicate-based rust inhibitors. These inhibitors can gel and foul, coating heat exchanger surfaces and reducing their efficiency. Silicates have also been shown to reduce significantly the lifespan of pump seals. While the wrong inhibitors can cause significant problems, the right inhibitors can prevent corrosion and significantly prolong the life of a liquid cooling loop. Inhibited glycols are recommended over non-inhibited glycols.
  As the concentration of glycol in the solution increases, the thermal performance of the heat transfer fluid decreases. Therefore, it is best to use the lowest possible concentration of inhibited glycol necessary to meet corrosion and freeze protection needs. Dow Chemical recommends a minimum concentration of 25-30% EGW4. At this minimum concentration, the ethylene glycol also serves as a bacteriacide and fungicide. With recirculating chillers, a solution of 30% ethylene glycol results in only a 3% drop in thermal performance over using water alone but will provide corrosion protection as well as freeze protection down to -15°C.
  The quality of the water used in the glycol solution is also important. Even with an inhibited glycol, water ions can cause inhibitor precipitation, resulting in fouling and corrosion.
- ***Propylene Glycol* (PG):** In its inhibited form, PG has the same advantages of low corrosivity shown by ethylene glycol. In addition, propylene glycol is considered non-toxic. Other than lack of toxicity, it has no advantages over ethylene glycol, being higher in cost and more viscous. Although EG has more desirable physical properties than PG, PG is used in applications where toxicity might be a concern. PG is generally recognized as safe for use in food or food processing applications, and can also be used in enclosed spaces.
- ***Methanol/Water*:** This is a low cost antifreeze solution, finding use in refrigeration services and ground source heat pumps. Similar to glycols, it can be inhibited to stop corrosion. This fluid can be used down to -40°C owing to its relatively high rate of heat transfer in this temperature range. Its main disadvantages as a heat transfer fluid are its toxicological considerations. It is considered more harmful than ethylene glycol and consequently has found use only for process applications located outdoors. Also, methanol is a flammable liquid and, as such, introduces a potential fire hazard where it is stored, handled, or used.
- ***Ethanol/Water*:** This is an aqueous solution of denatured grain alcohol. Its main advantage is non-toxicity. Therefore, it has found application in breweries, wineries, chemical plants, food freezing plants, and ground source heat pumps. As a flammable liquid, it requires certain precautions for handling and storage.
- ***Calcium Chloride Solution*:** Aqueous solutions of calcium chloride are used as circulating coolants in food plants. It is non-flammable, non-toxic and thermally more efficient than the glycol solutions. A 29% (by weight) calcium chloride solution has a freezing point below -40°C. The main disadvantage of this coolant is that it is highly corrosive even in the presence of corrosion inhibitors.
- ***Potassium Formate/Acetate Solution*:** Aqueous solutions of potassium formate and acetate salts are non-flammable and non-toxic as well as much less corrosive and thermally more efficient than calcium chloride solution. Therefore, even a with higher price than calcium chloride, they have found a large number of applications, in the food, beverage, pharmaceuticals, chemical and climatic chamber applications, and single-phase convection cooling of microprocessors.

- ***Liquid Metals*:** Liquid metals of Ga-In-Sn chemistry are utilized with a magneto-fluid-dynamic (MFD) pump. It utilizes the high thermal conductivity and density of the metal alloy to remove high heat flux from the heat source.

5.15 Direct and indirect liquid cooling

Application of liquid cooling for power electronics may be categorized as either indirect or direct.
Indirect liquid cooling is one in which the liquid does not contact the power electronic chips, nor the substrate upon which the chips are mounted. In such cases, a good thermal conduction path is provided from the power electronic heat sources to a liquid cooled cold-plate attached to the module surface, as shown in Figure 5.34. Since there is no contact with the electronics, water can be used as the liquid coolant, taking advantage of its superior thermo-physical properties.
Direct liquid cooling involves the components to be cooled being immersed in and in direct contact with the cooling fluid, as with oil-immersed transformers.

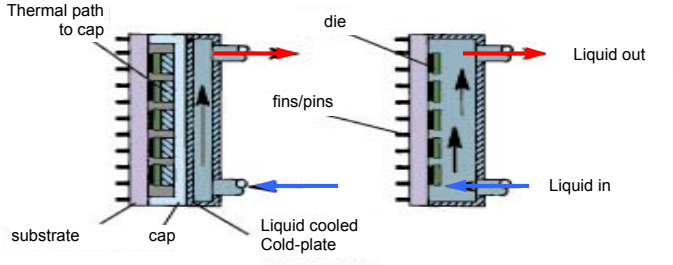


Figure 5.34. *Example of indirect and direct liquid immersion cooling for a disk package.*

5.16 Indirect liquid cooling

5.16.1 Heat pipes – indirect cooling

Heat pipes provide an indirect and passive means of applying liquid cooling. It is a two-phase device. A heat pipe can be used in situations when a heat source and a heat sink need to be placed apart. They are vacuum pumped (evacuated) and sealed vessels that are partially filled with a liquid, usually high purity water or alcohol in a saturated vapour form. The internal walls of the pipes are lined with a porous medium (the wick) that acts as a passive capillary pump. When heat is applied to one end of the pipe, the liquid starts evaporating. A pressure gradient exists causing the vapour to flow to the fractionally cooler regions. The vapour condenses (with the latent heat of vaporisation transferred to the condenser) back to the liquid state at the cooler regions and is transported back by the capillary wick structure (or gravity), thereby closing the fluid two-state thermodynamic loop.

This thermodynamic cycle can be summarised, with the aid of figure 5.35, as follows:

Stage 1-2: Heat applied to evaporator through the external source vaporizes the working fluid to a saturated (2') or superheated (2) vapour.
Stage 2-3: Vapour pressure drives vapour through adiabatic section to condenser.
Stage 3-4: Vapour condenses, releasing heat to a heat sink.
Stage 4-1: Capillary pressure created by wick menisci pumps condensed fluid into evaporator section.

Then the process starts over.

Heat pipes provide an enhanced means of transporting heat, much better than copper, from a source to a heat sink where it is rejected to the cooling medium by natural or forced convection. The effective thermal conductivity of a heat pipe can range from 50kW/mK to 200kW/mK, but is often lower in practice due to additional interface thermal resistances. The performance of heat pipes scales from 10 W/cm^2 to over 300 W/cm^2. A simple water-copper heat pipe has an average heat transfer capacity of 100 W/cm^2, a thermal conductivity that is in excess of 300 times better than that of an equivalently sized pure copper component. A heat pipe provides efficient transport of concentrated heat. An example of a typical application of a heat pipe for an electronics cooling application is given in Figure 5.36.

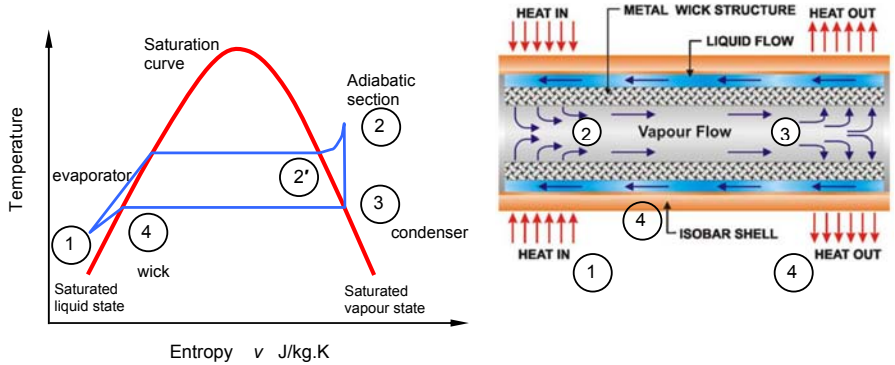


Figure 5.35. *Heat pipe thermodynamic T-v operational vapour power cycle.*

Heat pipes are an efficient, reliable, passive, silent, high thermal conducting, evacuated, sealed cylindrical device for quickly extracting, transporting, and remotely dissipating heat, in any orientation.
Although there is virtually no limit to the size of a heat pipe, the effectiveness of a heat pipe decreases with decreasing lengths. For heat pipes with a length less than about 1 cm the performance of a solid piece of metal, copper, is comparable. They are effective as efficient heat conductors to transport heat to locations were more area is available. 2D heat spreaders (otherwise known as vapour chambers) based on the heat pipe principle can achieve much higher effective thermal conductivities than copper. A thin planar heat spreader has a thermal performance greater than diamond.
Loop heat pipes (LHP) have the advantage over conventional heat pipes that the vapour and liquid paths are separated enabling much better performance of the liquid return loop, accommodate a heat flux of 625W/cm^2.

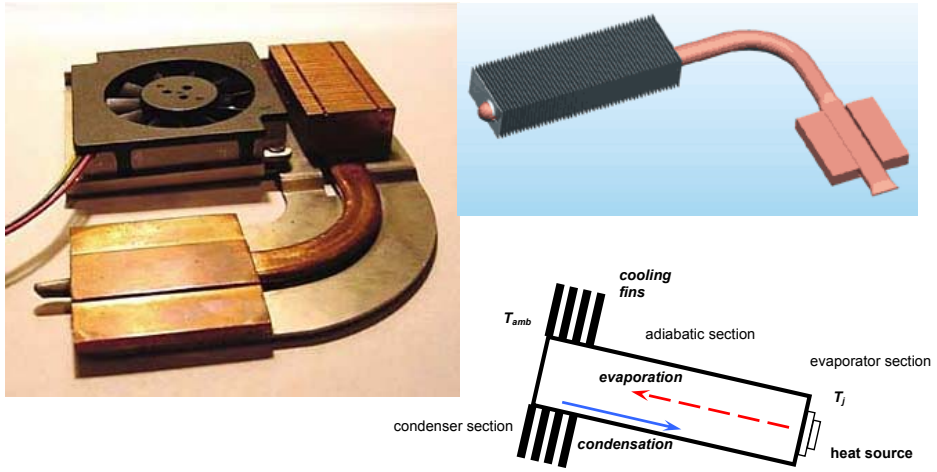


Figure 5.36. *Example of heat pipes used in a notebook application.*

A heat pipe is designed for a certain temperature range. Apart from the vapour temperature range, factors like thermal stability and thermal conductivity influence the choice of working fluid. Suitable working fluids, which desirably have a high surface tension, include:

- For ultra low temperatures: inert gases (helium), nitrogen, ammonia
- For usual temperatures to meet power electronics cooling requirements: distilled water with various additions, organic fluids like acetone, methanol, ethanol, toluene.
- For high temperatures: metals like mercury, sodium, silver.

The component to be cooled is mounted on the evaporator end (the hot end), where the heat boils and expands the liquid to the vapour phase, increasing the pressure. Boiling occurs because energy, in the form of heat, is taken from the surrounding area, which cools the heat source. This vapour rises through the adiabatic tube section (low-pressure drop, hence low temperature change) to the remote condenser end of the tube (the cold end), taking the heat within it. Effectively the heat is transported at the rate *P*:

$$P = L\frac{dm}{dt}$$

where *L* is the heat of vaporization per unit mass and *dm/dt* is the mass evaporation rate.

The vapour condenses back to the liquid phase, releasing its latent heat of vaporisation, and creating a pressure gradient, which helps draw more vapour towards the condenser. The temperature difference between the ends may only be a couple of degrees. The remotely situated condenser end is connected to an external heatsink or a radiator type grill, for cooling. The condensed working fluid runs back to the evaporator end due to gravity, or due to capillary pressure action, along porous capillaries that form a wick, depending on the physical application orientation design for the heat pump.

The quality and type of wick usually determines the performance of the heat pipe.
The wick is a porous structure made of materials like steel, aluminium, nickel or copper in various ranges of pore sizes. They are fabricated using metal foams, and more commonly, felts. By varying the pressure on the felt during assembly, various pore sizes can be produced. By incorporating removable metal mandrels, an arterial structure can be moulded in the felt.
The two most important properties of a wick are the pore radius and the permeability. The pore radius determines the pumping pressure (the maximum capillary head) the wick can develop. The wick permeability increases with increasing pore size. The permeability determines the frictional losses of the fluid as it flows through the wick. The heat transport capability of the heat pipe is raised by increasing the wick thickness. The overall thermal resistance at the evaporator also depends on the conductivity of the working fluid in the wick.

There are several types of wick structures available including (decreasing permeability and decreasing pore radius): grooves, screen, cables/fibred, and sintered powder metal.

The most common types of wicks (capillaries), shown in figure 5.37 and summarized in Table 5.20, that are used are:

Sintered powder: This wick provides high power handling, low temperature gradients and high capillary forces for anti-gravity applications. The complex sintered wick has several vapour channels and small arteries to increase the liquid flow rate. Tight bends in the heat pipe can be achieved with this type of structure.

Grooved tube: The small capillary driving force generated by the axial grooves is adequate for low power heat pipes when operated horizontally, or with gravity assistance. The tube can be readily bent. When used in conjunction with screen mesh the performance can be enhanced.

Screen mesh: This type of wick is the most common and provides readily variable characteristics in terms of power transport and orientation sensitivity, according to the number of layers and mesh counts used.

Fibre/spring: Fibrous materials, like ceramics, generally have smaller pores. The main disadvantage of ceramic fibres is minimal stiffness and requires support by a metal mesh. Carbon fibre filaments have many fine longitudinal grooves on their surface, have high capillary pressures, are chemically stable, and show a greater heat transport capability.

Table 5.20: Wick material properties

Wicking Material	Conductivity (Straight)	Overcome Gravity	Thermal Resistance	Stability	Conductivity lost (bended and flatten)
Axial Groove	Good	Poor	Low	Good	Low/Average
Screen Mesh	Average	Average	Average	Average	Low/average
Fine Fibre	Poor	Good	High	Poor	Average
Sintering (powder)	Average	Excellent	High	Average	High

Grooved wicks have a large pore radius and a high permeability, as a result, the pressure losses are low but the pumping head is also low. Grooved wicks can transfer high heat loads in a horizontal or gravity aided position, but cannot transfer large loads against gravity. The powder metals have small pore radii and relatively low permeability. Powder metal wicks are limited by pressure drops in the horizontal position but can transfer large loads against gravity.

An advantage of the sintered powder wick is its ability to handle high heat fluxes. Since sintered powder wicks are generally 50% porous, there is accordingly a large surface area available for evaporation. Typical sinter powder wicks handle 50 W/cm^2, and up to 250 W/cm^2. In comparison, a groove wick nominally handles 5 W/cm^2 and a screen wick will nominally handle 10 W/cm^2. Since a sintered powder wick is integral with the heat pipe envelope, and the fluid charge is only enough to saturate the wick, the heat pipe can be subjected to freeze/thaw cycles with no degradation in performance.

Working fluid or coolant

A first consideration in the identification of a suitable working fluid is the operating vapour temperature range. Within the approximate temperature band, several possible working fluids may exist, and a variety of characteristics must be examined in order to determine the most acceptable of these fluids for the application considered. The prime requirements are:

- compatibility with wick and wall materials
- good thermal stability
- wettability of wick and wall materials
- vapour pressure not too high or low over the operating temperature range
- high latent heat
- high thermal conductivity
- low liquid and vapour viscosities
- high surface tension
- acceptable freezing or pour point

The typical temperature operating range is within the bounds -55°C to over 200°C, depending on the coolant, as shown in Table 5.21. Heat pipes can be designed to operate over a broad range of temperatures from cryogenic (< -243°C) applications utilizing titanium alloy/nitrogen heat pipes, to high temperature applications (>2000°C) using tungsten/silver heat pipes. In power electronic cooling applications where junction temperatures below 125-150°C are desired, copper/water heat pipes are used. Copper/methanol heat pipes are used if the application requires heat pipe operation (and importantly, start up) below 0°C. Water heat pipes, with a temperature range from 5 to 230°C, are less sensitive than methanol to orientation, and are most effective for power electronics cooling applications, plus copper vessels are compatible with water.

Table 5.21: Heat pipe fluids, in increasing operating temperature range

Medium	Melting point	Boiling pt @ atm pressure	Measured axial heat flux	Measured surface heat flux	Vessel material	Operating range
	°C	°C	kW/cm^2	W/cm^2		°C
Helium (ℓ)	- 271	- 261				-271 to -269
Nitrogen (ℓ)	- 210	- 196	0.067 @ -163°C	1.01 @ -163°C	Stainless steel	-203 to -100
Ammonia (ℓ)	- 78	- 33	0.295	2.95	Ni, Al, stainless steel	-60 to 100
Acetone	- 95	57			Cu, Ni, Al, stainless steel	0 to 120
Methanol	- 98	64	0.45 @ 100°C	75.5 @ 100°C	Cu, Ni, stainless steel	0 to 120
Flutec PP2	- 50	76				10 to 160
Ethanol	- 112	78				0 to 130
Water	0	100	0.67 @ 200°C	146 @ 170°C	Cu, Ni	10 to 210
Toluene	- 95	110				50 to 200
Mercury	- 39	361	25.1 @ 360°C	181 @ 360°C	Stainless steel	220 to 650
Potassium			5.6 @ 750°C	181 @ 750°C	Stainless steel, Ni	400 to 900
Sodium	98	892	9.3 @ 850°C	224 @ 760°C	Stainless steel, Ni, Nb	500 to 1200
Lithium	179	1340	2 @ 1250°C	207 @ 1250°C	W, Ti, Niobium +1% Zirconium	900 to 1800
Silver	960	2212	4.1	413	Tantalum +5% Tungsten	1600 to 2300

The capillary or lifting height H is given by

$$H = \frac{2\gamma \cos\theta}{r_c g_f \rho_\ell} \tag{5.78}$$

where γ is the surface tension, N/m, unique for each working fluid at certain temperatures,
θ is the contact angle, rad, differs for each working fluid,
r_c the effective capillary radius, m, is a unique characteristic of the wick type used,
g_f is the gravitational acceleration, m/s^2, assumed constant, and
ρ_ℓ is the working fluid density, kg/m^3, dependant on the application conditions.

The heat power transfer capabilities of a heat pipe are related to its cross-sectional area A_x and length ℓ according to equation (5.3)

$$P_D = k\frac{A_x}{\ell} \quad \text{(W)} \tag{5.79}$$

while the temperature difference ΔT between the hot and cold ends is

$$\Delta T = k' P_D \left(\frac{1}{A_e} + \frac{1}{A_c}\right) \quad \text{(K)} \tag{5.80}$$

where A_e and A_c are the effective evaporator and condenser surface areas.

Heat pipes have no moving parts its mttf is estimated to be over 100,000 hours of use. Improper bending and flattening of the pipe may also cause the leakage on the pipe seal. There are some external factors that may also shorten the life of a heat pipe such as shock, vibration, force impact, thermal shock and corrosive environment.

Unlike Peltier elements (see 5.21.1):

- a heat pipe does not consume energy or produce heat itself and
- it is not possible to cool a device below ambient temperature using a heat pipe.

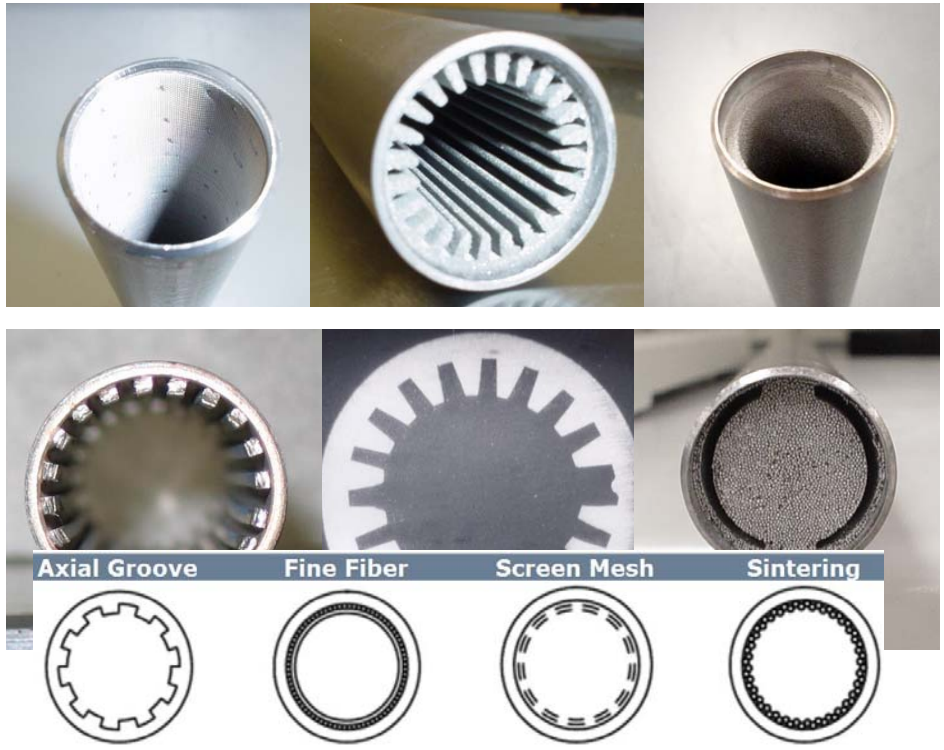


Figure 5.37. *Examples of heat pipes wicks.*

A solid rod of copper conducts heat by diffusion, and a constant, geometry-independent thermal conductivity can be defined for the material. A heat pipe, however, conducts heat by transport of the vapour and therefore is more correctly thought of as a heat carrying device rather than a simple thermal resistor. A heat pipe carries heat current with a certain temperature drop ΔT, but ΔT is virtually independent of the length of the heat pipe. Thus, heat pipes not only have low thermal resistance R (small ΔT) but also have the property that R is roughly independent of length. The advantage of using a heat pipe, rather than a rod of Cu metal, therefore increases with increasing length. The heat transfer or transport capacity of a heat pipe is specified by its *Axial Power Rating* APR. This is the energy moving axially along the pipe. The larger the heat pipe diameter, the higher the APR, while the longer the heat pipe, the lower the APR.
The fact that the heat pipe functions at all in an 'upside-down' orientation, as shown in figure 5.40, is due to the wicking of the condensed fluid against gravity.

There are five primary heat-pipe heat transport limitations, summarized in Table 5.22, which are a function of the heat pipe operating temperature, as shown in figure 5.38:

- *Viscous Limit* – Heat pipes will not function when the pipe temperature is lower than the freezing point (zero viscosity) of the working fluid. At low temperatures (just above freezing), the vapour pressure difference between the condenser and the evaporator may not be enough to overcome viscous forces. The vapour from the evaporator cannot move to the condenser and thus the thermodynamic cycle does not occur. Freezing and thawing may destroy the sealed joint of a heat pipe when place vertically.
- *Sonic limit* – The rate that vapour travels from the evaporator to the condenser. The limit occurs when the vapour velocity reaches sonic speed at the evaporator and any increase in pressure difference cannot speed up the flow. This usually occurs during heat pipe start-up.
- *Capillary pumping limit* – the rate at which the working fluid travels from the condenser to the evaporator through the wick. This limit occurs when the capillary pressure is too low to provide enough liquid to the evaporator from the condenser. Leads to dry-out in the evaporator, which prevents the thermodynamic cycle from continuing and the heat pipe no longer functions properly.
- *Entrainment or flooding limit* – Friction between the working fluid and vapour that travel in opposite directions. At high vapour velocities, droplets of liquid in the wick are extracted from the wick and deposited into the vapour. This results in dry-out.
- *Nucleated boiling limit* – the rate at which the working fluid vaporizes from the added heat. This limit occurs when the radial heat flux into the heat pipe causes the liquid in the wick to boil and evaporate causing dry-out.

Each limit has its own particular range of importance. However, in practice, the capillary and boiling limits are the most important. Figure 5.38 illustrates these limitation boundaries.

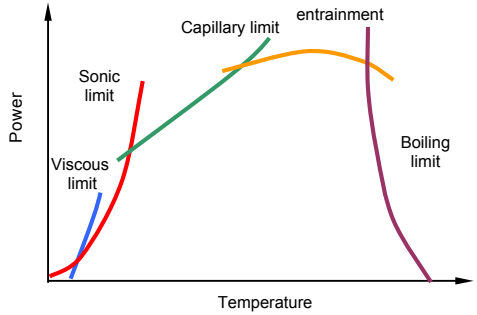


Figure 5.38. *Heat transport limitation boundaries of a heat pipe.*

For a heat pipe to function, the net capillary pressure difference between evaporator (heat source) and condenser (heat sink) must be greater than the sum of all pressures losses occurring throughout the liquid and vapour flow paths. This relationship, termed the *capillary limitation*, is expressed as:

$$\Delta P_{max} \geq \Delta P_{liquid} + \Delta P_{vapour} + \Delta P_g \qquad (5.81)$$

where ΔP_{cmax} is the maximum capillary pressure difference generated within the capillary wicking structure between the evaporator and condenser,
ΔP_{liquid} and ΔP_{vapour} are the viscous pressure drops occurring in the liquid and vapour phases, respectively, and
ΔP_g represents the hydrostatic pressure drop.

Table 5.22: Heat-pipe mechanisms and limitations

Heat Transport Limit	Description	Cause	Potential Solution
Viscous	Viscous forces prevent vapour flow in the heat pipe	Heat pipe operating below recommended operating temperature	Increase heat pipe operating temperature or find alternative working fluid
Sonic	Vapour flow reaches sonic velocity when exiting heat pipe evaporator resulting in a constant heat pipe transport power and large temperature gradients	Power/temperature combination, too much power at low operating temperature	This is typically only a problem at start-up. The heat pipe will carry a set power and the large ΔT will self correct as the heat pipe warms up
Capillary	Sum of gravitational, liquid and vapour flow pressure drops exceed the capillary pumping head of the heat pipe wick structure	Heat pipe input power exceeds the design heat transport capacity of the heat pipe	Modify heat pipe wick structure design or reduce power input
Entrainment/Flooding	High velocity vapour flow prevents condensate from returning to evaporator	Heat pipe operating above designed power input or at too low an operating temperature	Increase vapour space diameter or operating temperature
Boiling	Film boiling in heat pipe evaporator typically initiates at 5-10 W/cm^2 for screen wicks and 20-30 W/cm^2 for powder metal wicks	High radial heat flux causes film boiling resulting in heat pipe dry-out and large thermal resistances	Use a wick with a higher heat flux capacity or spread out the heat load

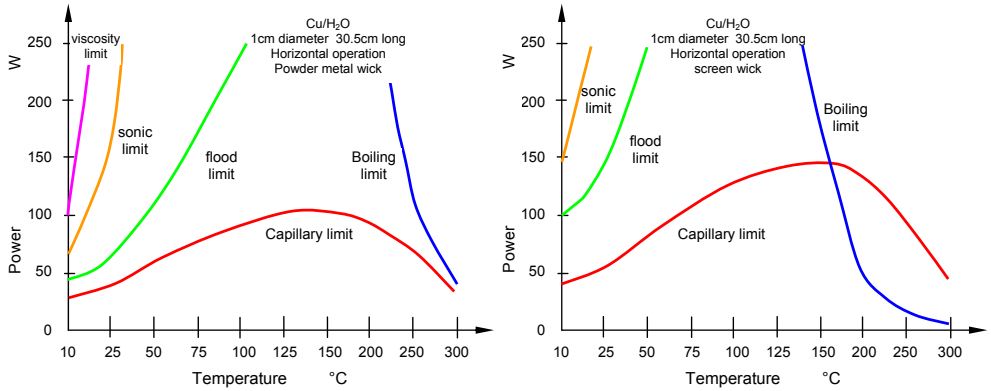


Figure 5.39. *Predicted heat pipe limitations, where the capillary limit is usually the limiting factor: (a) powder metal wick and (b) screen wick (viscosity limit off-scale). Note temperature scale entrainment.*

When the maximum capillary pressure is equal to or greater than the sum of these pressure drops, the capillary structure returns an adequate amount of working fluid (priming or repriming of the heat pipe) to prevent the evaporator wicking structure from drying out. When the sum of all pressure drops exceeds the maximum capillary pumping pressure, the working fluid is not supplied rapidly enough to the evaporator to compensate for the liquid loss through vaporization, and the wicking structure becomes starved of liquid and dries out (depriming of the heat pipe). This condition, referred to as *capillary limitation*, varies according to the wicking structure, working fluid, evaporator heat flux, operating temperature, and body forces.

Figure 5.39 shows graphs of the axial heat transport limits as a function of operating temperature for typical powder metal and screen wicked heat pipes.
The capillary limit is usually the limiting factor in a heat pipe design, when used in its optimal temperature range, and the limit is set by the pumping capacity of the wick structure. As shown in Figure 5.40, the capillary limit is a strong function of the operating orientation and the type of wick structure.

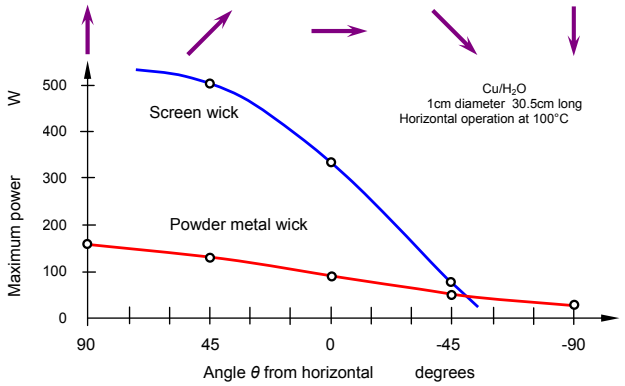


Figure 5.40. *Dependence of the maximum stable heat flow on orientation, specifically capillary limits versus operating angle. Arrows indicate the direction of the heat flow.*

Heat pipe effective thermal resistance

A primary heat pipe design consideration is effective pipe thermal resistance or overall heat pipe ΔT at a given design power. As the heat pipe is a two-phase heat transfer device, a constant effective thermal resistance value cannot be assigned. The effective thermal resistance is not constant but a function of many variables, such as heat pipe geometry, evaporator length, condenser length, wick structure, and working fluid.

The total thermal resistance of a heat pipe is the sum of the resistances due to

- conduction through the wall,
- conduction through the wick,
- evaporation or boiling, axial vapour flow,
- condensation, and
- conduction losses back through the condenser section wick and wall.

Because heat pipes are two-phase heat transfer devices that do not have relatively constant thermal conductivities like solid materials, an effective thermal conductivity is used. The equation used to calculate the effective thermal conductivity for a heat pipe is:

$$\lambda_{eff} = \frac{P_T \times L_{eff}}{A_x \times \Delta T} = \frac{1}{R_{ij}} \times \frac{L_{eff}}{A_x} \qquad (5.82)$$

where: $L_{eff} = ½(L_{evaporator} + L_{condenser}) + L_{adiabatic}$

λ = thermal conductivity

A_x = the cross-sectional area of the heat pipe

P_T = power transported by the heat pipe

ΔT = the measured temperature difference across the heat pipe.

A few rules of thumb can be used for first pass design considerations. A rough guide for a copper/water heat pipe with a powder metal wick structure is to use 0.2°C/W/cm^2 for thermal resistance at the evaporator and condenser, and 0.02°C/W/cm^2 for axial resistance.

The evaporator and condenser thermal resistances are based on the outer surface area of the heat pipe. The axial thermal resistance is based on the cross-sectional area of the vapour space. This design guide is only useful for powers at or below the design power for the given heat pipe.

These equations for thermal performance are only rule of thumb guidelines and should only be used to help determine if heat pipes will meet your cooling requirements, not as final design criteria.

It can be seen that the groove heat pipe has the lowest capillary limit among the four but functions best under gravity-assisted conditions.

Example 5.15: Heat-pipe

To calculate the effective thermal resistance for a 1.27 cm diameter copper/water heat pipe 30.5 cm long with a 1 cm diameter vapour space, the following assumptions are made.

The heat pipe is dissipating 75W with a 5cm evaporator and a 5cm condenser length.

Solution

The evaporator heat flux p equals the power P_T divided by the heat input area

$$p_{evap} = \frac{P_T}{A_{evap}} = \frac{P_T}{\pi D_o L_{evap}} = \frac{75\text{W}}{\pi \times 1.27\text{cm} \times 5\text{cm}} = 3.8\text{W/cm}^2$$

$$p_{cond} = \frac{P_T}{A_{cond}} = \frac{P_T}{\pi D_o L_{cond}} = \frac{75\text{W}}{\pi \times 1.27\text{cm} \times 5\text{cm}} = 3.8\text{W/cm}^2$$

The axial heat flux equals the power divided by the cross sectional area of the vapour space

$$p_{axial} = \frac{P_T}{A_{axial}} = \frac{P_T}{¼\pi D_i^2} = \frac{75\text{W}}{¼\pi \times 1.0^2} = 95.5\text{W/cm}^2$$

The temperature gradient equals the heat flux times the thermal resistance.

$$\Delta T = p_{evap} \times R_{evap} + p_{axial} \times R_{axial} + p_{cond} \times R_{cond}$$

$$\Delta T = 3.8\ \text{W/cm}^2 \times 0.2°\text{C/W/cm}^2 + 95.5\ \text{W/cm}^2 \times 0.02°\text{C/W/cm}^2 + 3.8\ \text{W/cm}^2 \times 0.2°\text{C/W/cm}^2$$

$$\Delta T = 3.4°\text{C}$$

♣

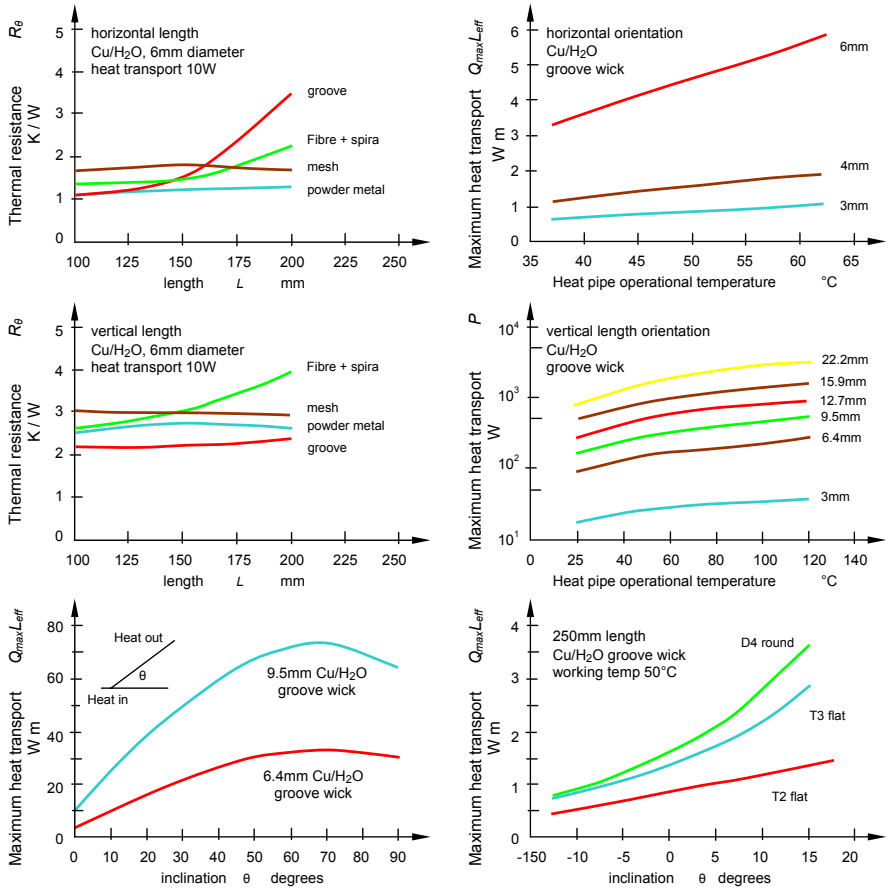


Figure 5.41. *Heat pipe parameter performance characteristics.*

Length and diameter affect the performance of the heat pipe

The rate of vapour travelling from the evaporator to the condenser is governed by the difference in vapour pressure between them. It is also affected by the diameter and the length of the heat pipe. In a large diameter heat pipe, the large cross sectional area allows a higher vapour volume to be transported from the evaporator to the condenser, than in a small diameter pipe. The cross sectional area of a heat pipe is a direct function of both the sonic and entrainment limit of the heat pipe. However, the operational temperature of the heat pipe also affects the sonic limit of the heat pipe. Fig 5.41 compares the heat transport for heat pipes with different diameters and shows that the heat pipes transport more heat at higher operational temperatures.

The rate at which the working fluid returns from the condenser to the evaporator is governed by the capillary limit and is a reciprocal function of the heat pipe's length. A longer heat pipe transports less heat than shorter heat pipes. In Fig 5.41, $Q_{max}L_{eff}$ (W-m) (that is, Y-axis) represents the amount of heat a pipe will carry per meter length. Therefore, if the pipe is half a meter, it can carry twice the wattage a metre long heat pipe would carry.

Orientation affect on performance of heat pipe performance

A wick structure with a higher capillary limit can transport more working fluid from the condenser to the evaporator against gravity. But as previously mentioned, the groove heat pipe, with the lowest capillary limit, works best under gravity-assisted conditions where the evaporator is located below the condenser. Fig 5.41 shows the effect of gravity on groove wick heat pipes.

Affect on performance if the heat pipe flattening or bending

If a heat pipe is flattened or bent, the sonic limit and entrainment limit will be reduced in relation to the flattened thickness, the number of bends and the angle of each bend. Therefore, any flattening or bending to a heat pipe will reduce the amount of heat that can be transported. Fig 5.41 shows the effect of flattening on a heat pipe.

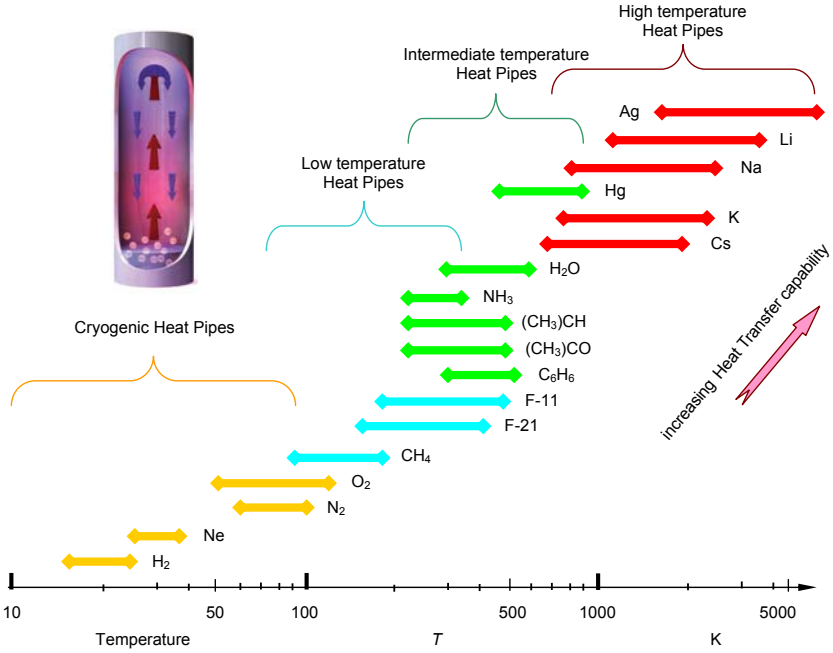


Figure 5.42. *Operating temperature ranges for various heat pipe fluids.*

Operating temperature range

The operating temperature ranges of heat pipes are referred to as *Cryogenic* (0 to 100K), 'Low Temperature' (100 to 250K), 'Intermediate Temperature' (250 to 600K), and 'High Temperature" (600 to 3000K). Working fluids are elemental gases in the cryogenic range, mainly polar molecules or halocarbons in the low temperature range, simple organic molecules in the intermediate temperature range, to liquid metals in the high temperature range.

The approximate useful heat pipe fluid range of some working fluids is indicated in Figure 5.42. Also indicated are the limits of the four defined temperature regimes. The range limits are approximate since some of the fluids overlap into the next temperature range.

Low temperature heat pipe technology

Low order hydrocarbons are common low temperature heat pipe fluids, as specifically shown for ethane and ammonia (in an aluminium alloy) based constant conductance heat pipes in figure 5.43.

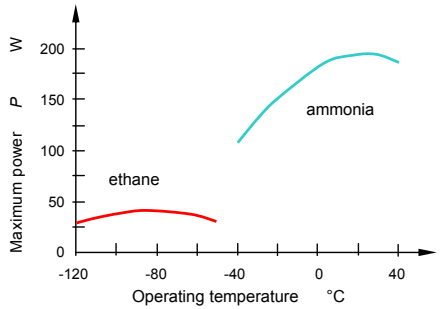


Figure 5.43. *Low temperature constant conductance heat pipe performance for Al/NH_3 and Al/C_2H_6 with slight adverse elevation, 12mm internal diameter, 60cm adiabatic section and 15cm evaporator/condenser.*

Intermediate temperature heat pipe technology

The intermediate temperature range extends from 450 to 750K. The alkali metals, such as caesium, potassium and sodium, are suitable working fluids at temperatures above this range. In the intermediate temperature range, the alkali metal vapour density is so low that the vapour sonic velocity limits the heat transfer, and the heat pipe vapour space becomes too large to be practical for alkali metals.

Water is commonly used at temperatures up to about 425K. Higher temperature water heat pipes can be used with titanium or Monel envelopes at temperatures up to 500K. Their effectiveness starts to fall off after 500K, due to the decrease in the surface tension of water. Potential intermediate temperature working fluids including Sulphur/Iodine mixtures, Iodine, Naphthalene, Phenol, Toluene, Mercury, and several halides. Sulphur has a temperature dependent polymerization property at 475K, which increases its liquid viscosity to approximately three orders of magnitude higher than the maximum level for effective heat pipe operation. The addition of 3-10% of iodine reduces the viscosity of sulphur to a level sufficient for effective heat pipe operation. A disadvantage of iodine is its low liquid thermal conductivity.

Another set of potential working fluids is the halide salts of titanium, aluminium, boron, phosphorus and silicon. Salts such as $TiCl_2F_2$ working fluids are polar, which increases the latent heat and the liquid transport factor.

Vapour Pressure and Merit Number are two parameters used to screen potential working fluids. Vapour pressures for some intermediate temperature working fluids are shown in Figures 5.44. Note that the vapour pressure for water is too high, and the vapour pressure for caesium is too low in this temperature range, so a vapour pressure between these two extremes is desirable. Most of the fluids discussed have a suitable vapour pressure.

The merit number (liquid transport factor) M_ℓ is a means of ranking heat pipe fluids, with a higher merit number more desirable:

$$M_\ell = \frac{\rho_\ell \sigma \lambda}{\nu} \qquad (5.83)$$

where: M_ℓ = merit number, W/m^2
ρ_ℓ = liquid density, kg/m^3
σ = surface tension, N/m
λ = latent heat, J/kg
υ = liquid viscosity, Pa

Figure 5.44 show the merit number as a function of temperature for a number of fluids. While water and caesium have good merit numbers, their vapour pressures eliminate them from contention.

Figure 5.45 compares the theoretical heat transfer capability (power) for heat pipes with five working fluids: water, iodine, $BiCl_3$, $SbBr_3$ and caesium. Water is the best fluid at the low temperature end and caesium at the upper end. Iodine and $SbBr_3$ offer good performance in the middle. However, Iodine has two potential problems: low liquid thermal conductivity and high corrosiveness.

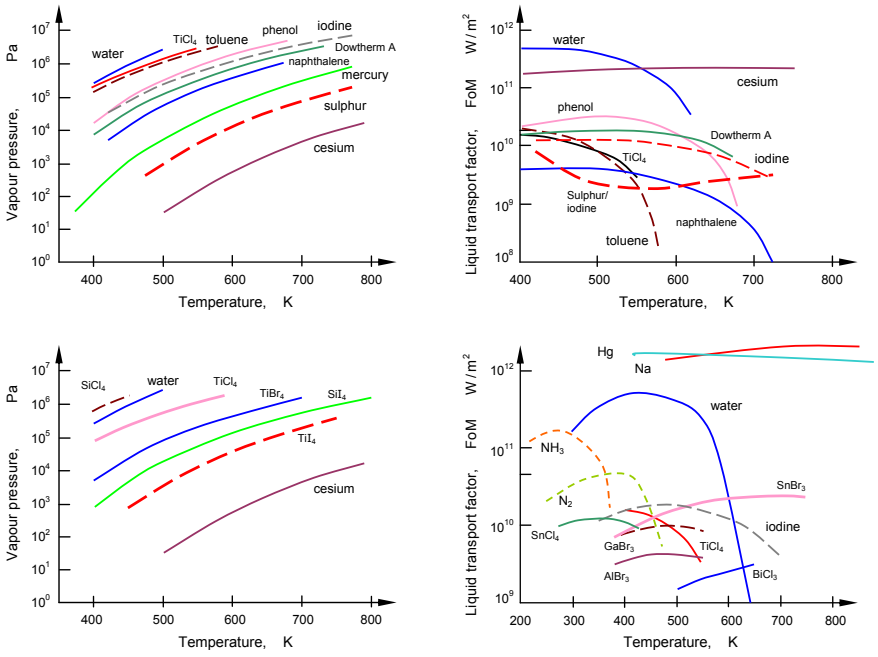


Figure 5.44. *Vapour pressure and figure of merit for intermediate temperature heat-pipe fluids, at a constant latent heat and normal boiling temperature.*

There are many candidate fluids in the intermediate temperature range. While some of the fluids have sufficient physical property data to allow their use, none of the fluids has adequate life test data at appropriate conditions.

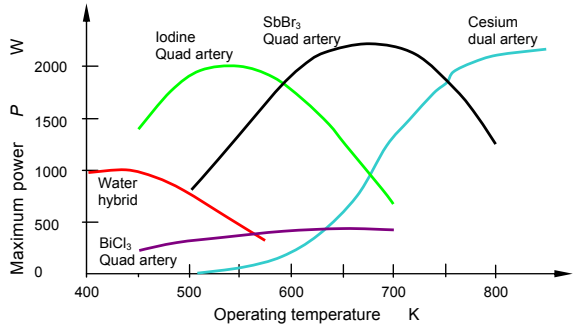


Figure 5.45. *Theoretical heat pipe powers for different fluids. Heat pipe dimensions: 2.54mm diameter, 220mm length.*

Evaporator cooling

Heat removal at the heat pipe remote evaporator can be affected with conventional aluminium finned heatsinks, with fan assistance if necessary. Any alternative cooling method in this chapter is applicable.

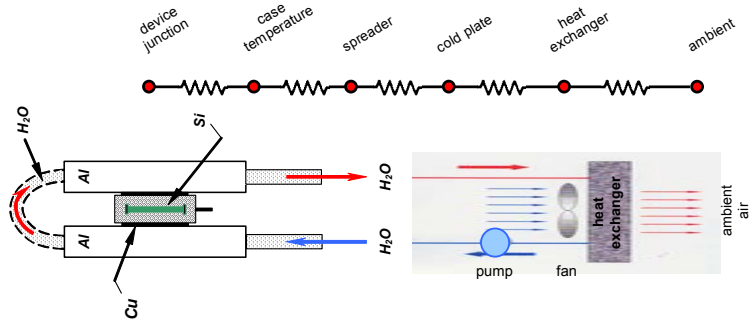


Figure 5.46. *Improved cooling with compact indirect cold-plate spreader water-cooling.*

5.16.2 Cold plates – indirect cooling

Liquid-cooled cold plates perform a function analogous to air-cooled heat sinks by providing an effective means to transfer heat from a component to a liquid coolant. Unlike heat pipes, they are active devices in that liquid is usually forced through them by the action of a mechanical pump. Vacuum-brazed fin-stock cold-plates and copper-based superalloy structures are used and a liquid-cooled microchannel heat sink can remove 790W/cm^2 with a temperature increase of 71°C for a 600ml/min flow rate with a pressure drop of 207kPa.

Liquid cooling can reduce effective thermal resistance to as low as 0.01K/W and may provide a much more compact clamp cold-plate spreader heat-sink arrangement, as shown in figure 5.46. Such equipment compact form factors require efficient yet cost-effective modes of removing excess heat.

Liquid cooling via cold-plate technology combines a high capacity for heat rejection with an ability to move heat remote from the power electronics to the ambient room air or facility water.

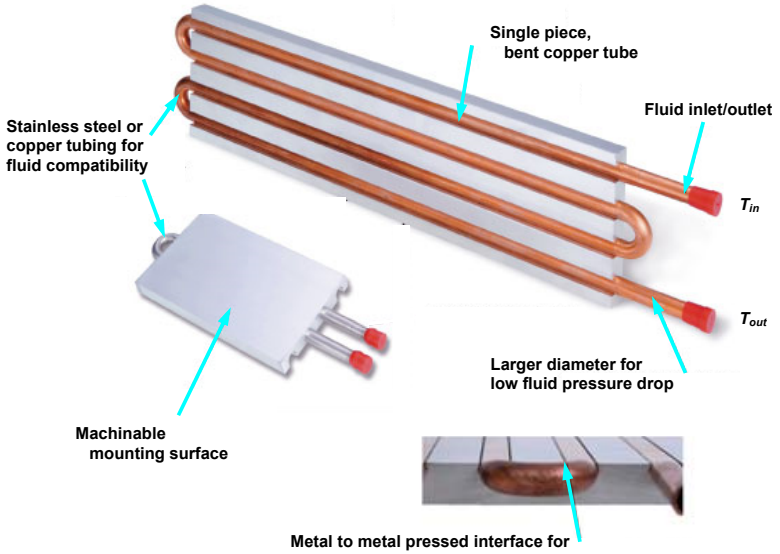


Figure 5.47. *Compact indirect cold-plate spreader water-cooler.*

Cold plates are used in closed systems where pumped coolants continuously cycle, conveying excess heat away from the devices to be cooled. This heat is then dumped into the ambient air via a radiator heat exchanger or recirculating chiller, or to facilities water via a liquid-to-liquid heat exchanger or liquid-cooled recirculating chiller, as shown in figure 5.46.
Depending on the specific cold-plate design, the components may be mounted on one or both sides of the plate, which is usually aluminium, as shown in figure 5.47.
In transfer the conducted (dissipated) heat P_D to the liquid coolant, a cold plate must satisfy

$$Q_{flow} = P_D$$
$$mc_p\left(T_{out} - T_{in}\right) = P_D = hA\left(T_s - T_m\right)$$
$$\text{where } T_m = \tfrac{1}{2}\left(T_{out} + T_{in}\right)$$

The heat transfer coefficient, from the heated surface, area A, at temperature T_s, is

$$h = \frac{\Delta T}{P_D} = \frac{Nu \times k}{D_H} \qquad (5.84)$$

where the Nusselt number is $Nu = 4.44$ for fully developed flow conditions.

$$\text{pressure drop}: \Delta P = \tfrac{1}{2}\frac{fL\rho V^2}{D_H} \quad (\text{Pa})$$
$$\text{pump power (W)}: P_w = \Delta P \times G \quad (\text{W})$$
$$\text{heat exchanger length}: L = -\frac{m_f c_p}{\pi h D_H}\ell \text{n}\frac{\Delta T_{out}}{\Delta T_{in}} \quad (\text{m}) \qquad (5.85)$$
$$\text{outlet temperature}: T_{out} = T_s - \left(T_s - T_{in}\right)e^{-\frac{1}{m_f c_p R_t}} \quad (\text{K})$$

where L = length of the cold plate
m_f = mass flow rate
f = friction factor
c_p = specific heat
D_H = diameter (hydraulic)
R_t = thermal resistance of one channel
G = volumetric flow rate, m^3/s
See section 5.18 on microchannels.

A number of cold plate technologies are available. Generally, the cold plate's cost will increase with improving performance. The technologies are listed in order of what is typically increasing cold plate efficiency and cost.

- ***Tubed cold plates***
  Tubed cold plates are the most common designs, giving dependable high performance at a comparatively low cost. A typical device is a flat metal plate, 15 to 30 mm long, with a series of channels on one or both sides. Into these channels is secured a length of serpentine metal tubing through which the liquid coolant flows. Fittings at the inlet and outlet of the coolant tubing connect to the user's coolant source.
  Usually aluminium, a tubed cold plate may either be cut to size (with channels for coolant tubing later machined into it), or extruded to size (with the channels formed at the same time the plate is extruded). Some manufacturers use a channel geometry that holds the coolant tubing in place, ensures good metal-to-metal contact, and optimizes thermal transfer. Others secure the tubing to the channels with a performance-limiting thermal epoxy.
  Tubing material is predominantly either 5 or 10 mm. out-side diameter copper or stainless steel. The larger diameter offers a lower pressure-drop. The smaller tubing increases fluid turbulence and can be formed into tighter radii for tighter tube packing and more-compact lower-profile cold plates. If the coolant is tap water, copper will suffice. If the coolant is deionised water or another corrosive fluid, stainless steel is preferable because it is non-reactive.
- ***Gun-drilled cold plates***
  A gun-drilled cold plate is usually fabricated by drilling a series of holes through the length of an aluminium plate, inserting copper or stainless-steel tubing, and then expanding the tubing to ensure a secure metal-to-metal contact with efficient thermal transfer properties. If aluminium corrosion is not a risk, it is possible to flow coolant with an ethylene glycol additive directly through the gun-drilled holes without inserting the tubing.
  For the return fluid path, the holes are drilled perpendicular to the main fluid path and then partially plugged to create a continuous coolant path. Before the development of two-sided tubed cold plates, gun drilling was the preferred approach for two-sided applications. Tubed cold plates are preferred because the inserted copper tubing eliminates the concern of aluminium corrosion. One additional benefit of gun-drilled cold plates is that they can have tighter tolerances than tubed cold plates, specifically for flatness requirements.

- ***Vacuum-brazed inner finned cold plates***
  Vacuum-brazed aluminium cold plates are reserved predominantly for high-performance custom designs that provide low thermal resistance, and superior leak-free reliability. They afford the greatest flexibility in specifying thermal resistance, thermal flow, pressure drop, fluid path, size, shape, material hardness, surface geometry, and dual-sided component mounting.
  To build these cold plates, high-performance corrugated aluminium-fin material is brazed into the liquid cavity below the mounting surface, which is also made of aluminium. The internal fins add heat transfer surface area and create turbulence in the coolant to minimize the fluid boundary layer and reduce thermal resistance.
  Depending on requirements, these cold plates can be surface machined to optimize heat-load and/or cold-plate contact. The size range of vacuum-brazed cold plates is from about 20 cm^2 to about 4000 cm^2.
- ***Extruded cold plates***
  Extruded cold plates represent the highest performance designs with the best thermal performance characteristics. A dimensionally compact structure (typically 5 cm wide x 4 mm. thick), this design features an aluminium or copper extrusion with many internal, discrete, parallel microchannels.
  The open ends of the extrusion are welded to connector tubes. The microchannel design provides a large internal surface area and a thin mounting surface that minimizes thermal resistance (see section 5.18).
  The channels create turbulence, which minimizes the fluid boundary layer and reduces thermal resistance. This approach also yields excellent thermal uniformity because coolant flows below the entire cold-plate surface. Depending on the direction the header manifolds are mounted, the extruded cold plates may have a 'U' or 'Z' fluid inlet/outlet configuration. Multiple extruded cold plates can be assembled into a single unit or pressed into an aluminium plate.

Cold-plate performance comparison

Cold-plate performance is normally expressed as thermal resistance, in °C/W. The lower the thermal resistance, the better the cold plate performance and the cooler the surface. Figure 5.48 compares normalized results with respect to area, allowing the various cold-plate technologies to be compared independent of individual part geometries.

Fluid pressure drop increases exponentially with fluid flow rate and viscosity.

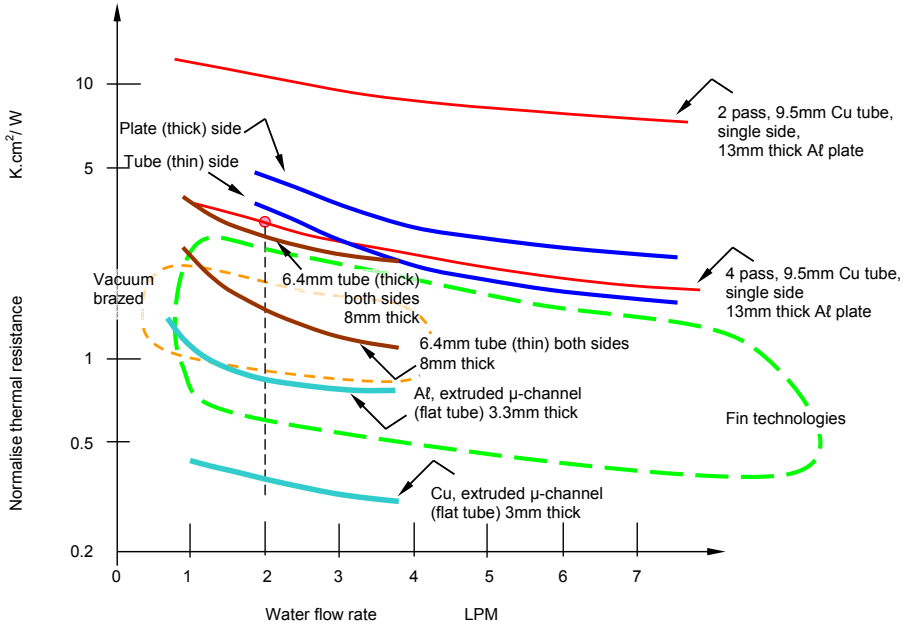


Fig. 5.48. *Performance comparison of various types of cold plates.*

Heat transfer fluids for cold plate liquid cooling

One of the most important factors when choosing a liquid cooling technology for an application is the compatibility of the heat transfer fluid with the wetted surfaces of the cooling components or system and the application. Heat transfer fluid compatibility is critical in ensuring long-term system reliability. Some other requirements for a heat transfer fluid may include high thermal conductivity and specific heat, low viscosity, low freezing point, high flash point, low corrosivity, low toxicity, and thermal stability. Based on these criteria, the most commonly used coolants for indirect liquid cooling applications are:

- Water
- Deionised Water
- Inhibited Glycol and Water Solutions
- Dielectric Fluids

By selecting a compatible pairing of heat transfer fluid and wetted materials, the risk of corrosion as well as the thermal performance are optimised. Copper is compatible with water and glycol/water solutions and aluminium is compatible with glycol/water solutions, dielectric fluids, and oils. When using deionised water or other corrosive fluids, however, stainless steel is used since it is more corrosion resistant than other metals. (See Tables 5.19 and 5.23.) Most cooling systems are compatible with water or glycol/water solutions but require special plumbing for compatibility with deionised water or a dielectric fluid such as polyalphaolefin, PAO.

Table 5.23: Cold plate metals and compatible fluids

Materials and Transfer Fluid Compatibility	Water	Glycols EGW	Deionised Water	Oil	Dielectric Fluids (Fluorinert)	Polyalphaolefin, (PAO)
Copper tubing	X	X				
Stainless steel tubing	X	X	X			
Aluminium flat tube or plate fin		X		X	X	X
Copper flat tubing, copper etched or copper brazed	X	X		X	X	X
Nickel brazed			X			

Selecting a cold plate

To determine the maximum allowable thermal resistance of the cold plate hence to specify a cold plate, it is necessary to know the cooling fluid flow rate, the fluid inlet temperature, the heat load of the devices attached to the cold plate, and the maximum desired cold plate surface temperature, T_{max}.
First, calculate the maximum temperature of the fluid when it leaves the cold plate, T_{out}. This is important because if T_{out} is greater than T_{max}, there is no solution to the problem.

$$T_{out} = T_{in} + \frac{P_d}{\dot{m} \times c_p} = T_{in} + \frac{P_d}{\rho_f v \times c_p} \qquad (5.86)$$

where T_{out} = temperature of fluid leaving cold plate, K
T_{in} = inlet temperature of fluid, K
P_d = heat load of power devices, W
ρ_f = density of the fluid, kg/m^3
v = cooling fluid flow rate, m^3/s
c_p = specific heat of the cooling fluid, J/kgK

Alternatively, heat capacity graphs describe the change in temperature, ΔT, that occurs along the fluid path. To find T_{out}, add ΔT to the inlet temperature, T_{in}.
Assuming T_{out} is less than T_{max}, the next step is to determine the required normalized thermal resistance R_θ for the cold plate using:

$$R_\theta = (T_{max} - T_{out})\frac{A}{P_d} \qquad (5.87)$$

where R_θ = thermal resistance, K/W
T_{max} = maximum desired cold plate surface temperature, K
A = area being cooled, m^2

Any cold plate technology that provides a normalized thermal resistance less than or equal to the calculated value will be a suitable solution.

The graph in figure 5.48 compares normalized thermal resistance expresses in °C.cm^2 / W of cold plate technologies. Normalization of the data allows thermal performance comparison across technologies without regard to physical size. Technologies with lower thermal resistance values offer the highest performance in the application.

Example 5.16: *Cold plate design*

A cold plate is used to cool a 5cm x 10cm copper base IGBT that generates 500W of heat. It is cooled with 20°C water at a 2 lpm flow rate. The surface of the cold plate must not exceed 55°C. Calculate the plate outlet header temperature, hence necessary plate thermal resistance.
Use c_p = 4,184 J/kgK and ρ_f = 998 kg/m^3

Solution

T_{in}: 20°C, T_{max}: 55°C, P: 500 W, Area: 5x10=50 cm^2

First calculate T_{out}. Using the heat capacity equation (5.86), the temperature change for 500W at a 2 lpm flow rate is 4°C. Therefore

$$T_{out} = T_{in} + \frac{P}{\rho v \times c_p}$$

$$= 20°C + \frac{P}{998\text{kg/m}^3 \times \frac{2}{60}\text{l/s} \times 10^{-3}\text{m}^3/\text{l} \times 4{,}184\text{J/kgK}} = 20°C + 3.6°C = 23.6°C$$

T_{out} is less than T_{max} so the required thermal resistance is determined by equation (5.87):

$$R_\theta = (T_{max} - T_{out})\frac{A}{P}$$

$$= (55°C - 23.6°C) \times \frac{50\text{cm}^2}{500\text{W}}$$

$$= 3.14°C.\text{cm}^2/\text{W} \text{ at } 2 \text{ lpm}$$

This point is plotted on the normalized thermal resistance graph in figure 5.48 where any technology below this point will meet the thermal requirement. But because the cooling fluid is water, aluminium tube or plate fin, or nickel-brazed technologies should not be employed.

♣

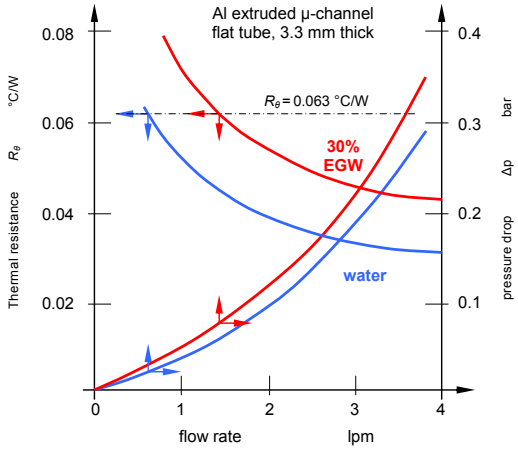


Figure 5.49. *Effects of flow rate and coolant variation on thermal resistance.*

In example 5.15 the cold plate thermal resistance requirement is 3.14 Kcm^2/W / 50cm^2 = 0.063K/W. This requirement can be met by an aluminium µ-channel cold plate with the characteristics shown in figure 5.49. This cold plate readily fulfils the thermal resistance requirements with de-ionised water. The aggressive nature of deionised water can be avoided by the using 30% ethylene glycol - water but at the expense of a decreased thermal conductivity, as seen in Table 5.19. From figure 5.49, the necessary flow rate is 1.5 lpm, with a cold plate input to output fluid pressure drop of 0.08 bar.

5.17 Direct liquid cooling

5.17.1 Immersion cooling – direct cooling

Direct liquid or immersion cooling is a well-established method for accommodating high heat flux. With natural convection two-phase flow, generally termed nucleate pool boiling, the critical heat flux using FC-72 is in the range of 5 to 20W/cm^2. However, much higher heat fluxes up to 100W/cm^2 can be accommodated with surface enhancement of the heat source. A device submerged in a pool of dielectric liquid and the heat dissipated in the device produces vapour bubbles that are driven by buoyancy forces into the upper region of the container, where the vapour condenses and drips back into the liquid pool. A disadvantage of this technique is the need for a liquid compatible with the device. Often, water cannot be used because of its chemical and electrical characteristics.
Direct liquid cooling may also be termed direct liquid immersion cooling, since there are no physical walls separating the devices and the surface of the substrate from the liquid coolant. This form of cooling offers the opportunity to remove heat directly from the package with no intervening thermal conduction resistance, other than that between the device heat sources and the package surfaces in contact with the liquid.
Direct liquid immersion cooling offers a high heat transfer coefficient which reduces the temperature rise of the package surface above the liquid coolant temperature. As shown in Figure 5.50, the relative magnitude of a heat transfer coefficient is affected by both the coolant and the mode of convective heat transfer (that is, natural convection, forced convection, or boiling). Water is the most effective coolant and the boiling mode offers the highest heat transfer coefficient. Direct liquid immersion cooling also offers greater uniformity of package temperatures than is provided by air-cooling.

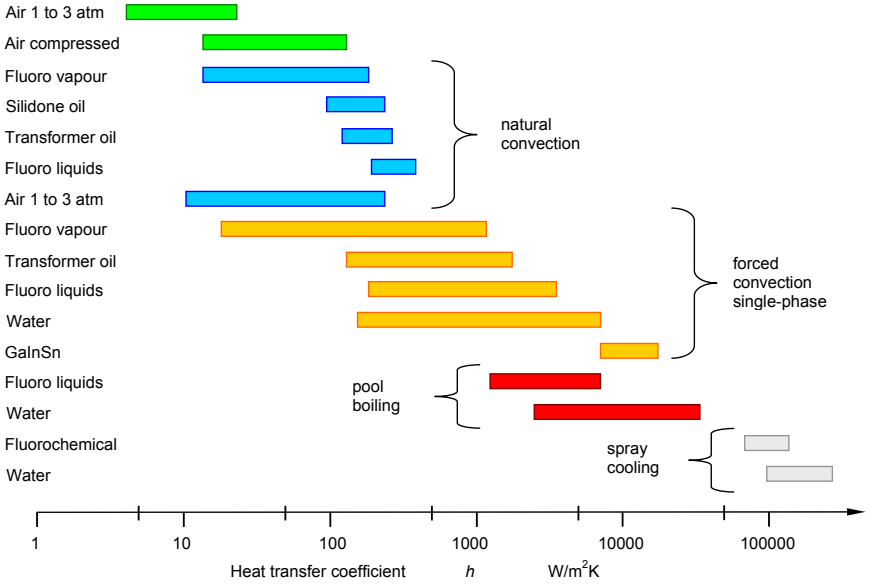


Figure 5.50. *Relative magnitude of heat transfer coefficients for various coolants and modes of convection. (See figure 5.19)*

Coolant considerations
Immersion cooling involves more than just the selection of a direct immersion liquid based on heat transfer characteristics alone. Chemical compatibility of the coolant with the mounting and packaging materials exposed to the liquid are the primary consideration.
There may be several coolants which can provide adequate cooling, but a limited few will be chemically compatible. Water is an example of a liquid which has desirable heat transfer characteristics, high thermal conductivity for example, but is generally unsuitable for direct immersion cooling because of other chemical characteristics, such dielectric constant. The fluorocarbon liquids listed in Table 5.24 are generally considered the most suitable liquids for direct immersion cooling, despite their poorer thermo-physical properties.

As shown in Table 5.24, the thermal conductivity, specific heat, and heat of vaporization of fluorocarbon coolants are lower than water. These coolants are clear, colourless per-fluorinated liquids with a relatively high density and low viscosity. They also exhibit a high dielectric strength and a high volume resistivity. The liquid therefore serves to both cool and insulate the components. The boiling points for available 'fluorinert' liquids range from 30 to 253 °C.
All of the high power and high voltage power electronics can be immersed in the fluorinert. As a component heats up, the fluorinert in contact with it vaporizes and it is this liquid to vapour phase transition which effectively removes the excess heat from the components. The fluorinert vapour is cooled by a heat exchanger located in the area above the fluid. This cooling technique allows high power electronics to operate continuously and reliably in a small volume.
Why use a dielectric liquid, specifically FC-72 instead of water?

- Same weight of water will extract more than three times the heat for the same temperature rise (higher specific heat).
- Water is eleven times more conductive.
- Same weight of water upon evaporating will extract 27 times more heat (higher heat of vaporization).
- However, FC-72 has a dielectric constant that is 2% that of water, that is, much less unwanted current will flow.

Table 5.24: Comparison of thermo-physical properties of some fluorocarbon coolants and water

Property		units	FC-87	FC-72	FC-77	H_2O
Boiling Point		@ 1 Atm, °C	30	56	97	100
Density ρ_ℓ	$\times 10^3$	kg/m^3	1.633	1.680	1.780	0.997
Specific Heat	$\times 10^3$	W.s/kg.K	1.088	1.088	1.172	4.179
Thermal Conductivity		W/m.K	0.0551	0.0545	0.057	0.613
Dynamic Viscosity υ	$\times 10^{-4}$	kg/m.s	4.20	4.50	4.50	8.55
Heat of Vaporization $_L$	$\times 10^4$	W.s/kg	8.79	8.79	8.37	243.8
Surface Tension	$\times 10^{-3}$	N/m	8.90	8.50	8.00	58.9
Thermal Coefficient of Expansion	$\times 10^{-3}$	K^{-1}	1.60	1.60	1.40	0.20
Dielectric Constant		=1 for a vacuum	1.71	1.72	1.75	78.0

These liquids should not be confused with the 'Freon' coolants which are chlorofluorocarbons (CFCs). Although some of the 'Freons' exhibit similar cooling characteristics, concern over their environmental effect preclude their use.

Modes of heat transfer
The convective heat transfer processes upon which liquid immersion cooling depends may be classified as natural convection, forced convection, or boiling modes. The relative magnitude of heat fluxes which can be accommodated by each mode is shown in Figure 5.51, as a function of "wall superheat" or surface-to-liquid temperature difference for a typical fluorocarbon coolant.

Natural convection: As in the case of air-cooling, natural convection is a heat transfer process in which mixing and fluid motion is induced by coolant density differences caused by the heat transferred to the coolant. As depicted in Figure 5.51, this mode of heat transfer offers the lowest heat flux or cooling capability for a given 'wall superheat'. Nonetheless, the heat transfer rates attainable with liquid natural convection can exceed those attainable with forced convection of air. Natural convection would typically be employed within a closed container to transfer heat from die or modules to liquid, and then from the liquid to the walls of the container. Heat could then be transferred from the walls to outside air by natural or forced convection.

Forced convection: Higher heat transfer rates may be attained by utilizing a pump to provide forced circulation of the liquid coolant over the die or module surfaces. This process is termed forced convection; and as with air-cooling, the allowable heat flux for a given surface-to-liquid temperature difference can be increased by increasing the velocity of the liquid over the heated surface. Depending upon the surface geometry and the nature of the flow (that is, laminar or turbulent), the heat transfer coefficient will be proportional to the velocity to a power between 0.5 and 0.8. The penalty for increasing cooling performance in this way, is a higher-pressure drop. This can mean a larger pump and higher system operating pressures. Although forced convection requires the use of a pump and the associated piping, it allows removal of heat from high power modules in a confined space; and then transports the heat via the liquid coolant to a remote heat exchanger to reject the heat to air or water.

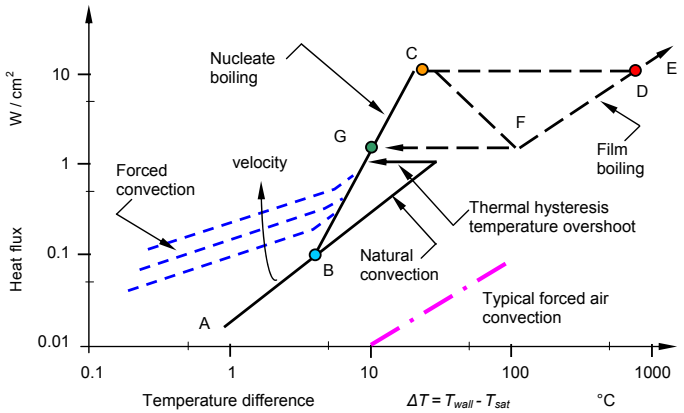


Figure 5.51: *Typical heat transfer regimes for immersion cooling with a fluorocarbon liquid.*

Boiling: Boiling is a complex convective heat transfer process depending upon liquid-to-vapour phase change with the formation of vapour bubbles at the heated surface. It is commonly characterized as either *pool boiling* (occurring in a stagnant liquid) or *flow boiling*. The pool boiling heat transfer rate Q usually follows a relationship of the form

$$Q = C'_{sf} \times A \times (T_{wall} - T_{sat})^n \qquad (5.88)$$

where C'_{sf} is a constant depending on each fluid-surface combination,
A is the heat transfer surface area,
T_{wall} is the temperature of the heated surface,
T_{sat} is the saturation temperature (that is, boiling point) of the liquid, and
n is an exponent, typically about 3.

The boiling curve for a particular surface and fluid of interest (for example, silicon and FC-72) is usually obtained experimentally. An example of a boiling curve depicting the cooling path from natural convection to film boiling is shown in Figure 5.51. If die power is gradually increased in small steps, cooling occurs first by natural convection (A - B). Eventually a power level is reached at which sufficient superheat is available to initiate the growth of vapour bubbles on the surface and boiling starts (B). As power is increased, more nucleation sites become active and the frequency of bubble departure increases. The region between B and C is termed the nucleate boiling regime. Vigorous agitation of the hot boundary along the heated surface, and gross fluid circulation caused by the motion of the vapour bubbles, provide the ability to accommodate substantial increases in heat flux with minimal increases in surface temperature. As power is increased to point C, the critical heat flux condition is reached. So many bubbles are generated at this point that they begin to form a vapour blanket inhibiting fresh liquid from reaching the surface. Further increases in power will result in a transition to film boiling (D - E). In this regime, heat transfer from the surface to the liquid is dependent on thermal conduction through the vapour and it is poor. In most power electronic cooling applications, transition to film boiling will result in failure due to high temperatures. To take advantage of boiling to cool electronic devices, it is desirable to operate in the nucleate boiling regime (B - C).
A problem often associated with pool boiling of fluorocarbon liquids is that of temperature overshoot. This behaviour is characterized by a delay in the inception of nucleate boiling (that is, beyond point B), such that the heated surface continues to be cooled by natural convection; with increased surface temperatures unless a sufficient superheat is reached for boiling to occur. This behaviour is a result of the good wetting characteristics of the fluorocarbon liquids and the smooth nature of silicon die and their metallisation. There is minimal temperature overshoot associated with flow boiling cooling applications.
The typical critical heat fluxes encountered in saturated (that is, liquid temperature saturation temperature) pool boiling of fluorocarbon liquids ranges from about 10 to 15 W/cm^2, depending upon the nature of the surface (that is, material, finish, geometry). The allowable critical heat flux may be extended by subcooling the liquid below its saturation temperature. For example, the critical heat flux can be increased to as much as 25 W/cm^2 by reducing the liquid temperature to -25°C.
Higher critical heat fluxes may be achieved using flow boiling. For example, heat fluxes from 25 to over 30 W/cm^2 have been reported for liquid velocities of 0.5 to 2.5 m/s over the heated surface. Heat fluxes in excess of 100 W/cm^2 have been obtained with a FC-72 liquid jet impinging upon a 6.5 mm x 6.5 mm chip at a flow rate of 2.2 cm^3/s.

Other considerations
Since fluorocarbon liquids are expensive, they should only be used in closed systems. Whether the application is in a self-contained module or a forced flow scheme, the seal materials must be compatible with the liquid. If boiling is to take place, then the design must incorporate a means to condense the resulting vapours. A finned surface may be designed for this purpose, or a remote finned condenser surface cooled by air or water might be used. In flow systems, care must be taken in selecting a pump. The relatively high vapour pressure of the low boiling point fluorocarbons generally require that a higher suction head be provided to prevent cavitation in the pump. In forced circulating liquid systems, it may be desirable to add a particulate and a chemical filter to ensure the long-term purity of the coolant.

5.17.2 Liquid jet impingement – direct cooling

Jet impingement cooling, as shown in figure 5.52, is similar to spray cooling but is performed with a lower nozzle pressure drop and a higher fluid flow rate. This lower pressure requirement, in conjunction with a high flow volume, reduces nozzle clogging and reduces noise levels. A cooling of 90W/cm^2 with a 100°C temperature rise using a flow rate of only 8ml/min is possible. A closed loop impingement jet gives cooling of 180W/cm^2 using water and a flow of 0.3 l/min at 300kPa. The micropump used 7W to drive the fluid flow.

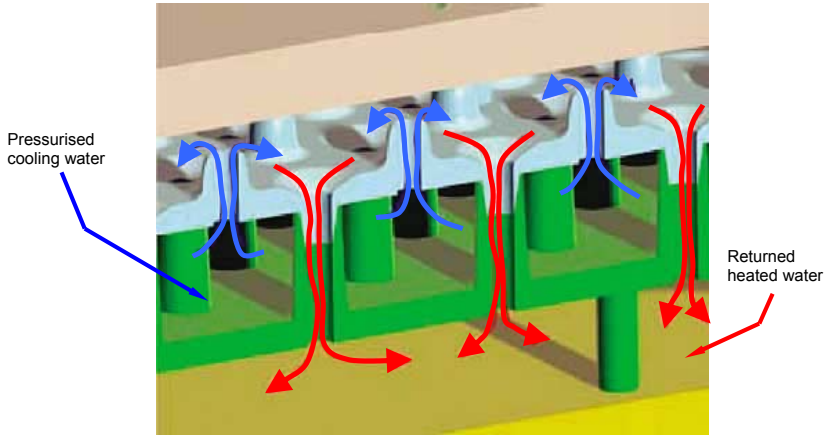


Figure 5.52. *Available multiple jet impingement liquid cooling.*

5.17.3 Spray Cooling – direct cooling

Spray cooling breaks up the liquid into fine droplets that impinge individually on the heated wall surface. Cooling of the surface is achieved through a combination of thermal conduction through the liquid in contact with the surface and evaporation at the liquid-vapour interface (latent heat of vaporization). The droplet impingement both enhances the spatial uniformity of heat removal and delays the liquid separation on the wall during vigorous boiling. The hot vapour is recovered by removing the waste heat to the ambient in a heat exchanger where the vapour condenses back to a liquid. The fluid is continually recycled within a closed system. Using the vapour for heat transport eliminates all resistance between the heat source and the ultimate heat destination.
Spray cooling technology enables all surfaces exposed to the liquid-vapour environment to remain close to the saturation temperature of the fluid. The result is an isothermal environment around the power devices which effectively reduces hot spots and thermal cycling – the primary cause of power device failure. Spraying reduces the flow rate requirements but requires a high nozzle pressure drop, than jets. Spray pattern types include an array of swirling turbulent hollow-cone sprays, with a wide range of droplet sizes. Drawbacks include likelihood of nozzle clogging, repeatability of impact patterns, high sensitivity to nozzle to surface distance, and the need for filters and pumps.
Spray cooling and jet impingement (as shown in Figure 5.53) are often considered competing options for electronic cooling. Spray evaporative cooling with a Fluorinert coolant can maintain junction temperatures between 70 and 85°C for heat fluxes from 15 to 55 W/cm^2. Spray cooling improves thermal management, increases system-packaging density, and reduces weight. Die-level spray cooling allows a maximum heat flux of over 160 W/cm^2.

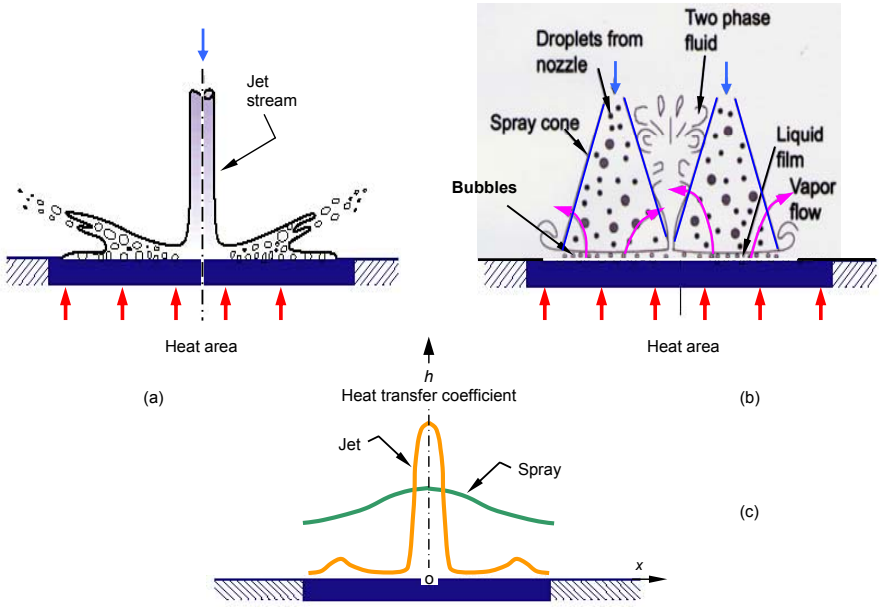


Figure 5.53. *Illustration of (a) spray cooling; (b) jet impingement cooling; and (c) resultant heat transfer coefficient, h.*

5.18 Microchannels and minichannels

The concept of microchannels applied to the thermal management of high heat-flux power electronics is simple. Because heat-transfer coefficients generally increase with decreasing size, the passage size (microchannels) should be made as small as possible. This results in a dense package with higher heat transfer and a larger surface area-to-volume ratio than a conventional cooling device.
The term 'micro' is applied to devices having hydraulic diameters of ten to several hundred micro-metres, while 'mini' refers to diameters of the order of one to a few millimetres. A low flow rate within micro-channels produces laminar flow resulting in a heat transfer coefficient inversely proportional to the hydraulic diameter. In other words, the smaller the channel, the higher the heat transfer coefficient. Unfortunately, the pressure drop increases with the inverse squared of the channel width, whilst maintaining the mass flow constant. Stacking of microchannel layers is used to decrease the pressure drop, but flow non-uniformity across the channel remains a problem, producing non-uniform cooling.
A 50μm wide, 300μm deep, 1cm long microchannel passage, experiences a 30-psi drop with a 0.66 lpm water flow, can dissipate 790 W/cm^2 with a 71°C temperature rise. With 5 to100um channel sizes, the heat transfer coefficient can reach 80kW/m^{2}K.

Integrated microchannel heatsink
Figure 5.54 displays the fractions of the thermal resistivity attributed to different individual layers for a typical power module. In the case of microchannel cooling, the largest contributor to the thermal resistivity is the thermal grease layer between the base-plate and the heatsink.
Because of the increased importance of the conductive resistance of the stack, a better solution would be to eliminate some of the layers from the structure. An integrated heatsink with a series of microchannels fabricated directly into the bottom copper layer of the active metal braze (AMB) substrate is shown schematically in figure 5.55, and has advantages in reducing both the convective and conductive resistances of the module.

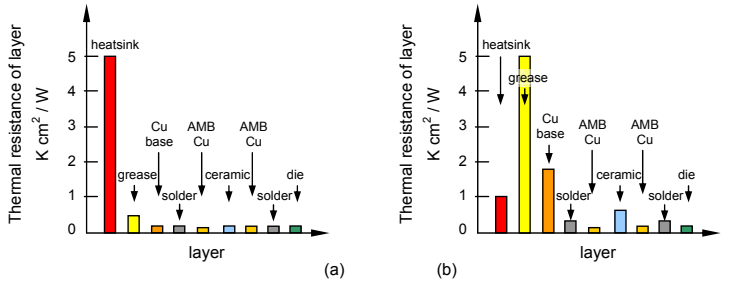


Figure 5.54. *Thermal resistivities of a typical power module with: (a) a conventional heatsink and (b) a microchannel heatsink, indicate that layers other than the heatsink can dominate the performance of microchannel designs.*

By using microchannels, the convective thermal resistivity is reduced dramatically. In addition, this stack removes the base-plate solder, the copper base-plate and, most importantly, the thermal grease from the conductive path. As seen in figure 5.54, this results in the elimination of the two largest resistances in the structure. A thermal grease or epoxy layer for attachment, is a thin 75μm layer a with relatively high thermal conductivity of 9 W/m/K. The overall result is a reduction of the total stack resistivity by a factor of two when compared to a microchannel heatsink.

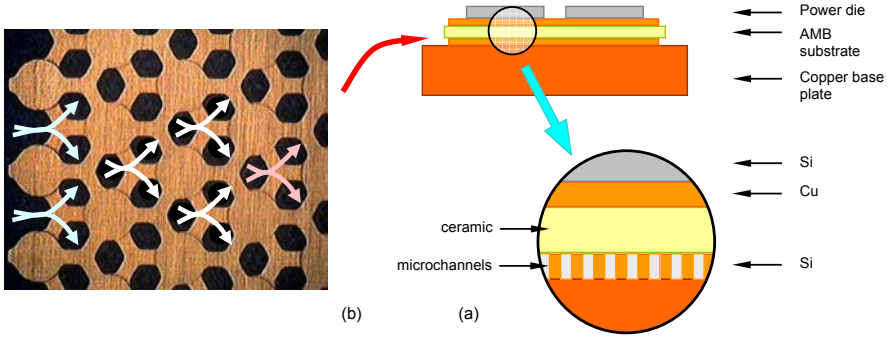


Figure 5.55. *Integrated microchannel heatsink design, where microchannels are (a) fabricated directly into the bottom copper layer of the AMB substrate or (b) created between at least two thin (0.3mm) copper etched mesh layers, bonded by a melting process.*

Analytical optimization
First-order heat transfer is governed by the thermal resistivity R_θ, which is defined as the temperature rise divided by the heat flux. For convective heat transfer in channels having hydraulic diameter D_H, the thermal resistivity is calculated as:

$$R_\theta = \frac{\Delta T}{P_d} = \frac{D_H}{k \times Nu} \qquad (5.89)$$

where k is the working fluid thermal conductivity, and
Nu is the Nusselt number for the appropriate flow condition ($Nu = A \times Re^m Pr^n$).

$$Nu = \frac{h \times L}{k}$$

where h is the convective heat transfer coefficient, W/m^{2}K and
L is the characteristic passage length of the microchannel.
Pr is the Prandtl number
For example, for laminar, fully developed flow in a circular passage with constant heat flux, Nu_{DH} = 4.36. [Nu_{DH} = 3.66 for convection with a constant uniform surface temperature.]
In addition, the pressure loss ΔP is calculated using the friction factor f as:

$$\Delta P = \tfrac{1}{2}\rho v^2 f \frac{L}{D_H} \qquad (5.90)$$

with

$$f = \frac{64}{Re}$$

where ρ_ℓ is the fluid density and
f is the friction factor (loss coefficient), in laminar, fully developed flow in a circular passage

Re is the Reynolds number

$$Re = \frac{\rho_\ell \times v \times D_H}{\upsilon} \qquad (5.91)$$

where v is the fluid velocity (volumetric flow rate) and
υ is the fluid kinematic viscosity.

These basic expressions for R_θ and ΔP are used to select the optimal channel sizing. The passage dimensions are chosen to minimize the thermal resistivity R_θ subject to pumping constraints on the maximum pressure loss and flow rate. Typical performance characteristic are as shown in figure 5.49.

The primary variables to be optimized are the channel width and pitch. These values vary beneath a heat source of 2×2 cm size between 10 to 200 cooling passages and a ratio of wall thickness to channel width of 0.1 to 2. These dimensions result in a range of flow conditions, including laminar, turbulent, developing and fully developed regimes. In addition, the channel height in the AMB substrate is varied from 0.05mm to 0.3mm, which is the maximum depth allowed due to the thickness of the bottom copper layer. The typical coolant is water at room temperature. The pump constraints are specified as 4-litre per minute maximum flow rate and 25-psi maximum pressure loss, which are representative values for power electronics cooling applications.
The channel width is at least 100µm, given the difficulty of manufacturing narrower passages in copper. For the integrated microchannel heatsink, the preferred channel shape has a width of 100µm, a depth of 300µm and a wall thickness between channels of 100µm. The narrow channel width and small pitch results in a high surface area-to-volume ratio, while the tall channel height tempers the pressure loss through the passage. The calculated thermal resistivity for this design is 0.042K.cm^2/W.
The performance is effective and superior to existing heatsinks since no thermal grease layer is needed. The overall thermal resistivity of a power module equipped with this heatsink is less than 0.15K.cm^2/W, resulting in less than 75°C junction-to-coolant temperature rise for a heat flux of 500W/cm^2. This thermal performance is better than any existing heatsink using a comparable material stack.

Microchannel cooling is more effective for areas smaller than 7 x 7 cm. Integrated single and two-phase micro heat sinks are able to cool about 450 W/cm^2 using both single and two-phase heat transfer. For two-phase flow, the pumping power is about ten times lower and the required flow rate is considerably lower. By using off-set strip fins and a split-flow arrangement, cooling of over 300 W/cm^2 at 24 kPa is possible with a flow of 1.5 l/min. A silicon microchannel cooler can be used for high-power chips. A separate microchannel cold plate is bonded to the back of the chip. This requires a low interface thermal resistance. If the microcooler is based on silicon, a rigid bonding means that silver-filled epoxies or solder should be used, giving power densities in excess of 400W/cm^2, for a flow of 1.2 l/min at 30kPa.
It may be possible to push microchannel heat transfer even further by utilizing boiling. In addition to offering higher heat transfer coefficients, boiling convection in microchannels requires less pumping power than single-phase liquid convection to achieve a given heat sink thermal resistance. For the same heat flux, the pressure drops by a factor of 20. A 1000W/cm^2 cooling system based on boiling heat transfer in microchannel heat sinks using a flow rate of 500ml/min is possible. Local heat transfer coefficients may change appreciably over time leading to local temperature changes of 10-15°C. Also backflow of already heated flow due to expansion of bubbles is observed.
If fluid impinges on the surface to be cooled, performance can reach 1000W/cm^{2}K, 14-21kPa and 0.05K/W/cm^2. Pumped liquid (both single and two-phase) cooling technologies in addition to loop heat pipes for space applications, in a single-phase solution, incorporates an oscillating flow heat transfer mechanism, capable of cooling over 1300W/cm^2.
A different way of making microchannels is to use metal foams or metal made porous, which accommodated a heat flux of 500W/cm^2 for a 50K difference at a pressure drop of 115kPa, using water.

5.19 Electrohydrodynamic and electrowetting cooling

As an alternative to a continuous flow set into motion either by temperature differences or by mechanical means, liquid can be formed and moved in droplets of nano-to-millilitre size by means of electric fields. Electrowetting on a dielectric film, in which the surface property of a dielectric film can be modified between hydrophobic and hydrophilic states using an electric field, can be used to provide the basis for a direct micropumping system. Electrowetting involves control of the surface tension of a liquid and can cause a droplet of liquid to bead or spread out on the surface depending upon its surface state.

With 0.4ml/min it is theoretically possible to cool 90 W/cm^2. The application of electrowetting to liquid metals has the advantage, besides a better heat transfer capability, of necessity a much lower voltage, 2V instead of 50V, to produce the electric field.

5.20 Liquid metal cooling

Table 5.25 shows properties for various liquid metals, where the viscosity is given at room temperature and water is included for comparison. Advantages of liquid metals are their much lower thermal expansion coefficient compared to water and the fact that freezing introduces fewer problems. Ideally, the metal should be non-poisonous, non-caustic, low viscosity, high thermal conductivity, and high heat capacity.

Pure Bismuth, Bi, melts at 271°C but some of it alloys have considerable lower melting points. Woods metal is probably the most well known, melting at 70°C. By alloying with metals such as lead, tin, cadmium and indium, it is possible to get a lower limit of 47°C. Such alloys have the disadvantage of high melting points.
Pure gallium, Ga, melts at 29.7°C and has a latent heat of 80.1 J/g, but several of its alloys have much lower melting points. Although non-toxic and relatively cheap, the main drawback is its aggressiveness towards most metals. All gallium alloys must therefore be enclosed within ceramic walls, which is difficult to realise. Because its surface tension is much higher than water, liquid Gallium is immune to the presence of small cracks or channels in imperfect seals that would cause leaks if water were the cooling fluid.

Table 5.25: Approximate thermal properties for liquid metals. Water is included for comparison

Metal	Melting point	Boiling point	Density	Specific heat	Conductivity λ	Viscosity υ	Kinematic viscosity	Prandtl number
	°C	°C	kg/m^3	J/kgK	W/mK	Ns/m^2 ×10^{-3}	m/s ×10^{-8}	$c_p \mu / k$
bi-alloys	47--271		9,800	142	8.4			
Ga	29.7	2205	5,900	334	28	2.04	32	0.0261
Hg	-38.8	356.8	13,546	140	7.8	0.15	11.4	0.0278
FC-72		56	1680	1088	0.0545	0.45		
H_2O	0	100	998	4184	0.613	0.86	85.5	6.62
NaK alloy (22/78%)	-11.1	783.8	892	1058	25.3	0.94	53.8	0.0213

The heat transfer coefficient h is the rate that thermal energy is removed from a surface per unit surface area per temperature differential, and is related to Nusselt number Nu, by

$$h = \frac{Nu \times \lambda}{k_D}$$

where k_D is the characteristics dimension of the geometry and
λ is thermal conductivity, shown in table 5.24 and 5,25.
In a circular tube characterised by convection with a uniform surface temperature and laminar fully developed conditions, Nusselt number is constant, Nu = 4.36, and 3.66 for a uniform heat flux condition.
Apart from heat pipes based on liquid metals, mainly for the high-temperature range, Ga-Sn-In eutectics that remain liquid down to minus 19°C are possible. The thermal heat transfer coefficient h is ten times larger than for water. High-performance liquid metal cooling loops use magnetofluiddynamic MFD pumps, with over 200 W/cm^2 cooling capacity, using a flow of 0.3 l/min at 15 kPa.
Mercury, Hg, has attractive thermal properties and has been used as a working fluid for power generating purposes. It has environmental drawbacks.
The best liquid metal candidate is an eutectic solution of sodium and potassium, NaK. The melting point is as low as -12°C. Its density and viscosity are similar to water but has a lower specific heat and a much higher thermal conductivity. It can be used with nickel, chrome and steel but is aggressive to cadmium, antimony, bismuth, copper, lead, silicon, tin, and magnesium. It also reacts violently with air and water. This alloy is associated with material and handling problems. Liquid sodium has nevertheless been used as a coolant for nuclear reactors, which shows that these drawbacks can be managed.

Any form of liquid cooling is associated with leakage hence reliability problems. Consideration of emersion cooling, etc., should be restricted to applications requiring precise temperature control and heat dissipation rates that are too high for effective removal by conduction and air cooling.

5.21 Solid state cooling

5.21.1 Thermoelectric coolers

A thermoelectric module is a highly reliable, small, light, solid-state, active device that can operate as a heat pump or as an electrical power generator. When used to generate electricity, the module is called a thermoelectric generator (TEG). When used as a heat pump, a refrigerator, the module utilizes the Peltier effect to move heat and is called a thermoelectric cooler (TEC).
When current passes through the junction of two different types of conductors, a temperature change results at the junction. However, the practical application of this concept requires semiconductors that are good conductors of electricity but poor conductors of heat. Anisotropic orientated polycrystalline bismuth telluride is mainly used as the semiconductor material, heavily doped to create either an excess (n-type) or a deficiency (p-type) of electrons, as shown in figure 5.56a.
If the current is reversed, the heat is moved in the opposite direction. In other words, what was the hot face will become the cold face and vice-versa.

A thermoelectric device consists of a number of p and n type pairs (couples) connected electrically in series and sandwiched between two ceramic plates, as shown in figure 5.56. When connected to a DC power source, dc current causes heat to move from one side of the TEC to the other. This creates a hot side and a cold side on the TEC. The device to be cooled is mounted on the cold side of the TEC and the hot side is thermally connected to another TEC or a heatsink which dissipates the heat into the environment. A heat exchanger with forced air or liquid may be required. A thermoelectric cooler does not absorb heat, it only transfers or moves it.

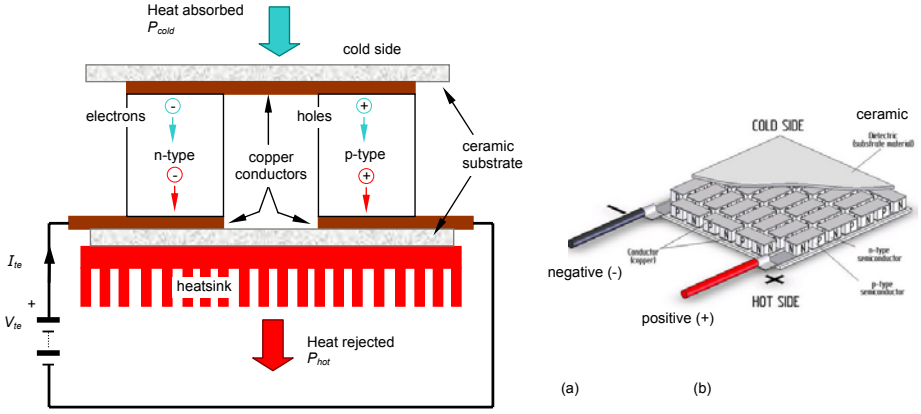


Figure 5.56. *The thermoelectric cooler: (a) principle and (b) module.*

Design involves the initial specification of three parameters, the hot and cold side temperatures, T_{hot} and T_{cold}, (or T_h and T_c) hence the temperature gradient or difference $\Delta T = T_h - T_c$, ($\Delta T > 0$) and the amount of heat, in Watts, to be absorbed at the cold surface of the TEC, P_{cold}.
The cold surface temperature is the desired temperature of the object to be cooled, directly in contact with the TEC.

The hot surface temperature is defined by two major parameters:

- The temperature of the ambient environment to which the heat is being rejected.
- The efficiency of the heat exchanger that is between the hot surface of the TEC and the ambient.

The third parameter required is the amount of heat, the thermal load, to be removed or absorbed by the cold surface of the TEC. The thermal load includes the active I^2R type losses of the device to be cooled, as well as parasitic loads such as conduction through any mechanical object in contact with both the cold surface and any warmer environment, like conduction through mounting bolts and plates (and the radiation from the plates). Figure 5.57 show the thermal resistance components and system model.

The *coefficient of performance*, CoP, is useful in selecting a module (the larger CoP the better), and is defined by

$$CoP = \frac{\text{heat absorbed at the cold junction}}{\text{electrical dc input power}} = \frac{P_{cold}}{V_{te} \times I_{te}} = \frac{P_{cold}}{P_{tec}} \qquad (5.92)$$

where I_{te} is the current drawn by the TE module
V_{te} is the voltage applied to the TE module

A maximum CoP represents the minimum input dc power P_{tec}, therefore minimum total heat to be rejected by the heat exchanger on the hot side P_{hot}, that is $P_{hot} = P_{cold} + P_e$.
Performance characteristic charts, as in figure 5.58, are usually provided. These allowing the terminal dc voltage and dc current requirements to be determined from the temperature difference ΔT and heat to be absorbed on the cold side, P_{cold}. The maximum ΔT is about 67°C for a single TEC, higher than this requires cascading (stacking) of TECs. The negative quadratic shape in the lower plot, represents the optimal operating curve. Further TE effect and module technical details can be found in Chapter 22.10.

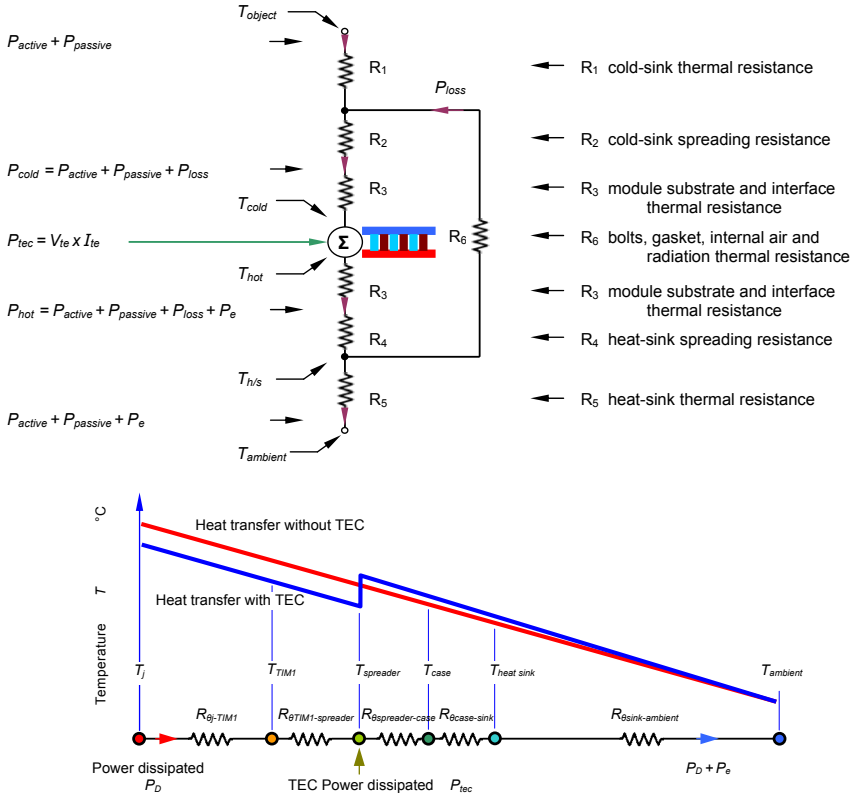


Figure 5.57. *The thermal resistance TEC cooling model.*

Example 5.17: ***Thermoelectric cooler design***

A semiconductor junction dissipating 100W has a steady-state thermal resistance to its case mounting of $R_{\theta j} = 0.15$ K/W. If the TEC cold-side temperature is not to exceed 5°C in a 25°C external ambient, determine the heatsink requirement to dissipate the transferred TEC heat, if the TEC has the characteristics shown in figure 5.58, where $T_h = 35$°C. Assuming $P_{loss} = 0$W, determine:

i. the semiconductor junction temperature;
ii. the coefficient of performance CoP for the TE module; and
iii. heatsink requirements.

Solution

i. The junction temperature is T_{cold} + Power dissipated × $R_{\theta\,j\text{-}c}$ = 5°C + 100W×0.15K/W = 20°C.

ii. The temperature differential across the TE module is

$$\Delta T = T_{hot} - T_{cold} = 35°C - 5°C = 30°C$$

Only one stage (no TEC cascading is required) should be necessary since the maximum $\Delta T < 62$°C.
Since thermal leakage losses are assumed zero, the cold-side power losses to be transferred are 100W, which is shown as the horizontal plotted line in the performance graph in figure 5.58.
The vertical intersection of the 100W horizontal line and the ΔT = 30°C curve gives the necessary TE module input current 18.2A, and associated terminal voltage, 9.6V, from the upper curve intersection.
The TE module electrical input power P_e is therefore 9.6V×18.A = 74.7W
From equation (5.92), the TE module coefficient of performance is

$$CoP = \frac{P_{cold}}{V_{te} \times I_{te}} = \frac{P_{cold}}{P_{tec}} = \frac{100}{74.7} = 1.39$$

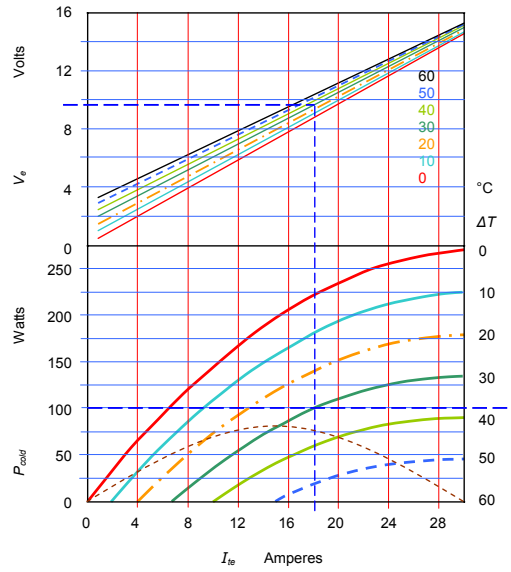


Figure 5.58. *Thermoelectric module characteristics, P_{max} = 270W, with 127 couples for a 35°C hot-side temperature T_{hot}, ΔT = 62°C, and I_{max} = 30A.*

iii. The heatsink thermal resistance requirement is

$$R_{\theta hs} = \frac{T_{hot} - T_{amb}}{\text{Power}} = \frac{35°C - 25°C}{100W + 18.2A \times 9.6V} = \frac{10°C}{100W + 18.2A \times 9.6V} = 0.036°C/W$$

♣

Example 5.18: ***Thermoelectrically enhanced heat sink***

A switching semiconductor device dissipates 120W. Figure 5.59 shows the thermal resistor equivalent network and associated resistances for cooling system designs based on:

T_{amb} = 40°C
$R_{\theta hs\text{-}a}$ = 0.18 K/W
$R_{\theta s\text{-}hs}$ = 0.08 K/W
$R_{\theta TIM2} = R_{\theta TIM3} = R_{\theta TIM4}$ = 0.2 cm^2 K/W
A_{TE} = 64cm^2, representing the area of four TECs, each 4x4 cm
A_{switch} = 9cm^2

Using the TEC characteristics in figure 5.57, calculate the switch case temperature with and without a TEC, and the effective thermal resistance in each case.

Solution

The TEC adds additional thermal interfaces but also provides a negative ΔT in the thermal path.

For the ***no-TEC case***, the switch heat load is 120W from a 30 x 30 mm package. The package is mounted to a heat sink using thermal grease with a TIM2 thermal resistance of 0.2 m^2.°C/W. The sink to ambient thermal resistance comprises two components; a spreading component due to the small heat input area, and the fin to ambient component as shown in Figure 5.59. When the TECs are added, we can use a simpler heat sink, one designed for a uniform heat input, and this spreading in the heat sink is not necessary.

From an Ohms law calculation, the case temperature without a TEC is

$$T_c = 120W \times (0.2cm^2K/W \times 9cm^2 + 0.08K/W + 0.18K/W) = 73.9°C.$$

This equates to a case to ambient thermal resistance of

$$R_{\theta c\text{-}amb} = \frac{(T_{case} - T_{amb})}{P} = \frac{(73.9°C - 40°C)}{120W} = 0.283°C/W.$$

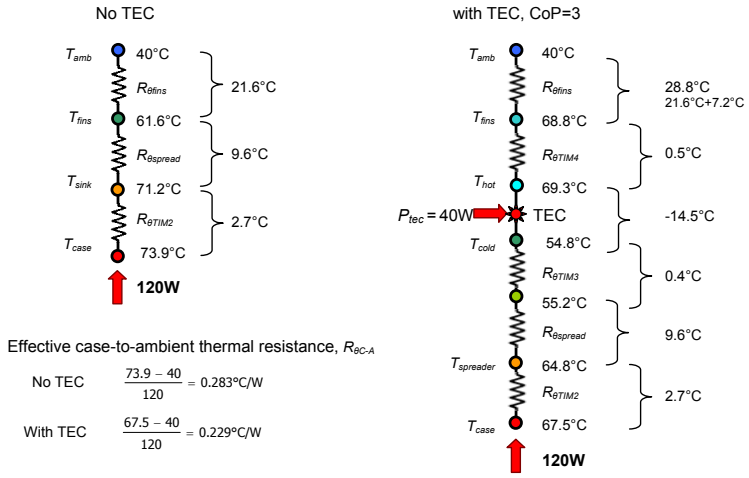


Figure 5.59. *Typical thermal resistor equivalent networks with and without a TEC.*

With a TEC, assume the TECs operates with a COP of 3. For practical applications of switch cooling, the *COP* will likely be between 2 and 3.5, as shown in figure 5.60a, corresponding to TEC ΔT's between 10 and 20°C.

- For a COP greater than 3.5, sufficient ΔT is simply not generated by the TEC to offset the additional interface and heat sink rise and still provide measurable performance improvement.
- For a COP less than 2, considerable additional input electrical power must be supplied and consequently dissipated by the heat sink.

For the 120W switch heat dissipation, a *COP* of 3 equates to a TEC input power of 40W and a TEC ΔT of about -14.5°C. This 40W must be dissipated by the heat sink in addition to the switch 120W heat load, thus raising the heat sink temperature by 7.2°C. Multiple TECs cover the heat sink area (total TEC area of 64cm^2). Thus the heat sink has a much larger heat input area compared to the no-TEC case thereby eliminating the need for a thick copper base plate or embedded multiple heat pipes to spread the heat. The need for spreading is moved from the heat sink to the cold side of the TEC. This cold side spreader could be a heat pipe assembly, vapour chamber, or solid copper spreader, with are cost, performance, and weight tradeoffs for each case. For this example, a spreading resistance of 0.08°C/W is assumed.
From Figure 5.60a (also see Chapter 22), an optimally designed TEC operating at a *COP* of 3 has a ΔT of about -14.5°C. With the TEC, the case temperature decreases from 73.9°C to 67.5°C, a reduction of approximately 6.4°C, resulting in a 19% reduction in $R_{\theta c\text{-}amb}$ for the 120W case. Note that the TEC

actually operated with a 14.5°C negative ΔT, yet the case temperature was reduced by only 6.4°C. More than half of the negative ΔT from the TEC must be used to offset the additional interface losses and additional rise in heat sink temperature due to the TEC input power.

$$\Delta T_{net} = \Delta T_{TE} + \Delta T_{interfaces} + \Delta T_{heatsink}$$
$$= -14.5 + 0.5 + 0.4 + 7.2 = -6.4°C$$

Also, it should be noted that for the baseline system without the TEC, the 0.283°C/W thermal resistance is a constant with respect to heat load. For the TEC, the 0.229°C/W is not a constant with respect to heat load and accurately represents the thermal resistance for the 120W heat load only.

♣

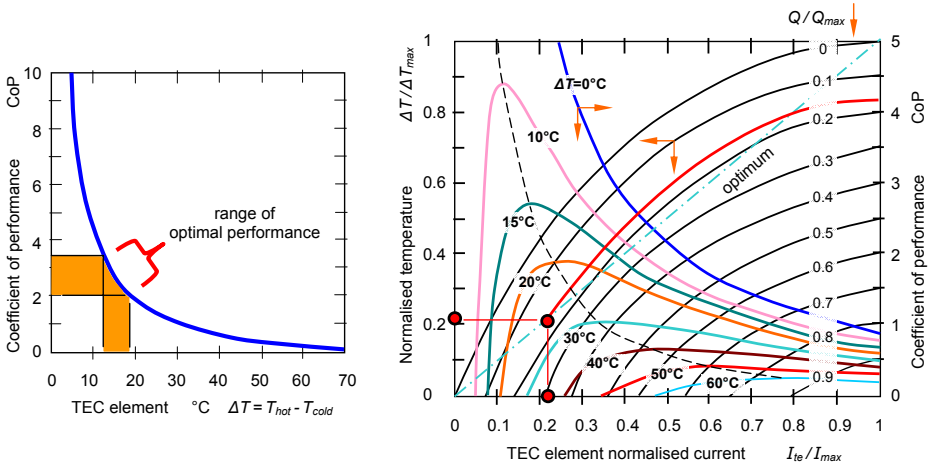


Figure 5.60. *Generic Bi_2Te_3 thermoelectric characteristics: (a) theoretical optimum (maximum) COP vs ΔT and (b) COP as a function of current and generic thermoelectric performance curves.*

Using example 5.17 as a benchmark case, it should be noted when considering utilising a TEC:

- The TEC should operate near its maximum (optimum) COP for the ΔT chosen. As shown in Figure 5.60b, off-optimum operation results in significantly higher TEC power consumption (lower *COP*). This results in a larger ΔT across the hot side thermal interface and a larger additional rise in heat sink temperature, thus reducing the net gain from the negative ΔT from the TEC. In addition, for the same COP and thus the same input power, an off optimum TEC will not achieve the same ΔT, thus further reducing the ΔT_{net}. The performance penalty for selecting non-optimum TECs and running them off-optimum is severe.
- Analysis of Figure 5.59 shows there will be cases when a TEC should not be used - even if it is optimally designed. Even for an optimally designed TEC that behaves according to Figure 5.60a, if the heat sink performance is poor, the additional rise in heat sink temperature, $\Delta T_{hs\text{-}amb}$ will offset more of the negative ΔT from the TEC. For example, a poor heat sink with a fin to ambient thermal resistance of 0.3°C/W instead of 0.18°C/W (with all other parameters in the example kept the same), the additional rise in temperature of the heat sink due to the necessary dissipation of the TEC power would increase from 7.2 to 12.0°C (0.3°C/W x 40W). The net gain from adding the TEC drops to only -1.6°C. Such a small gain would not justify the added cost and complication of adding a TEC.
- The calculations in Example 5.17 were performed at the design power heat dissipation at the maximum ambient conditions 40°C. Typically, the switch may not operate for prolonged periods at these extreme conditions. When the switch heat dissipation is less than 120W or when the ambient is less than the maximum, the realized TEC *COP* can be significantly greater, rising exponentially as the required TEC ΔT is decreased (see Figure 5.60a). COPs higher than 5 would be likely under typical operating conditions when only small ΔT would be required by the TEC in order to maintain the maximum junction temperature below its specification. Likewise, operating the TEC at a constant input power under these non-peak conditions offers the opportunity for significant reductions in fan speed with favourable acoustic benefits.

TEC requirements

In the description of the TE enhanced heat sink concept, multiple TECs and a large spreader were utilized. For four TECs pumping 120W, each TEC in Example 5.17 pumps 30W. The generalized thermoelectric performance curve given in Figure 5.60c give an indication of the type of TEC required. For a *COP* of 3 operation, the optimum ΔT is 14.5°C (figure 22.17). For a typical TEC with a ΔT_{max} of around 68°C, the ratio of $\Delta T/\Delta T_{max}$ is about 0.21. From Figure 5.60c, for optimum (lowest power) operation, P/P_{max} is approximately 0.17. Therefore, for a 30W TEC heat load per TEC, the P_{max} for each TEC should be approximately 30W/0.17 = 176W.

In addition, the four TECs and a large cold side spreader add cost and weight to the system. A single TEC with partial spreading of the heat prior to the TEC and additional spreading between the TEC and heat sink could offer some performance/cost tradeoffs. However, if the entire heat from the switch were pumped by a single TEC, the P_{max} of that single TEC would need to be approximately 700W, (120W/0.17), which exceeds available limits of about 550W.

5.21.2 Superlattice and heterostructure cooling

The strategy to improve thermoelectric cooling has turned to the nano scale level, where coherent and incoherent transport plays an important role in electron and phonon diffusion. ZT values (a figure of merit parameter, see Chapter 22) between 2 and 3 at room temperature are obtained with Bi_2Te_3/Se_2Te_3 superlattices. Cooling power density is as high as 700W/cm^2 at 353 K compared to 1.9W/cm^2 in the bulk material, figure 5.61.
Thin-film, based on SiGe/Si, gives a cooling power density of almost 600 W/cm^2 for a temperature difference of 4K below ambient for a 40 x 40 micrometer area. A superlattice approach produces a ZT larger than 1.4.

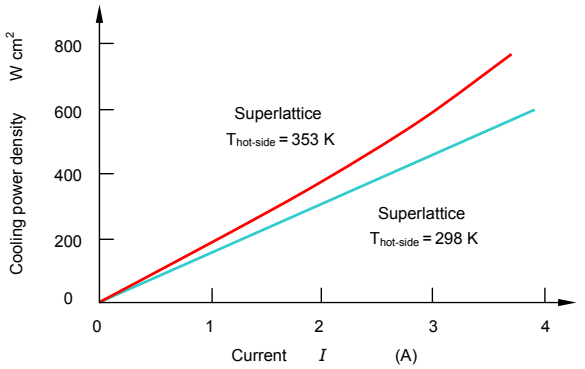


Figure 5.61. *Estimated power density for superlattice devices as a function of current.*

5.21.3 Thermionic and thermotunnelling cooling

Thermionic cooling is based on the principle that a high-work-function cathode preferentially emits hot electrons. Materials available have a work function of 0.7eV or higher, which limits the use to the higher temperature ranges (>500K). Vacuum thermionic devices based on resonant tunnelling have cooling capabilities of 20 to 30°C with kW/cm^2 cooling power densities achievable. However, since the operating currents for the device are as high as 105A/cm^2, effects such as Joule heating (I^2R) at the metal-semiconductor contact resistance and reverse heat conduction limit cooling to <1°C.
Devices based on quantum tunnelling through a small gap, with the spacing between the cathode and the anode of the order of 10 nm provide much larger cooling power than thermoelectric superlattice coolers, specifically 10kW/cm^2 for 50K cooling at room temperature.

5.22 Cooling by phase change

Phase change materials and heat accumulators
Phase change materials use for electronics thermal management is limited to applications where time-dependent phenomena play a role.

In the case of *heat accumulators,* the use of composite materials based on a granulated open-porous matrix filled with a hygroscopic substance is an approach to accumulate heat. The advantage is a significant increase in the heat that can be stored as compared to sensible heat (surrounding heat, potential energy) and latent heat. For example, for a 100°C temperature rise, copper absorbs 40kJ/kg. Evaporation of water is associated with an absorption of 2260kJ/kg. The enthalpy of a reversible chemical reaction can reach a value of 7000kJ/kg. A principal advantage of reversible chemical reactions for heat accumulation is their ability to store the accumulated energy for a long time, if the reaction is controlled by the presence of either a catalyst or a reagent.

Phase change material thermal properties

When power electronics is operated under transient conditions, increasing the thermal capacitance is a useful technique for limiting temperature increases and/or minimizing the performance requirements of a heat sink. An effective method of increasing thermal capacitance is to utilise a material that undergoes a change of phase at a desirable temperature – at which temperature there is thermal energy transfer associated with the change of phase. Utilizing phase change for temporary thermal energy storage has the benefit of allowing heat rejection to occur over a longer time period, which in turn allows a smaller heat sink. Examples of phase change that have been used in power electronics cooling include solid-liquid, liquid-vapour, and solid-solid (for example, crystalline structure to amorphous). A solid-liquid phase change is most common for systems that require reuse of the phase change material (PCM). The thermal energy required to melt the PCM is described as the *latent heat of fusion* and *latent heat of vaporisation* is applicable to a liquid to gas phase change. Directing the waste heat from the power electronics into the PCM can result in a near isothermal heat sink while the PCM is changing state - melting or vaporising.

Selecting a PCM requires knowing the range of expected temperatures (the melt or vapour temperature of the PCM must be high enough such that state change does not occur until needed). The list of candidates is restricted when issues such as material compatibility, toxicity, availability of thermal property data, and cost are considered.

Table 5.26: General solid-liquid PCM thermo-physical characteristics

property		Organic paraffin	Organic non-paraffin	Inorganic salt hydrate	Inorganic metal eutectic
Latent heat of fusion h_f	kJ / kg	230 - 290	120 - 240	170 - 340	30 - 90
Latent heat of vap h_{fv}	J / m^3 x 10^6	190 - 240	140 - 430	250 - 660	300 - 800
Density ρ_f	kg / m^3	≈ 810	900 -1800	900 - 2200	≈ 8000
Thermal conductivity λ	W / mK	≈ 0.25	≈ 0.2	0.6 - 1.2	≈ 20
Thermal expansion		high	moderate	low	low
Congruent melt		yes	some do	most do not	yes
Supercool		no	no	most do	no
Corrosion		low	some are	highly	some are
Toxicity		no	some are	highly	some are

Heat of fusion (solid to liquid)

Desirable characteristics of a solid-liquid PCM include high heat of fusion per volume, congruent melting and freezing characteristics, high thermal conductivity, minimal supercooling, and low thermal expansion. Table 5.26 lists four categories of solid-liquid PCMs with melting temperatures in the range of 40° to 120°C.

The main drawback of metal eutectic PCMs is high density, resulting in heavy package solutions. Salt hydrates have handling and safety issues. Two salt hydrates with high volumetric heat of fusion are Lithium Nitrate Trihydrate (679×10^6 J/m^3) and Barium Hydroxide Octahydrate (656×10^6 J/m^3). The first is a severe oxidizer with a moderate health rating and the second carries a severe health. The organic non-paraffin category has the largest variability. Acetamide is an example of a non-paraffin organic PCM that melts at 81°C and has only moderate handling and material compatibility issues.

The organic paraffins represent a reasonable compromise between handling and performance and are available in a wide range of melt temperatures. In general, higher melt temperature paraffins are more expensive, especially for high purity levels. The low thermal conductivity of these materials usually requires the use of an imbedded matrix material to help conduct heat into the PCM. The volume change between solid and liquid states limits their packaging density.

Heat of vaporization (liquid to vapour)

The energy required to change a substance from a liquid state to a vapour state is termed the *heat of vaporization*, and is directly related to overcoming the intermolecular bonding force in the liquid state. The energy released when a vapour is condensed to a liquid is numerically equal to the heat of vaporization but has an opposite sign and is commonly termed the *heat of condensation*. The most common applications of this thermodynamic property in power electronics cooling are associated with boiling heat transfer.

Compared to solid to liquid phase change, the volume change associated with liquid to vapour phase change is significantly larger, often greater than two orders of magnitude. Even though heats of vaporization are typically larger than heats of fusion for a particular material, the large volume change limits the ability to take advantage of this property for simple thermal energy storage. When it is not required to contain and reuse the vapour, such as in a single use application, expendable coolants that undergo a liquid-vapour phase change as they absorb heat are a consideration.

More common in power electronics cooling, the use of heat of vaporization occurs in closed systems where heat removal is augmented by the vaporization property and then released elsewhere in the system when the vapour is condensed. Examples for cooling hardware using this property include heat pipes, pumped refrigerant, and spray cooling. The operating pressure of these systems can be adjusted to provide an attractive boiling point for the coolant. Desirable qualities for a coolant that will operate in a two-phase mode include a high heat of vaporization, acceptable boiling temperature and pressure, low corrosion potential, low toxicity, environmentally friendly, and low cost. The heat of vaporization and melting temperature for several common substances at standard temperature and pressure are listed in Table 5.27.

Table 5.27: Heat of vaporization at standard temperature and pressure

Substance	Boiling Point	Heat of Vaporization
	°C	kJ / kg
Helium	-268.9	21
Hydrogen	-252.8	461
Nitrogen	-195.8	199
Propane	-42.1	427
R12	-38.8	165
Ammonia	-33	1369
R134a	-26	178
Fluoroketone	49	88
Acetone	56	518
FC-72	56	88
Methanol	64.7	1100
Ethanol	78.4	846
Water	100	2256
Ethylene glycol	197	800
Glycerin	290	974
Mercury	347	295

Reading list

Fishenden, M. and Saunders, 0. A., *An Introduction to Heat Transfer,* Oxford University Press, 1982.

http://www.electronics-cooling.com

http://www.qats.com/qpedia.asp

http://www.lytron.com/

http://www.aavidthermalloy.com/

http://www.1-act.com/

5.23 Appendix: Comparison between aluminium oxide and aluminium nitride

Aluminium Nitride is made by nitridation of aluminium or by conversion of alumina Al_20_3 to aluminium nitride. It is a covalently bonded material and has a hexagonal crystal structure. Because of its resistance to sintering, an oxide-forming additive such as Y_20_3 is needed to form a substrate. General comparison properties of aluminium oxide and nitride are listed in Table 5.28.

Aluminium nitride is a cost effective, non-toxic alternative to beryllium oxide and has a thermal conductivity nearly eight times higher than alumina (Aluminium oxide). Advantages of aluminium nitride include good thermal performance, low thermal expansion, and non-toxicity. Aluminium nitride offers:

- High thermal conductivity
- Low thermal expansion coefficient closely matching Silicon
- Good dielectric strength
- High electrical sensitivity
- Low toxicity and therefore excellent replacement for Be0
- Good shock and corrosion resistance
- Low dielectric loss
- High temperature stability
- High flexure strength and light weight
- Resistant to wafer processing gasses and plasma erosion
- Conducive to finishing operations such as lasering, lapping, and polishing
- Substrate for direct bond copper DBC and filled vias
- Good adhesion for thin and thick film applications
- Uniform lapped and polished surfaces for resistor networks
- Polished and lapped surface finishes to 12nm, R_a, with minimum pullouts
- Lapped surface finishes to 150nm, R_a, where is the roughness average profile

$$R_a = \frac{1}{n}\sum_{i=1}^{n}|y_i|$$

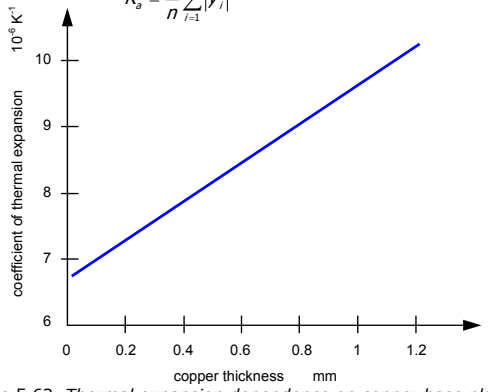


Figure 5.62. *Thermal expansion dependence on copper base plate thickness.*

Table 5.28: Substrates properties

Material		Aluminium oxide	Aluminium nitride	Beryllium oxide	Silicon carbide
symbol		Al_3N_4	Al_2O_3		
Colour		Tan/Gray	white	white	
Melting point/ maximum use temperature	°C	>2200/800 oxidizing	2054/1700	2507/1800	-/1650
Purity	(wt %)	98%	99.6%	99.5	
Density	(g/cm^3)	>3.25	>3.65	2.85	3.2
Thermal Conductivity	(W/m. K)	100 - 300	27	265	270
Thermal Expansion	(x10^{-6}/K) @ >20°C	<4.3	<7.7	8.0	3.7
Specific Heat	(J/kg K)	740	880		750
Dielectric Strength	(kV/mm)	>15	>12	9.5	
Dielectric Constant	(at 1MHz) @ 20°C	8.7	9.2	6.6	40
Dissipation Factor	(x10^{-4} @ 1 MHz)	3 - 7	3	4	
Volume Resistivity	(ohm-cm)	>10^{14}	>10^{12} at 20°C >10^{8} at 500°C	10^{15}	10^{2}-10^{6}
Flexural Strength	(kg$_f$/mm) MPa	>250 300	>280 345	120-150 150-200	450 550
Modulus of elasticity	GPa	331	372	345	410
Hardness (Knoop)	GPa	11.8	14.1	9.8	27
Poisson's ratio		0.22	0.21	0.26	0.14
toxicity		nontoxic	nontoxic	toxic	nontoxic
Substrate Specification					
Maximum Dimension	(mm)	140 x 100mm,	100 x 200		
Thickness	(mm)	0.63 - 0.2 mm	0.63 - 0.30mm		
Surface Roughness	(micron)	as fired - 0.3 as lapped - 0.075 as polished - 0.025	as fired - 0.3 as lapped - 0.075 as polished - 0.025	as fired - 0.3 as lapped - 0.1 as polished-0.08	

At room temperature, the thermal conductivity of aluminium nitride ceramics is independent of Al_3N_4 grain size or number of grain-boundaries, but is controlled by the internal structure of the grains, such as the degree of oxidation (oxygen contamination).
The coefficient of thermal expansion for direct copper bonded (DCB) substrates with a layer of 0.6 mm alumina sandwiched between Cu layers of various thicknesses, as shown in figure 5.62.

5.24 Appendix: Properties of substrate and module materials

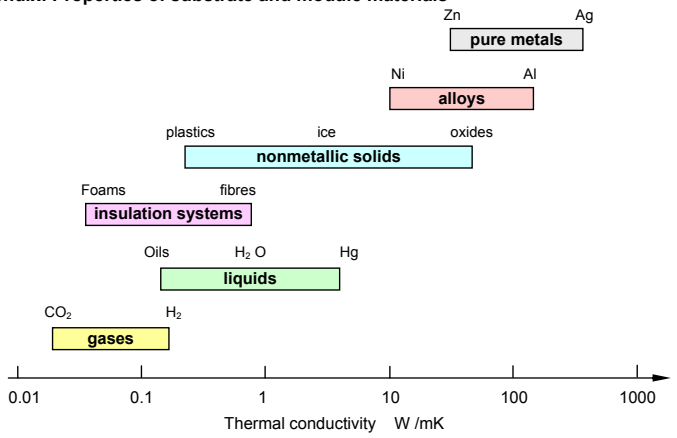


Table 5.29: Coolant properties

Material	Specific Heat	Density	Thermal Conductivity
	c_p	ρ_ℓ	λ
	J/kgK	kg/m^3	W/mK
Coolanol 25	1,838	903	0.12
Diala-X	1,840	870	0.14
20/80 EGW Solution	3,817	1023	0.57
50/50 EGW Solution	3,283	1064	0.39
Fluorinert®, FC-77	1,028	1771	0.063
Hydraulic Oil	1,842	868	0.12
Polyalphaolefin	2,180	794	0.14
SAE 10W Oil	1,901	875	0.12
SAE 30W Oil	1,901	875	0.12
Stainless Steel, 316	500	8025	16.20
Water, H_2O	4,184	998	0.59
Air	1,008	1.1	0.27
Mercury	137	13,800	137

Table 5.30: Properties of module materials

material	relative permittivity	dielectric loss factor	specific thermal conductivity	linear thermal expansion coefficient	temperature coefficient of ε_p	type
	@1MHz	10GHz, 25°C		@25°C		
	ε_r	$tan\delta_\varepsilon$	λ	$\Delta l / l \ / \Delta T$	$\Delta\varepsilon / \varepsilon\Delta T$	
		$\times 10^{-4}$	W/m K	10^{-6}/K	10^{-6}/K	
$A\ell_2 0_3$ 99.5%	9.8	1	37	6.5	136	insulator
sapphire	9.4	1	42	6	110	insulator
Quartz glass	3.78	1	1.7	0.55	13	insulator
Beryllium oxide ceramic Be0	6.3	60	210	6.1	107	insulator 8900kg/m^3
GaAs	12.9	20	46	5.7		semiconductor
Silicon $\rho = 10^3$ Ωcm	11.9	150	145	4.2		semiconductor
PTFE	2.2	3	0.2	106	350	plastic
polyolefin	2.32	7	0.5	108	480	plastic
copper	1.0		393	17	24.1	metal
aluminium	$-1300+i1.3\times10^{14}$		220	23.8		metal

Table 5.31: Power electronic component properties

material		thermal conductivity	specific heat capacity	density	temperature coefficient of expansion	Melting point
20°C		λ	c	ρ_ℓ	CTE	
		W /m K	J/ kg K	kg/m^3	ppm/K	K
air (STP)		0.026	1004	1.2	-	
silicon	Si	120	700	2330	3.5	1685
solder	PbSn	50	150	8400	24.1	183
copper	Cu	385	385	8930	17	1358
alumina	$A\ell_2 0_3$	22	80	3720	6.5	
aluminium nitride	$A\ell_3 N_4$	170	725	3300	4.5	
aluminium silicon carbide	AℓSiC	170		3000	7	
polyimid		0.2	1100	1400		
dielectric layer		0.3	1400	1120		
encapsulation		0.5		2000		
aluminium	Aℓ	205	900	2710	22.5	775
gold	Au	315	126	19320	14.2	1336
platinum (90%) Iridium (10%)		31.1				
platinum	Pt	70.9	1448	21450	9.0	1728
tungsten	W	188	130	19300	4.6	3410
molybdenum	Mo	140	250	10200	4.9	2610
lead	Pb	235	130	11340	23	327
tin	Sn	66	227	7300	23	232
Cu/Mo/Cu					5.8	
silicon carbide	SiC	700	250	3.21	3.7	
sapphire		2700	419	3900	8.4	2040
diamond		2300	509	3500	2.4	3100
brass	CuZn	111	343	8490	18.0	920
steel (low carbon)		48	460	7850	11.5	1370
Mica K Mg_3 $AlSi_3$ O_{10} $(OH)_2$		0.6				1700
beryllium	Be	230	1088	2880	5.9	1280
silicon thermal grease		0.8	2093	2.8	-	
thermal conducting plastic		20				

Table 5.32: Substrate characteristics

Material	Al_2O_3, BeO, AlN, Quartz, Silicon, Sapphire, Ferrite
Surface Finish	Al_2O_3 As Fired, 0.05µm maximum;
Polished	to 0.12µm0
Dimensions	0.5mm x 0.5mm to 100mm x 100mm
Dimensional Tolerance	± 0.01mm scribed; (± 0.002mm saw cut)
Thickness	0.012mm inches to1.2mm
Thickness Tolerance	0.12mm standard to as tight as 0.04mm
Sputtered Resistor Material	NiCr, Ta_2N Sputtered Metallization Ti, TiW, Pd, Ni, Au, Al
Electro-Plated Metals	Au, Cu, Ni, Solder
Electroless Plated Metals	Sn, Ni, Au

5.25 Appendix: Emissivity and heat transfer coefficient

Emissivity is not only a material property but also a surface property, at least for opaque materials. Consequently, coatings (oxides, grease, and water film) influence the value measured under pristine conditions. For example, the emissivity of a copper surface covered with 2 µm oxide increases from 0.03 to 0.2. In practice, surfaces that are initially shiny are covered with oxide and dust after one year of operation. Additionally, surface texture can influence the emissivity because of a strong dependence on the angle.
For heat transfer calculations, total hemispherical emissivity is found by integration over all wavelengths and all angles. What is being measured with an IR camera is the normal spectral emissivity. To use an IR camera to measure the emissivity is not recommended because it returns the normal emissivity restricted to the wavelength band of the detector.

As a rule of thumb: for unpolished metals the ratio of hemispherical to total emissivity is 1.1 to 1.3, and for non-conductors 0.95 to 0.97. Another angle-dependent difference between metals and insulators is the fact that under a shallow incident angle, the emissivity of metals tends to one and insulators tend to zero.
To make things more complex, the emissivity of many materials of interest is strongly dependent on the wavelength. In other words, there is a big difference between the visible band and most common IR bands (0.8 to 3 µm, 3 to 8 µm, 8 to 14 µm). Si and Ge are notorious examples. It is relevant to low-temperature applications (<100°C) because the bulk of the radiation is in the long-wavelength region: 8 to 14 µm. This is the main reason why the colour of paints is not relevant for heat sinks and covers. All colours are black. Apart from the fact that it is simply impossible to account for all the physics of radiation in a practical situation, the problem is only relevant when radiation is an important contribution to the total heat transfer. For natural convection cases, this contribution might well be 40%. When the choice is between an emissivity of 0.05 and 0.1, the radiation contribution is 2 to 4% - minimal. When the choice is 0.7 to 0.9, matters are different. Most heat transfer computer codes that support radiation heat transfer, start from the following set of simplified assumptions:

- Surfaces diffuse, grey (sometimes specula [mirror-like]), opaque
- Surfaces isothermal
- Surfaces uniformly irradiated
- Medium is transparent for all relevant wavelengths

In summary, differences exist between your experimental data and simulation results when radiation is a major factor, even when all other data are known to within 1%.
Table 5.33 shows typical normal total emissivity values @ 20°C, unless stated otherwise.

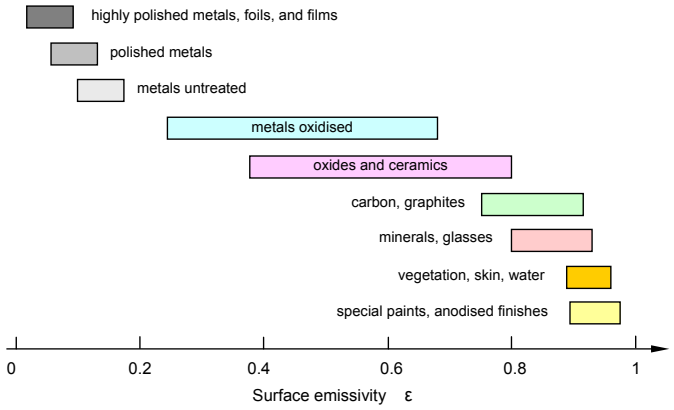
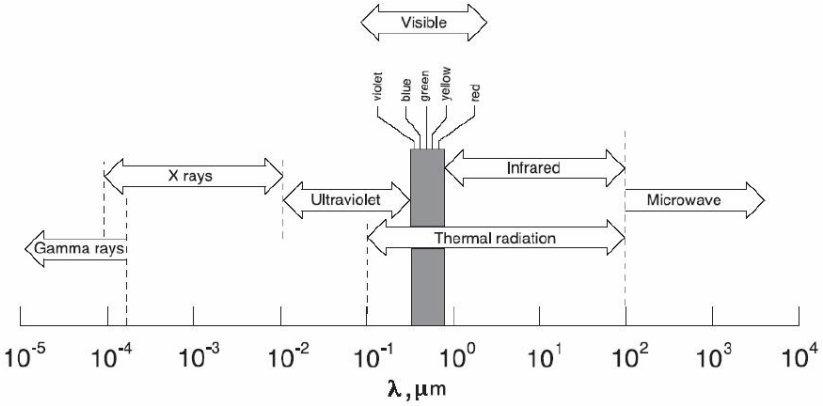


Table 5.33: Normal total emissivity values @ 20°C

Metals		ε	Coatings	ε
Aluminium	polished	0.04	Aluminium bronze	0.3
	sheet	0.09	Aluminium paint	0.35
Brass	polished	0.05	Blackbody paint	0.97
	oxidized	0.22	Enamel	0.82
Chromium	polished	0.06	Lampblack	0.95
	rough	0.74	Paint	0.89
Copper	polished	0.03		
Gold	polished	0.025	**Various**	
Graphite	polished	0.42	Glass, Quartz	0.93
Inconel	polished	0.2	Ice	0.98
Iron	polished	0.06	Plastics	0.8
	oxidized	0.85	Paper	0.8
	ground	0.24	Porcelain	0.92
	cast	0.16	Silk, Cotton, Wool	0.75
Mercury		0.09	Stone	0.8-0.9
Molybdenum	polished	0.05	Water (> 0.1mm)	0.95
	2600 K	0.29	Wood	0.9
Silver	polished	0.025		
Steel	polished	0.06		
	oxidized	0.6		
Silicon	difficult	0.3 - 0.8		
Tin	bright	0.07		
Tungsten	polished	0.05		



Electromagnetic spectrum

Natural convection heat transfer coefficient

The convective heat transfer coefficient h, for various geometry arrangements are given by equations (5.93) and (5.94), as applicable, in conjunction with Table 5.34.

$$h = k_h \left(\frac{\Delta T}{D}\right)^{1/4} \quad (5.93)$$

$$h = k_h \left(\frac{\Delta T}{L}\right)^{1/4} \quad (5.94)$$

where ΔT is temperature difference, K
L is length, m
D is diameter, m

Table 5.34: Heat transfer coefficient constant, k_h

Geometry	Dependant geometrical parameter		constant k_h	equation
Vertical cylinder or plate	Vertical height L		1.42	(5.94)
Horizontal cylinder	Diameter D		1.32	(5.93)
Horizontal plate	$L = \frac{4 \times area}{perimeter}$	Upper surface hot	1.32	(5.94)
		Lower surface hot	0.59	
Vertical populated PCB	Vertical height L		2.44	(5.94)
Sphere	Diameter D		1.92	(5.93)

5.26 Appendix: Ampacities and mechanical properties of rectangular copper busbars

Effect of emissivity and number of busses on ampacity (current carrying capacity) – data in Table 5.35 shows how higher emissivities improve ampacity. Multiple busses also affect ampacity in a nonlinear relationship. Ampacity may be raised by increasing heat dissipation with convection cooling or surface treatments. Surface treatments which improve emissivity are oxidation or thinly coated, flat, inorganic based spray paints.

Table 5.35: Ampcapacity

number of 6mmx100mm busses	Ampacity, A											
	30°C rise				50°C rise				65°C rise			
	emissivity ε				emissivity ε				emissivity ε			
	0.15	0.4	0.7	0.9	0.15	0.4	0.7	0.9	0.15	0.4	0.7	0.9
1	1100	1250	1400	1600	1500	1700	1900	2000	1700	1950	2200	2300
2	1900	2050	2200	2300	2550	2750	2950	3100	2950	3200	3400	3600
3	2500	2700	2850	3000	3400	3600	3850	4000	3950	4200	4500	4600
4	3100	3300	3450	3600	4200	4400	4700	4800	4900	5100	5400	5600
6mm spacing. Ampacities of bus bar systems of other configurations must be calculated taking into account size, spacing, number of bus bars, and overall skin-effect ratio.												

5.27 Appendix: Isolated substrates for power modules

Currently used isolated substrates for power modules are:

Isolation material

Ceramic: aluminium oxide Al_2O_3
aluminium nitride AlN
(beryllia oxide Be0)
(silicon nitride Si_3N_4)

Organic: epoxy
polyimide (Kapton)

Substrates

Metal sheets: *DCB* (**D**irect **C**opper **B**onding)
AMB (**A**ctive **M**etal **B**razing)

Metal sheets: IMS (**I**nsulated **M**etal **S**ubstrate)
Multilayer-IMS

Thick film layers: TFC (**T**hick **F**ilm **C**opper)

*DCB (**D**irect **C**opper **B**onding)*
Power modules with IGBTs (or MOSFETs) and freewheel diodes commonly use substrates made of *DCB*-ceramics with Al_2O_3 or AlN isolation that combine good thermal conductivity and high isolation voltage. For DCB, copper surfaces 200µm to 600µm thick, 300µm thick, are applied to the top and bottom surfaces of the isolation substrate material (0.25mm to 0.85mm thick, typically 0.5mm thick) by eutectic melting at between 1065°C and 1083°C. The sandwiched copper oxide layer helps adjust for the different thermal expansion rates. After the necessary track structure for the module circuitry has been etched into the top side copper surface, the chips are soldered on, and contact connection on the chip top side is effected by bonding. The bottom side copper of the DCB-ceramic substrate is fixed to the module base plate (usually 3mm thick copper) usually by soldering, as seen figure 5.8. Other module types do not necessarily require a base plate and the soldering procedure may be avoided. In these modules, the DCB-substrate is pressed on to the heatsink by means of a suitable case.

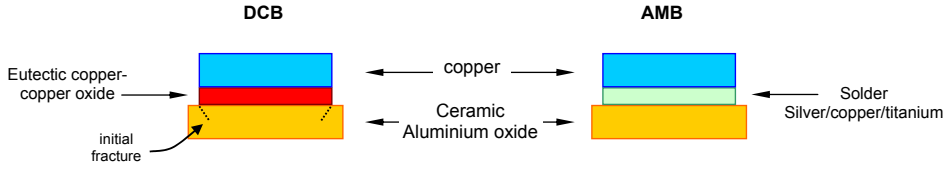


Figure 5.63. *Direct copper bonding, DCB and active metal brazing, AMB.*

Advantages of the DCB-technology compared to other structures are mainly the high current conductivity due to the copper thickness, good cooling features due to the ceramic material, the high adhesive strength of copper to the ceramic (reliability), and the optimal thermal conductivity of the ceramic material. Possible failure due to cracking, termed conchoidal fracture, starts at the copper edge, as shown in figure 5.63, and progressively extends under the copper interface area.

*AMB (**A**ctive **M**etal **B**razing)*
The AMB process (*brazing* of metal foil to a substrate) has been developed based on DCB technology. The advantages of AMB-substrates with AlN-ceramic materials compared to substrates with Al_2O_3-ceramic materials are lower thermal resistance, lower coefficient of expansion, and improved partial discharge capability.
Figure 5.63 illustrate the differences between DCB and AMB.

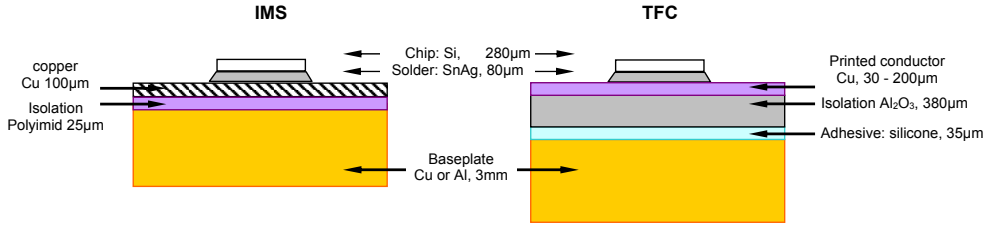


Figure 5.64. *Basic module structure of: (a) an IMS power module and (b) a TFC power module.*

*IMS (**I**nsulated **M**etal **S**ubstrate)*
IMS is mainly used in the low-cost, low-power range and is characterized by direct connection of the isolation material to the module base plate. For insulation, polymers (such as epoxies, polyamides) are applied to an aluminium base plate, as seen in figure 5.64a. The upper copper layer is produced in foil form and glued onto the isolation substrate (similar to PCB production) and is patterned by etching.
Advantages of IMS are low costs, filigree track structure (possible integration of driver and protection circuitry), substrate high mechanical robustness, and relatively wide substrate areas, compared to DCB.

The thin isolation layer, however, leads to comparably high coupling capacitances associated with the mounting surface. Also the thin upper copper layer only provides a comparably low heat spreading, which is improved by additional metallised heat spreading layers under the chips or by adding Al-particles to the isolation layer.

*TFC (**T**hick-**F**ilm-**C**opper)-thick film substrates*
Just as with DCB, the basic material for thick film substrates is an isolation ceramic, which is glued directly on to the base plate or a heatsink by means of silicone or applied by soldering, as shown in figure 5.64b. The tracks on the top of the ceramic substrate are made of copper and are applied by screen printing. The power semiconductor chips or other components are soldered or glued on to the track pads.

TFC technology can also be combined with standard thick film technology. Since low resistances may be produced by the paste materials which are usually applied in thick film technology, and since isolated tracks can be arranged on top of one another and connected together, quite a number of system components may be densely integrated. However, the filigree tracks, typically 15µm thick, limit the current capability of such structures to about 10A.

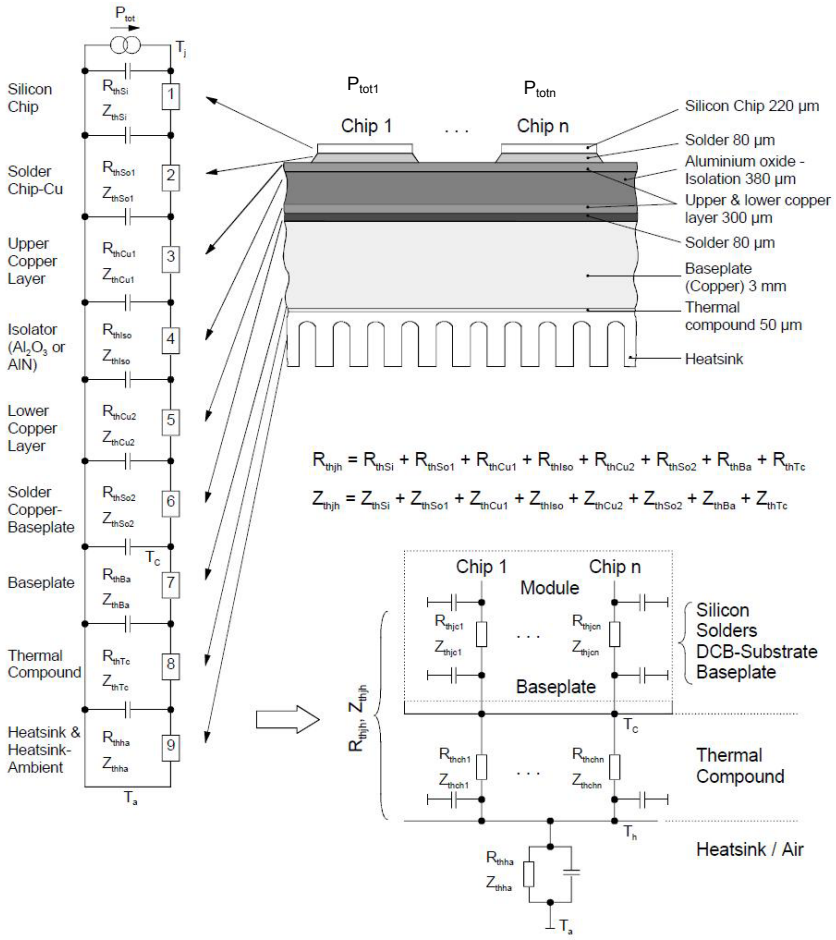


Figure 5.64. *Power module with thick film substrate: (a) basic structure and (b) thermal model.*

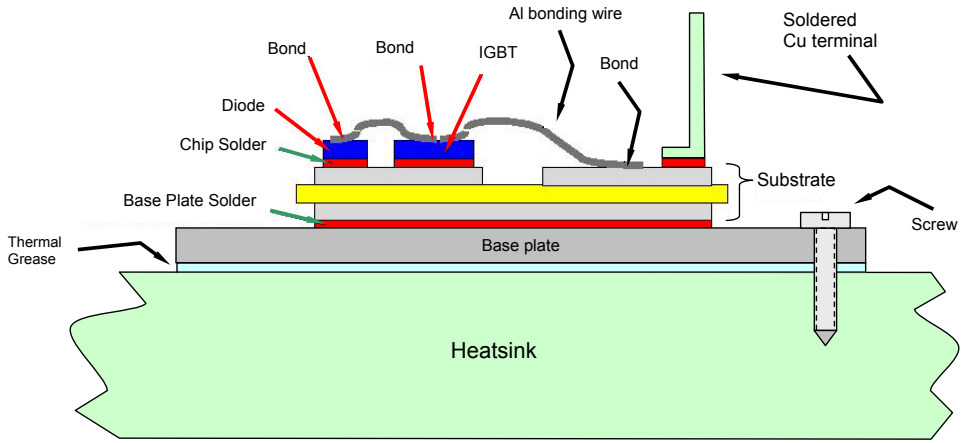


Figure 5.65. *Power module with DCB substrate: (a) basic structure and (b) thermal model.*

High-temperature lead-free transient liquid phase (TLP) die and substrate attach methods

While silicon semiconductor technology is limited to junction temperatures of about 200°C, emerging SiC technology could exploit 600°C operating temperatures, were it not for die and substrate attachment limitations and aluminium thermal bonding stressing.

The high-temperature, lead-free silver-tin transient liquid phase (TLP) die attach process for connecting the SiC power devices to a nickel-plated direct bond copper (DBC) or direct bond aluminium (DBA) power substrate (aluminium nitride or silicon nitride) shown in figure 5.66a, allows junction temperature operation to in excess of 400°C.
Similarly the high-temperature, lead-free nickel-tin TLP attachment process for connecting the nickel-plated DBC or DBA power substrate to the MMC base-plate allows operation to temperatures in excess of 400°C.

The baseplate of the power module utilizes a lightweight copper-molly (CuMo) metal matrix composite (MMC) that has a coefficient of thermal expansion (CTE) characteristic closely matching that of the SiC power die. This CTE matching reduces thermal-stress mismatches, thus improving the long-term thermal stressing reliability of the power module.

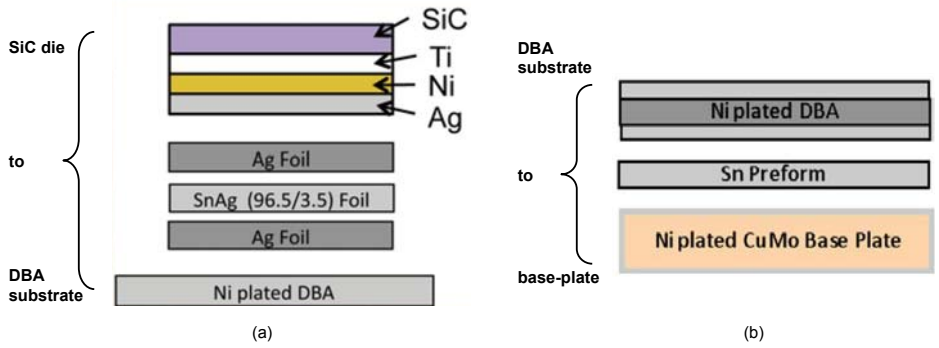


Figure 5.66. *Cross-section of various layers in the lead-free 400°C high temperature SiC power module: (a) Ag-Sn TLP die attach and (b) Ni-Sn TLP substrate attach.*

CONVERSIONS

STATIC PRESSURE

1 mmH_2O = 0.0394 inch H_2O
1 mmH_2O = 9.8 Pa
1 mmH_2O = 25.4 mm H_2O
1 Pa = 0.102 mm H_2O
1 inch H_2O = 249 Pa

1 Pa ≡ 1 N/m^2 = 10^{-5} bar = 10^{-6} N/mm^2 = 0.102 kp/m^2 = 0.987×10^{-5} atm = 0.0075 Torr = 145.04×10^{-6} psi

AIRFLOW

1 m^3/min = 35.31 ft^3/min (cfm)
1 cfm = 0.0283 m^3/min
1 m^3/min = 16.67 litre /s
1 cfm = 0.472 litre /s
1 litre /s = 0.06 m^3/min

1 cfm = 1.7m³/h
1 litre/s = 3.6 m³/h
1 m³/s = 3600 m³/h

Problems

5.1. A thyristor bridge switches at 1 kHz and the total energy losses per thyristor are 0.01 Joule per cycle. The thyristors have isolated studs and a thermal resistance of 2 K/W. The heat sink has a thermal resistance of 1.8 K/W. Calculate the maximum number of thyristors that can be mounted on one heat sink if the thyristor junction temperature is not to exceed 125°C in an ambient of 40°C. What is the heat sink temperature?
[3 devices, T_s= 94°C]

5.2. A transistorised switch consists of two IGBTs and two 1 Ohm current-sharing resistors, as shown in figure 5.67, mounted on a common heat-sink. Each transistor has a thermal resistance $R_{\theta j\text{-hs}}$ of 2 K/W, while each resistor has a thermal resistance $R_{\theta\ r\text{-hs}}$ of 1 K/W. The maximum switching frequency is 1 kHz and the maximum duty cycle is 99.99 per cent. The heat-sink thermal resistance $R_{\theta\ hs\text{-}a}$ is 1 K/W. The energy losses per transistor are 5 mJ/A per cycle. If the ambient temperature is 30°C, maximum allowable junction temperature is 150°C, and the maximum allowable resistor internal temperature is 100°C, calculate the switch maximum current rating based on thermal considerations. What are the operating temperatures of the various components, assuming ideal current sharing?
[6.88 A, T_r = 100°C, T_{hs} = 88°C, T_j = 122.5°C]

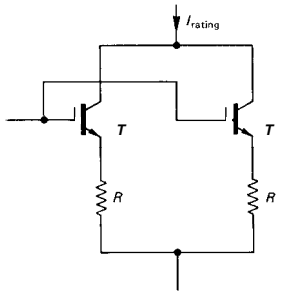


Figure 5.67. *Problem 5.2.*

5.3. Figure 5.68a shows the circuit diagram for a power current sink which utilises a 40V source. Both the IGBTs *T* and wire wound resistors *R* are mounted on a common heat-sink, of thermal resistance $R_{\theta\ hs\text{-}a}$ = 1 K/W. The transistor has a thermal resistance of 2 K/W from the junction to the heat-sink, and 10 K/W from the junction to air via the transistor casing exposed to the air. The resistor has a mounting thermal resistance from the insulated wire to the heat-sink of 1 K/W and 10 K/W from the wire to the air via its casing exposed to the air. The maximum transistor junction temperature is 423 K, the maximum resistor wire temperature is 358 K and the ambient air temperature is 303 K.

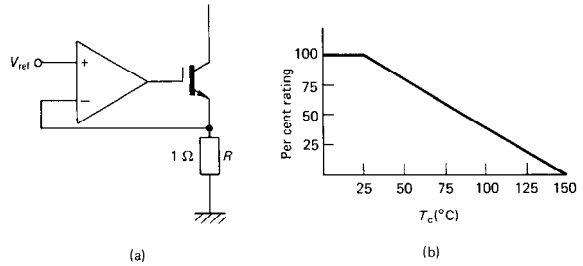


Figure 5.68. *Problem 5.3.*

Based on thermal considerations, what is the maximum current rating of the current sink and under such conditions, what is the heat-sink temperature?
What power rating would you suggest for the 1 Ohm current measurement resistor?
Are there any difficulties in operating the transistor in the linear region in this application if it is in a 120 W dissipation package which is derated according to figure 5.67b?
[1.36 A, 69°C, > 2 W]

5.4. A power IGBT switches a 600 V, 25 A inductive load at 100 kHz with a 50 per cent on-time duty cycle. Turn-on and turn-off both occur in 100 ns and the collector on-state voltage is to be 2 V.
Calculate the total power losses, P_d, of the switch.
The switch has a thermal resistance $R_{\theta j\text{-hs}}$ = 0.05 K/W, and the water-cooled heatsink provides a thermal resistance $R_{\theta hs\text{-}w}$ = 0.05 K/W. Calculate the operating junction temperature if the water for cooling is maintained at 35°C.
The 25 A steady state load current is stepped to 200 A. Calculate the surge power dissipation P_s, at 200 A, assuming transistor switching and on-state characteristics remain unchanged.
The junction temperature for a power surge during steady-state operation is given by case (e) in Table 5.11.
With the aid of figure 5.10, determine the junction temperature at the end of a 0.1s, 200 A pulse. How long is it before the junction temperature reaches $\hat{T}_j$ = 125°C, with a collector current of 200 A?
(Assume $R_{\theta c\text{-hs}}$ = 0).
[175 W, 52.5°C, 1400 W, 112.6°C, 0.5 s]

5.5 Rework example 5.6 finding the case temperature when the switching losses equal the on-state loss.

5.6 A 20kHz, step-down, 340V dc chopper feeds an inductive load with an average current of 20A and a peak-to-peak ripple of 20A. Thus the MOSFET switch on-state current rise from 10A to 30A while the freewheel diode current falls from 30A to 10A when the switch is off. The MOSFET on-state resistance is 0.1Ω and has switch on and off times of 100ns and 200ns respectively. The switch duty cycle is 75% and it has a thermal resistance $R_{\theta\ j\text{-}c}$ of 0.4K/W and is mounted on a heatsink of thermal resistance $R_{\theta c\text{-}a}$ of 0.6K/W in a maximum ambient temperature is 40°C.
Calculate:
i. switching losses, using equations 6.9 and 6.10
ii. switch on-state losses
iii. MOSFET junction operating temperature
[3.4W + 20.4W = 23.8W; I_{rms} = 15.8A, 25W; T_j = 88.8°C]

6

Load, Switch, and Commutation Considerations

Power switching devices are employed for controlling inductive, resistive or capacitive loads. Inductive loads include electrical machines, transformers, solenoids, and relays. High-current in-rush occurs with loads such as incandescent lamps, pulse-forming networks, snubbers, and motors. Incandescent lamps are essentially resistive, but the cold resistive in-rush current during turn-on is 12 to 18 times the steady-state current. This turn-on surge presents special switch-on problems. Capacitive loads, such as fluorescent lighting, also present high-current in-rush at turn-on. Electromechanical loads, such as shakers, present loads that vary between capacitive and inductive over their operating frequency range.

The interaction of the load circuit on the switch arrangement and its commutation depends on three inter-related factors.

- The type of load, usually inductive, and rarely purely resistive.
- Switching mechanism classification, how the load effects switching commutation, namely hard switching, resonant, etc.
- The switch characteristics required to fulfil the supply and load *I-V* requirements, such as a bidirectional current switch, an asymmetrical sustaining voltage switch, etc.

Each of the three factors and their interdependence with the switching mechanisms are considered separately.

6.1 Load types

The two principal load types of general interest in power electronics are

- the resistive load and
- the inductive load.

Turn-on and turn-off voltage and current switching waveforms, hence losses in a switch, depend on the type of load.

6.1.1 The resistive load

A purely resistive load is rarely encountered in power switching applications (other than at load resonance). Figure 6.1 shows a simple resistive load being switched by a common emitter-connected IGBT transistor, which could equally be another appropriate semiconductor switch, for example, a MOSFET. When the gate is driven by the voltage waveform shown in figure 6.2a, the resultant collector voltage and current waveforms are as shown in figures 6.2b and 6.2c. These figures show that at turn-on, as the collector current increases, the voltage across the resistive load increases proportionally, as the collector voltage v_{ce} decreases at the same rate. That is, at turn-on, $v_{ce}(t) = V_s - i_c(t)R_L$, while at turn-off the inverse process occurs. Figure 6.2d shows transistor instantaneous power loss during turn-on and turn-off, which in each case has a peak value of ¼V_sI_m when the collector voltage and current reach half their respective maximum values. The energy loss W during switching is given by

$$W = \int v_{ce}(t)\, i_c(t)\, dt \qquad \text{(J)} \qquad (6.1)$$

where the integration is performed over the switching transition period.

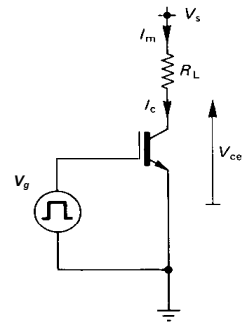


Figure 6.1. *A typical IGBT transistor switching circuit incorporating a resistive load.*

Figure 6.3 shows the safe operating area (SOA) characteristics for an IGBT, on logarithmic axes. Illustrated are the collector switch-on and switch-off trajectories, which are virtually coincident. In the off-state, point A on figure 6.2b, the transistor supports the supply rail voltage V_s while in the fully on-state, point C on figure 6.2b, the collector current I_m is V_s /R_L, neglecting the low on-state voltage of the transistor. During switching the collector voltage and current traverse the *I-V* switching trajectory between the steady-state operating conditions on → V_s /R_L and off → V_s, as shown in figure 6.3.

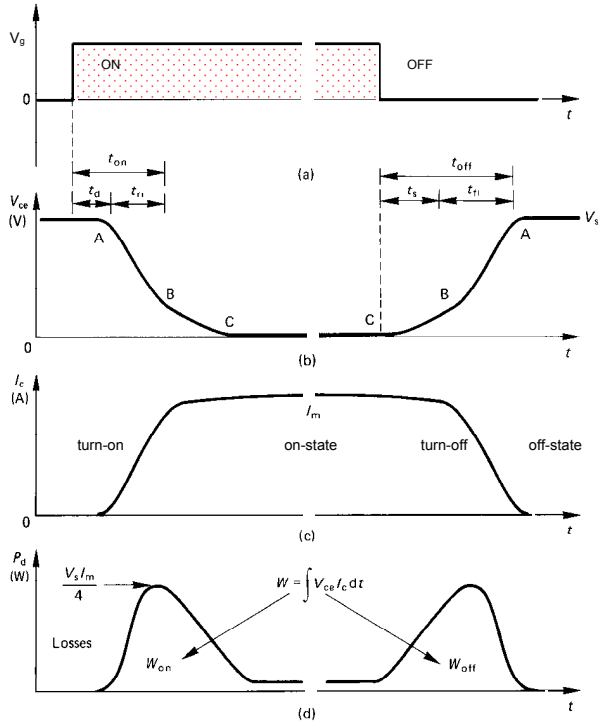


Figure 6.2. *Transistor switching waveforms for a resistive load: (a) on-off gate drive voltage; (b) collector-to-emitter voltage; (c) collector and load current waveform; and (d) instantaneous collector-emitter losses.*

It is important that this trajectory does not exceed the shown SOA bounds set by the device voltage and current limits, and that the SOA region be traversed rapidly. For slow transitions, greater than a few microseconds, power dissipation considerations become the limiting design factor, which is a thermal limitation.

In order to perform the required thermal design calculations (for heatsink determination) it is necessary to be able to specify device-switching losses. To simplify analysis, the switching waveforms shown in figure 6.2 are linearised as shown in figure 6.4. As indicated on these waveforms, the collector voltage fall at turn-on is given by $v_{ce}(t) = V_s(1 - t/t_{on})$ while the collector current rise is $i_c(t) = I_m t/t_{on}$, where $I_m = V_s/R_L$. Combining $v_{ce}(t)$ and $i_c(t)$ by eliminating time t, gives

$$i_c = V_s(1 - v_{ce}/V_s)/R_L \qquad (6.2)$$

As shown in figure 6.3, this describes the linear turn-on transition of slope -1/R_L from the on-state voltage with V_s/R_L collector current, shown as C, to the off-state at A where no current flows and the collector supports the supply V_s. Note figure 6.3 uses logarithmic axes, so the transition trajectory does not appear as a straight line (the inset figure is for linear axes).

Using equation (6.1), the switch-on loss for a resistive load is given by

$$W_{on}^r = \int_0^{t_{on}} V_s\left(1 - \frac{t}{t_{on}}\right) I_m \frac{t}{t_{on}}\, dt$$
$$= \tfrac{1}{6} I_m V_s t_{on} \quad \text{or} \quad \tfrac{1}{6}\frac{V_s^2}{R_L} t_{on} \quad \text{(J)} \qquad (6.3)$$

where $I_m = V_s/R_L$ and t_{on} is the period of the switch-on interval, as shown in figure 6.4.

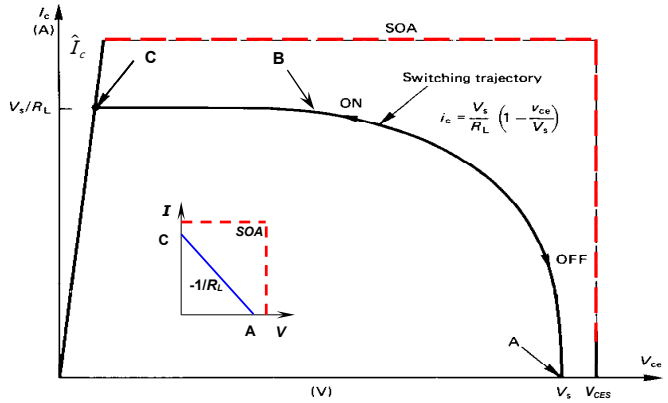


Figure 6.3. *Transistor I-V characteristics showing safe operating area and the switching trajectory with a resistive load, on logarithmic axes, and inset, on linear axes.*

Similarly, using the time dependant collector voltage and current equations shown on figure 6.4a, the turn-off switching loss is given by

$$W_{off}^r = \int_0^{t_{off}} V_s \frac{t}{t_{off}} I_m\left(1 - \frac{t}{t_{off}}\right) dt$$
$$= \tfrac{1}{6} I_m V_s t_{off} \quad \text{or} \quad \tfrac{1}{6}\frac{V_s^2}{R_L} t_{off} \quad \text{(J)} \qquad (6.4)$$

where t_{off} is the turn-off period as shown in figure 6.4.

The average power loss due to switching, which is required for the thermal design outlined in chapter 5, is obtained by multiplying energy loss W by the switching frequency f_s. That is, the turn-on switching loss is given by

$$P_{on} = \tfrac{1}{6} I_m V_s t_{on} f_s \quad \text{(W)} \qquad (6.5)$$

while the turn-off loss is given by

$$P_{off} = \tfrac{1}{6} I_m V_s t_{off} f_s \quad \text{(W)} \qquad (6.6)$$

Because of IGBT current tailing and voltage overshoot at turn-off, the practical switching losses will be larger than those given by the linear approximating methods outlined.

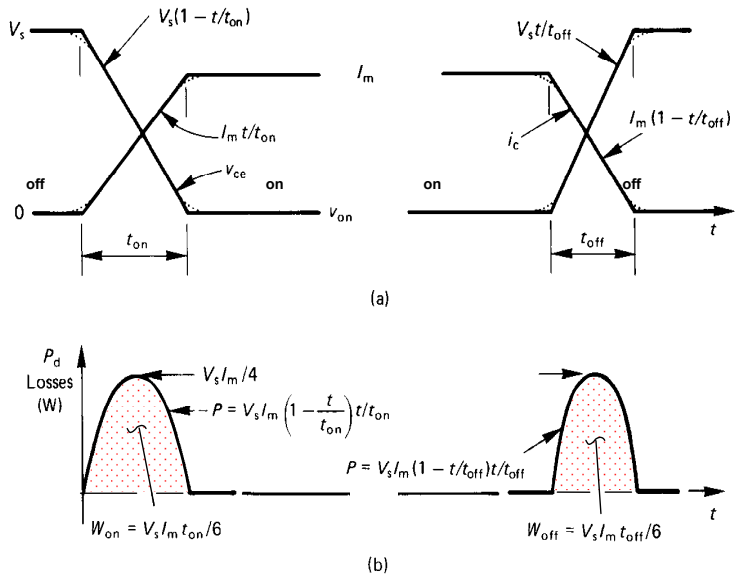


Figure 6.4. *Linear approximations of switching intervals for a purely resistive load: (a) collector voltage and current linear waveforms and (b) corresponding energy and power losses.*

Example 6.1: ***Resistive load switching losses***

An IGBT switches a 10 ohms resistive load across a 100V dc supply. If the switch on-state duty cycle is 25%, ($\delta = ¼$), calculate the average load voltage and current. Calculate the switch losses if the switch-on time is t_{on} =1μs, switch-off time is t_{off}=2μs, and the on-state voltage is 2V.

Solution

When the switch is on, the current in the resistor is $I_L = V_s/R$ = 100V/10Ω = 10A.

The average load voltage is

$$V_o = \delta V_s$$
$$= 0.25 \times 100\text{V} = 25\text{V}$$

The average load current is

$$\bar{I}_o = V_o/R = 25\text{V}/10\Omega = 2.5\text{A}$$

The total switch losses P_T are made up of three components.

P_T = on-state loss + loss at switch-on + loss at switch-off

$$P_T = \delta \times v_{ce} \times I_L + \tfrac{1}{6} V_s I_L t_{on} f_s + \tfrac{1}{6} V_s I_L t_{off} f_s$$
$$= ¼ \times 2\text{V} \times 10\text{A} + \tfrac{1}{6} \times 100\text{V} \times 10\text{A} \times 1\mu\text{s} \times 10\text{kHz} + \tfrac{1}{6} \times 100\text{V} \times 10\text{A} \times 2\mu\text{s} \times 10\text{kHz}$$
$$= 5\text{W} + \tfrac{5}{3}\text{W} + \tfrac{10}{3}\text{W}$$
$$= 10\text{W}$$

Since the off-state leakage current and gate power losses are not specified, it is assumed these are insignificant. Technically the load current should be calculated based on 98V across the load since the switch supports 2V. Also the switching loss calculations should use a voltage of 98V, rather than 100V and a load current of 9.8A rather that 10A. The percentage error is small, and becomes increasingly insignificant at higher voltages.

♣

Example 6.2: ***Transistor switching loss for non-linear electrical transitions***

Assume the transistor collector current at turn-off falls according to

$$i_c = \tfrac{1}{2} I_m (1 + \cos \pi t / T_0) \quad \text{for} \quad 0 \le t \le T_0 \qquad (6.7)$$

For a resistive load, R_L

i. Calculate transistor loss at turn-off.
ii. Show that the switching trajectory across the SOA is as for the linear current fall case, as given by equation (6.2) and shown in figure 6.3.
iii. Calculate the peak power dissipation and the time when it occurs.

Solution

i. The collector voltage for a resistive load, on a dc supply V_s, is given by

$$v_{ce}(t) = V_s - i_c(t) R_L$$
$$= V_s - \tfrac{1}{2} I_m (1 + \cos \pi t / T_0) R_L$$

and since $V_s = I_m R_L$

$$v_{ce}(t) = \tfrac{1}{2} V_s (1 - \cos \pi t / T_0)$$

The turn-off energy loss is given by

$$W_{off} = \int_0^{T_0} p(t)\, dt = \int_0^{T_0} i_c(t) v_{ce}(t)\, dt$$
$$= \int_0^{T_0} \tfrac{1}{2} I_m (1 + \cos \pi t / T_0) \times \tfrac{1}{2} V_s (1 - \cos \pi t / T_0)\, dt$$
$$= \tfrac{1}{8} V_s I_m T_0$$

ii. Combining $v_{ce}(t)$ and $i_c(t)$ so as to eliminate the time variable, yields

$$i_c = \frac{V_s}{R_L} \left(1 - \frac{v_{ce}}{V_s}\right)$$

which is the same straight line expression as in equation (6.2) and shown in figure 6.3, for the linear switching transition case.

iii. Instantaneous power dissipation is given by

$$P = v_{ce} i_c = v_{ce} \frac{V_s}{R_L} \left(1 - \frac{v_{ce}}{V_s}\right)$$

Peak power $\hat{P}$ occurs when $dP/dv_{ce} = 0$, that is, when $v_{ce} = \tfrac{1}{2} V_s$, whence on substitution into the power expression P, yields

$$\hat{P} = \tfrac{1}{4} V_s^2 / R_L = \tfrac{1}{4} V_s I_m \quad \text{at} \quad t = \tfrac{1}{2} T_0$$

♣

Turn-on loss can be similarly analysed to yield virtually identical expressions, as is required in problem 6.4.

6.1.2 The inductive load

The voltage spikes generated by inductive loads at turn-off may have high energy content, and the power generated may cause excessive device temperature, voltage stressing, and device failure.

At turn-off, the switch decreases the inductive load current from I_m to zero at a high *di/dt* and the resultant inductive voltage spike is given by

$$v(t) = L \frac{di}{dt} \qquad \text{(V)}$$

where L is the load inductance. The spike energy to be absorbed by the switch is given by

$$W = \tfrac{1}{2} L I_m^2 \qquad \text{(J)}$$

Both the voltage spike and its associated energy may be well outside the capabilities of the switching device. The peak voltage induced must be limited to a value below the breakdown rating of the device. Four commonly employed voltage limiting techniques are shown in figure 6.5.

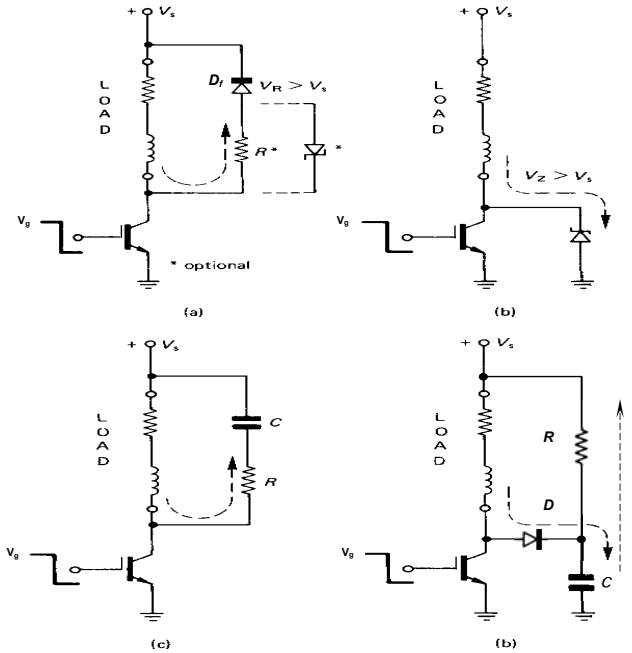


Figure 6.5. *Four methods of limiting inductive load turn-off voltage spike and of absorbing the associated energy: (a) freewheel clamping diode; (b) Zener diode clamp; (c) R-C snubber circuit; and (d) capacitor soft voltage clamp.*

The freewheel diode D_f in figure 6.5a is used to clamp the maximum device voltage to the supply rail voltage. The stored load energy is dissipated after turn-off as a result of the current that flows in the diode and load. The low impedance of the diode causes the current to decay slowly, since the inductor stored energy can only dissipate slowly in the freewheeling loop resistive components. A shorter current decay time can be achieved if series loop resistance R is added, as shown in figure 6.5a. Now the peak off-state voltage experienced by the switch is increased from V_s in the case of only the diode, to $V_s + I_m R$ because of the initial voltage drop across the optionally added resistor. This extra voltage drop, $I_m R$, decreases exponentially to zero. The resistor in figure 6.5a can be replaced by a Zener diode, thereby clamping the switch voltage at turn-off to $V_s + V_Z$. The load now freewheels at a fixed voltage V_Z thereby improving the rate of current decay, which is now constant. The inductive load current will fall linearly from I_m to zero in a time given by

$$t = L I_m / V_z \qquad \text{(s)}$$

An alternative Zener diode clamping circuit, as shown in figure 6.5b, can be employed in low power applications. The Zener breakdown voltage V_z is selected between the rail voltage V_s, and the switch breakdown voltage $(V_s < V_z < V_{BR})$. At turn-off, the Zener diode clamps the switch voltage to a safe level V_Z and absorbs the stored inductive load energy. The higher the clamping voltage level, the faster the energy is dissipated. The inductive load current decays linearly to zero in a time given by

$$t = L I_m / (V_z - V_s) \qquad \text{(s)} \qquad (6.8)$$

The two different Zener diode approaches perform the same switch clamping function in the same current decay time, if the voltage experienced by the switch is the same, but with different Zener diode losses. The desirable feature in the case of the Zener diode in parallel to the switch as in figure 6.5b, is that the protection component is directly across the element to be voltage protected. When placed in parallel with the load as in figure 6.5a, the switch is indirectly voltage protected, relying on the supply decoupling being a low inductance path. A reverse blocking diode D_f in figure 6.5a is mandatory.

The parallel-switch Zener diode approach in figure 6.5b has a number of disadvantages

- The Zener diode voltage rating must be in excess of the supply rail, V_s, while any Zener value can be used when the Zener diode is in parallel with the load.
- At higher voltages, >280V, Zener diodes will have to be series connected, thus the low inductance advantage of clamping with just one component is diminished.
- Assuming no resistance in the load, the energy dissipated with the two Zener diode approaches differs. When in parallel with the load, the load energy $½LI_m^2$ is dissipated while in the second case, load and supply energy are dissipated in the clamping Zener diode. The extra supply energy, in addition to $½LI_m^2$, dissipated in the Zener diode, is $½LI_m^2 V_s/(V_z - V_s)$. This is derived by recognising that, assuming a purely inductive load, the dc supply V_s delivers a current I_m which linearly falls to zero over the period given by equation (6.8).

The *R-C* snubbing circuit shown in figure 6.5c is commonly used in power conversion circuits to limit spikes caused by transformer leakage inductance, diode recovery, and interconnection wire inductance. The stored load energy is resonated to the snubber capacitor at switch turn-off. The reset resistor *R* (non-inductive) must overdamp the *L-C-R* oscillation by absorbing the transferred energy. The resistor also limits the snubber capacitor discharging current to a maximum of V_s/R at switch turn-on. For a purely inductive load, the snubber resistor power losses are given by the sum of the turn-off and turn-on losses, that is

$$P = \left(½LI_m^2 + ½CV_s^2\right) f_s \qquad \text{(W)}$$

Figure 6.5d shows a capacitive voltage clamp used to soft clamp the switch voltage overshoot caused by the inductive energy stored in the load. The capacitor retains a charge of at least V_s. At switch turn-off, when the switch collector voltage reaches the capacitor (supply V_s) voltage level, the inductive stored load energy is transferred to the capacitor and concurrently, the capacitor discharges the energy in excess of V_s into the supply. When the capacitor is over charging, energy is taken from both the load inductance and the supply. When the capacitor discharges through the resistor back into the supply, the earlier energy taken from the supply is returned. The net effect is that only the energy $½LI_m^2$ is dissipated in the resistor. A reset resistor of low inductance is not necessary – a wirewound resistor can be used. This capacitive soft voltage clamp is analysed in detail in chapter 8.2.

Example 6.3: ***Zener diode, switch voltage clamping***

A reed relay coil of 1 mH inductance is switched at 20 kHz with a 20 per cent on-time duty cycle, across a 100 V dc rail. The energy stored in the coil at turn-off is dissipated in a 25 V Zener diode connected as shown in figure 6.5a.

i. Sketch the coil current and voltage, and the switch voltage waveforms.
ii. What is the average coil voltage?
iii. What Zener diode voltage is required for the circuit in figure 6.5b so as to produce the same coil current waveform as in figure 6.5a when using a 25 V Zener diode?
iv. For each circuit, calculate the power requirement of the Zener diode and the average power delivered from the 100 V supply.
v. Calculate the minimum resistance that replaces the Zener diode in figure 6.5a if the coil is to be switched on with almost zero current. Draw the coil current and switch voltage waveform, showing the switch peak voltage at turn-off.
vi. Discuss the relative features of each voltage clamping approach.

Solution

The three voltage clamping circuits being considered are shown in figure 6.6a.

i. With a 20kHz switching frequency, the coil current rises and falls every 50µs, with an on-state duty cycle representing 10µs for the current to increase in the coil and 40µs for the current reset decay to reach zero.
From *V=Ldi/dt*, in steady-state, with zero coil resistance and zero initial current, the peak coil current is $I = V_s t/L$ = 100V×10µs/1mH = 1A. Thus the coil current rises linearly from zero to 1A in 10 µs. During reset, the coil current waveform depends on the reset circuit. For Zener diode (constant voltage) reset, the current falls linearly, while with a resistor the reset current decays with an L/R exponential time constant, as shown in figure 6.6b, for each case.

The various circuit voltage and current waveforms are shown in figure 6.6b, where data derived from the rest of this example has been incorporated.

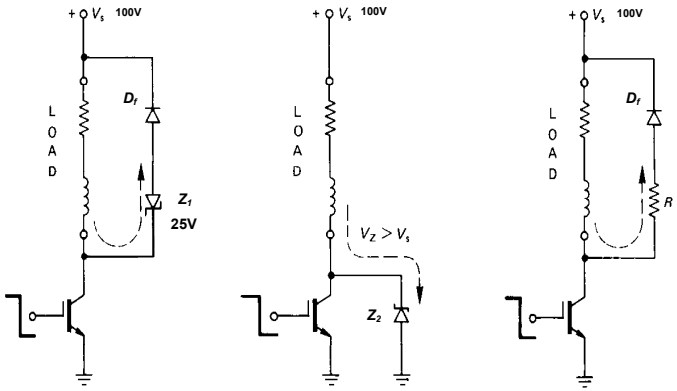


Figure 6.6a. *Three inductive load clamping circuits.*

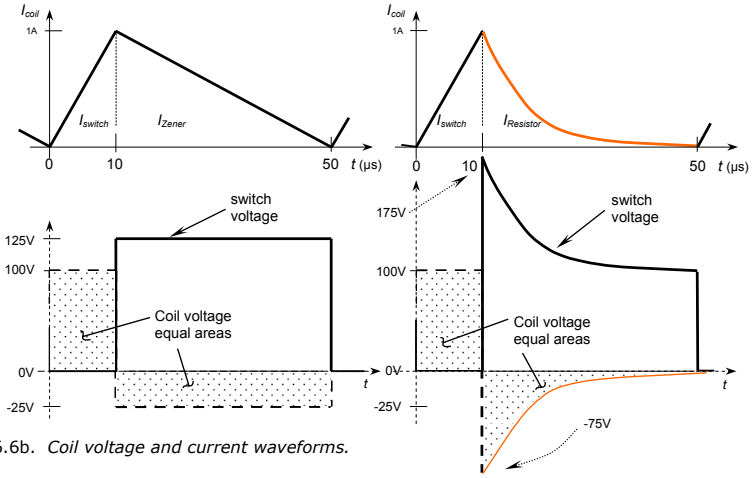


Figure 6.6b. *Coil voltage and current waveforms.*

ii. From *V=Ldi/dt*, for a steady-state continuous waveform, $\int V_L(t)dt = 0$, thus $1/T\int v(t)dt = V_{ave} = 0$, as shown on the coil voltage waveform (the coil voltage areas cancel to zero).

iii. The parallel Zener diode requirement is $V_{Z2} = V_s + V_{Z1}$ = 100V+25V = 125V.

iv. **Zener diode V_{Z1}** in the parallel-load reset circuit:
The energy $½LI^2$ is transferred from the coil to the Zener diode when the switch is turned off. The power dissipated in the Zener diode at 20kHz is therefore $½LI^2 f_s = ½\times 1\text{mH}\times 1\text{A}^2\times 20\text{kHz} = 10\text{ W}$. The total power drawn from the supply is the power stored by the coil at the end of the 10µs on-time, namely 10W.
Zener diode V_{Z2} in the parallel-switch reset circuit:
When the coil releases its stored energy (10W) into the Zener, current is also drawn from the supply. The total average power delivered by the supply over the 50µs period is given by $V_s I_{ave} = ½\times 100\text{V}\times 1\text{A} = 50\text{W}$. This comprises $½LI^2$ (10W) from the supply into the coil when the switch is on for 10 µs, and the remainder (40W) into the Zener diode (plus the coil energy, 10W), when the switch is off for 40 µs. The Zener diode losses are 50W during the switch off period.

v. When a resistor is used in the reset circuit, the current decays exponentially from 1A to 0A. The resistance determines the peak switch voltage. The resistance does not affect the amount of energy dissipated, only the period over which the coil energy is released, dissipated as heat. Assume the coil current to be near zero after three *L/R* time constants, that is $3L/R$ = 40µs = t_{off}.

For L = 1mH, this gives R = 75Ω, with a power dissipation rating of 10W from part iv. At switch turn-off the collector voltage rises to (100V+1A×75Ω) 175V and then decays to 100V. Use an 82Ω (preferred value, exceeding 75Ω which reduces the time constant), 15W metal oxide resistor for low inductance.

vi. A Zener diode approach gives a fixed over-voltage on the switch, independent of current or stored energy. When clamping is in parallel with the switch, only one clamping element is needed, but its power requirement is significantly higher than when the clamp (Zener plus diode) is in parallel to the load. Any resistive element must have low inductance. This is restrictive given the power levels involved, and may result in only less effective wire wound elements being viable.

♣

By far the most common technique used to limit inductive switch-off voltage spikes in power circuits involves the use of a freewheel diode without R_{opt}, as shown in figure 6.5a and 6.7a. Typical switching waveforms for an inductive load clamped by a freewheel diode are shown in figure 6.7.

- **At turn-off**, the switching device conducts the full load current as the collector voltage rises to the supply rail. When the collector voltage reaches the supply rail level the freewheel diode becomes forward-biased and begins to conduct. Only then can the switch current fall to zero. The freewheel diode conducts the load current.
- **At switch turn-on**, assuming the diode is still freewheeling load current, the switch current increases, displacing freewheeling diode current, while the load is clamped to the rail voltage by the conducting freewheel diode. Only when the switch conducts the full load current can the freewheel diode recovered (and block), so that the switch voltage can fall to the low on-state level.

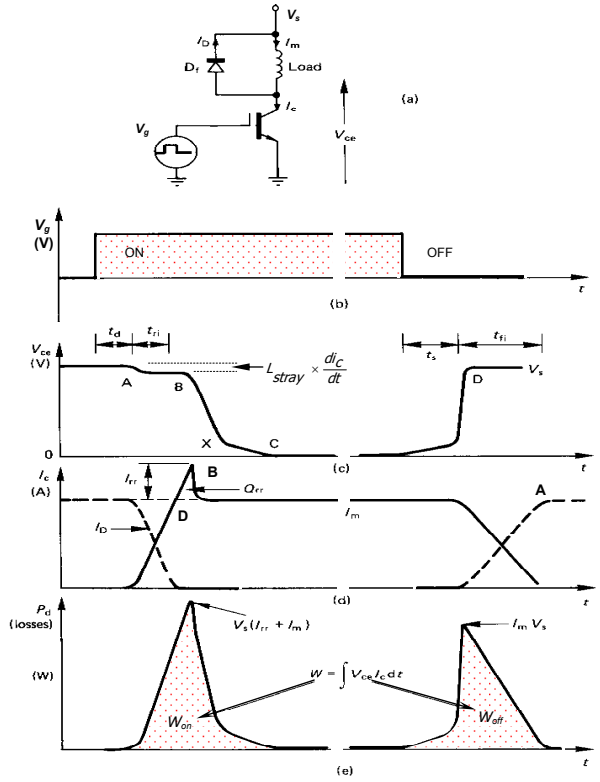


Figure 6.7. *Inductive load switching waveforms:*
(a) the circuit including the freewheel diode D_f; (b) on-off gate drive voltage; (c) collector-to-emitter voltage; (d) collector and freewheel diode current; and (e) switch instantaneous power losses.

It will be seen in figure 6.7 that during both turn-on and turn-off the switch must support instantaneously a maximum voltage, V_s, and full load current, I_m, condition. These severe electrical conditions are shown on the SOA characteristics in figure 6.8. In switching on from the operating point A to C, a maximum voltage and current condition (V_s, I_m) occurs at point D. Because of freewheel diode current reverse recovery effects at turn-on, an SOA trajectory point B is reached. At turn-off, due to stray inductance, voltage over shoot occurs and the point E is reached. By comparison with figure 6.2, it is seen that power losses during the switching intervals are higher for an inductive load than a resistive load.

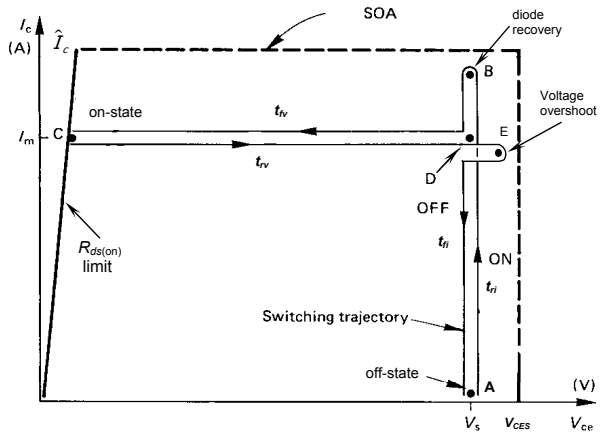


Figure 6.8. *I-V characteristics for an IGBT showing its safe operating area and switching trajectory for an inductive load (linear axes).*

Switching losses can be calculated by using linear approximations to the switching transitions. It can be assumed that a silicon carbide Schottky freewheel diode is employed so as to allow reverse recovery effects to be neglected. Figure 6.9 shows the linearised switching waveforms for an inductive load, where maximum voltage V_s and current I_m occur simultaneously during both turn-on and turn-off. The equations for the collector voltage and current at turn-on and turn-off are also shown in figure 6.9.

The turn-on switching interval loss is given by the time integral over the current rise period plus the voltage fall period,

$$W_{on} = \int_0^{t_{ri}} V_s I_m \frac{t}{t_{ri}}\, dt + \int_0^{t_{fv}} V_s \left(1-\frac{t}{t_{fv}}\right) I_m\, dt \qquad (6.9)$$

$$= \tfrac{1}{2} V_s I_m t_{on} \quad \text{(J)}$$

where $t_{on} = t_{ri} + t_{fv}$, as shown in figure 6.9. The current rise time at turn-on is termed t_{ri}, while the switch voltage fall time at turn-on is termed t_{fv}.

Similarly, from figure 6.9c, the turn-off loss is given by

$$W_{off} = \int_0^{t_{rv}} V_s \frac{t}{t_{ri}}\, I_m\, dt + \int_0^{t_{fi}} V_s\, I_m \left(1-\frac{t}{t_{fv}}\right) dt \qquad (6.10)$$

$$= \tfrac{1}{2} V_s I_m t_{off} \quad \text{(J)}$$

where $t_{off} = t_{rv} + t_{fi}$, as shown in figure 6.9c. The switch voltage rise time at turn-off is termed t_{rv}, while the switch current fall time is termed t_{fi}.

Comparison of switching losses for a resistive load, equations (6.3) and (6.4), and an inductive load, equations (6.9) and (6.10), shows that inductive switching losses are three times those for the resistive load case. The peak power experienced by the switch during switching of an inductive load, $V_s I_m$, is four times greater than that experienced with a resistive load, $\frac{1}{4}V_s I_m$. As for the resistive load switching circuit, actual switch losses with an inductive load are higher than those predicted by equations (6.9) and (6.10). The effects of current tailing, voltage over-shoot, and freewheel diode reverse recovery can together produce losses of the same order as those predicted for theoretical switching by equations (6.3), (6.4), (6.9), and (6.10).

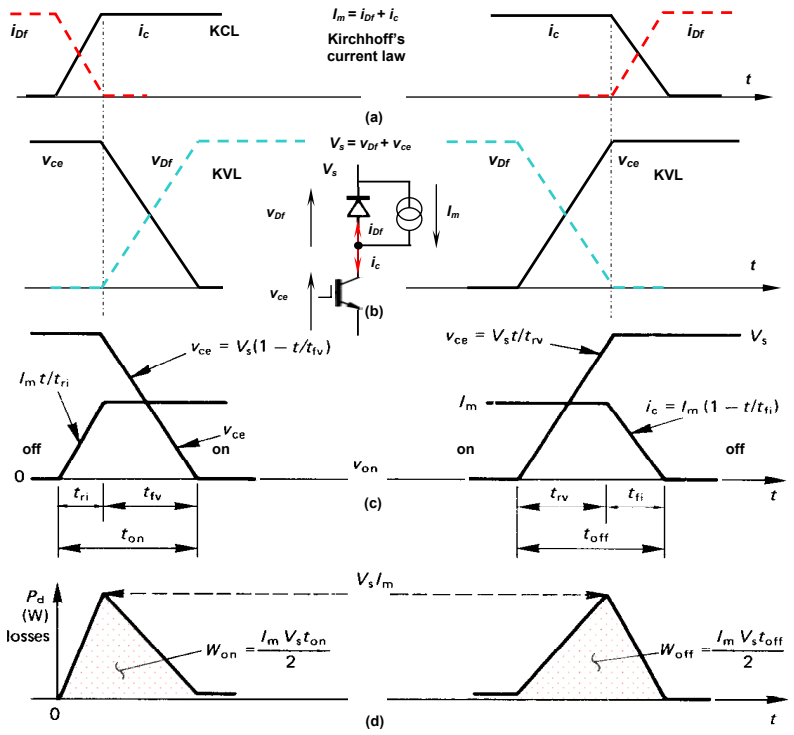


Figure 6.9. *Linear approximations of transistor switching intervals for an inductive load: (a) Kirchhoff's current law $I_m = i_{Df} + i_c$; (b) Kirchhoff's voltage law $V_s = v_{Df} + v_{ce}$; (c) collector voltage and current waveforms with switching parameters defined; and (d) corresponding switching losses.*

Example 6.4: ***Inductive load switching losses***

A power n-channel MOSFET switches a 10A, 100V dc, highly inductive load at 10kHz.
Calculate the worse case switch losses if the switch turn-on time is t_{on} = 1μs, switch turn-off time is t_{off} = 2μs, and the MOSFET channel on-state resistance is 0.2Ω at 10A.
Calculate the maximum instantaneous power dissipation in the switch, and determine when it occurs.

Solution

Maximum switch losses occur when the duty cycle approaches one ($\delta \rightarrow 1$) such the both turn-on and turn-off still occur.
The total switch losses P_T are made up of three components

P_T = on-state loss + loss at switch-on + loss at switch-off

$$P_T = \delta \times I_L^2 \times R_{ds(on)} + \tfrac{1}{2} V_s I_L t_{on} f_s + \tfrac{1}{2} V_s I_L t_{off} f_s$$

$$= 1 \times 10^2 \times 0.2\Omega + \tfrac{1}{2} \times 100V \times 10A \times 1\mu s \times 10kHz + \tfrac{1}{2} \times 100V \times 10A \times 2\mu s \times 10kHz$$

$$= 20W + 5W + 10W$$

$$= 25W$$

Since the off-state leakage current and gate power losses are not specified, it is assumed these are insignificant. The switching loss calculations should use a voltage of 98V, rather than 100V, since (10A×0.2Ω) 2V is dropped across the channel resistance of the MOSFET. The percentage error is small, and becomes insignificant at higher voltages.

Maximum switch loss occurs when during the switching transitions, the drain current is 10A and the drain voltage is 100V. The maximum instantaneous loss is 10A×100V=1000W, ($I_L \times V_s$).

♣

6.1.3 Diode reverse recovery with an inductive load

When a bipolar diode conducts the pn junction scl region accumulates charges. When the diode turns off and the current falls to zero, the junction retains charge that must recovery before diode reverse voltage can be supported. Negative diode current flows. This phenomenon was considered in chapter 4.2.2 and is shown in figure 6.10a. The maximum collector current at turn-on is increased above the load current level I_m by the reverse recovery current I_{rr}, to I_m+I_{rr}. The diode begins to support reverse voltage once the peak reverse recovery current is reached. As a consequence the turn-on losses are increased as shown in figure 6.10c.
The circuit current at peak recovery has a discontinuous derivative, and as a consequence, high circuit voltages are induced across circuit stray inductance due to $v = Ldi/dt$. High-frequency voltage ringing occurs as the stored energy in the stray inductance is dissipated and reverse voltages far in excess of V_s are experienced by the recovering diode.

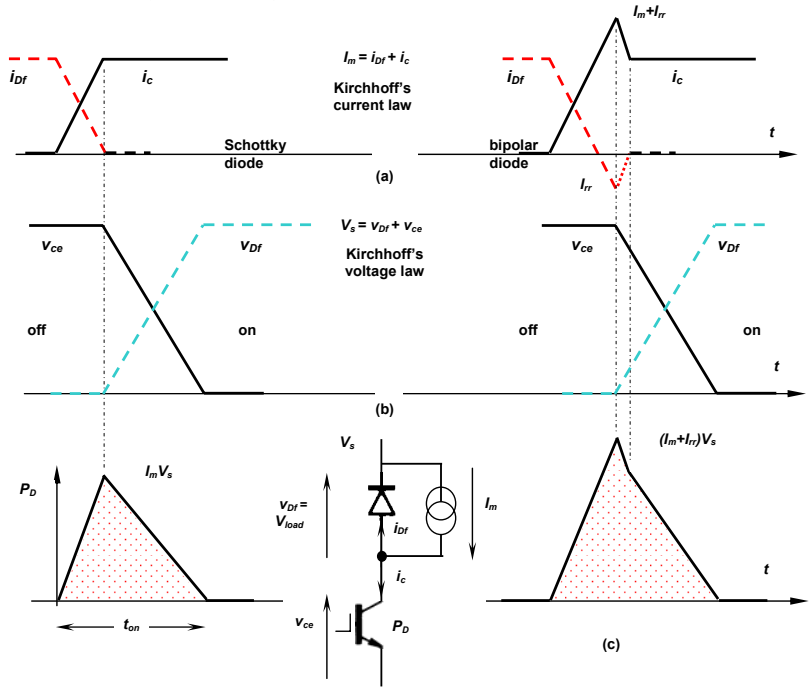


Figure 6.10. *Linear approximations of transistor switching turn-on interval for an inductive load showing freewheel diode reverse recovery effects on the right: (a) Kirchhoff's current law $I_m = i_{Df} + i_c$; (b) Kirchhoff's voltage law $V_s = v_{Df} + v_{ce}$; and (c) corresponding switching losses.*

Example 6.5: ***Inductive load switching losses with device models***

A MOSFET 340V dc chopper feeds an inductive dc motor load at 50kHz. In steady state the load current rises from 10A to 25A when the switch is on with a 75% on-state duty cycle (δ = ¾).
The MOSFET switch turn-on time is t_{on} = 100ns, switch turn-off time is t_{off} = 200ns, and the channel on-state resistance is $R_{ds\,on}$ = 0.025Ω.
The freewheel diode is modelled by a 1V on-state voltage and on-state resistance of 0.05 Ω.
Neglecting diode recovery and diode turn-on losses, calculate

i. the MOSFET total losses
ii. diode losses
iii. power delivered to the motor load, if the armature resistance is 1 Ω and back emf is 170V
iv. electromagnetic energy conversion efficiency and total circuit efficiency

Solution

i. The MOSFET losses comprise turn-on, turn-off, and conduction losses.
The rms current in the MOSFET is given by

$$I_{M-rms} = \sqrt{\frac{\delta}{3}\left(\hat{I}^2 + \hat{I}\times\check{I} + \check{I}^2\right)}$$
$$= \sqrt{\frac{0.75}{3}\times\left(25\text{A}^2 + 25\text{A}\times 10\text{A} + 10\text{A}^2\right)} = 15.6\text{A}$$

The MOSFET conduction losses are therefore

$$P_c = I_{rms}^2 R_{ds\,on} = 324.75\times 0.025\Omega = 8.1\text{W}$$

The switching losses are

at turn-on $P_{t\text{-}on} = \tfrac{1}{2} V_s \check{I}\, t_{on}\, f_s = \tfrac{1}{2}\times 340\text{V}\times 10\text{A}\times 100\text{ns}\times 50\text{kHz} = 8.5\text{W}$

at turn-off $P_{t\text{-}off} = \tfrac{1}{2} V_s \hat{I}\, t_{off}\, f_s = \tfrac{1}{2}\times 340\text{V}\times 25\text{A}\times 200\text{ns}\times 50\text{kHz} = 42.5\text{W}$

Total MOSFET losses are $P_{MOSFET} = P_c + P_{t\text{-}on} + P_{t\text{-}off} = 8.1\text{W}+8.5\text{W}+42.5\text{W} = 59.1\text{W}$

ii. The diode RMS current is

$$I_{D-rms} = \sqrt{\frac{1-\delta}{3}\left(\hat{I}^2 + \hat{I}\times\check{I} + \check{I}^2\right)}$$
$$= \sqrt{\frac{0.25}{3}\times\left(25\text{A}^2 + 25\text{A}\times 10\text{A} + 10\text{A}^2\right)} = 9\text{A}$$

The average diode current is

$$\bar{I}_d = \tfrac{1}{2}(1-\delta)\left(\hat{I} + \check{I}\right) = \tfrac{1}{2}\times(1 - \tfrac{3}{4})\times(25\text{A}+10\text{A}) = 4.375\text{A}$$

The total diode losses are

$$P_{diode} = I_{D-rms}^2 R_{D-on} + \bar{I}\times V_{D-on}$$
$$= 81.25\times 0.05\Omega + 4.375\text{A}\times 1\text{V} = 8.4\text{W}$$

iii. The power delivered to the load comprises losses in the 1Ω armature resistance and the power delivered into the 170V dc back emf.
The rms load current is given by

$$I_{M-rms} = \sqrt{\frac{1}{3}\left(\hat{I}^2 + \hat{I}\times\check{I} + \check{I}^2\right)}$$
$$= \sqrt{\frac{1}{3}\times\left(25\text{A}^2 + 25\text{A}\times 10\text{A} + 10\text{A}^2\right)} = 18\text{A}$$

The load resistor loss is $P_{Ra} = I_{a-rms}^2 R_a = 325\times 1\Omega = 325\text{W}$

The average load current is

$$\bar{I}_a = \tfrac{1}{2}\left(\hat{I} + \check{I}\right) = \tfrac{1}{2}\times(25\text{A}+10\text{A}) = 17.5\text{A}$$

The power delivered to the back emf is $P_{E-a} = \bar{I}_a E_a = 17.5\text{A}\times 170\text{V} = 2975\text{W}$

The total power delivered to the dc motor is $P_{motor} = P_{Ra} + P_{E-a} = 325\text{W}+2975\text{W} = 3300\text{W}$

iv. The dc motor efficiency is

$$\eta_{dc} = \frac{\text{power output}}{\text{power input}} = \frac{2975\text{W}}{3300\text{W}}\times 100 = 90.2\%$$

Including switch and diode losses yields total circuit efficiency, that is

$$\eta_{circuit} = \frac{\text{power output}}{\text{dc supply power input}} = \frac{\text{power output}}{\text{chopper circuit losses + dc motor power input}}$$
$$= \frac{2975\text{W}}{(59.1\text{W}+8.4\text{W})+3300\text{W}}\times 100 = 88.3\%$$

♣

6.2 Switch characteristics

Having considered the switching of inductive and resistive loads, the following are the electrical and thermal characteristics desirable of commutable switching devices (as well as low cost):

off-state *(ideally open circuit):*
- Low, temperature independent leakage current in the off-state, to minimise off-state power loss, and to simplify resistive networks for device series connection.
- High forward and reverse voltage blocking ratings to reduce the need for series device connection, which would otherwise complication control and protection circuitry requirements. Series connection increases the on-state voltage, hence on-state loss. When a diode is used in antiparallel across the switch to allow reverse principal current flow, the switch does not require a significant reverse voltage blocking rating.
- High static off-state avalanche capability to absorb transient overvoltage stresses.
- High static and re-applied *dv/dt* capability to withstand high applied off-state voltages without avalanche or false turn-on, with minimal displacement current.

on-state *(ideally short circuit):*
- Low on-state conducting voltage or low on-state resistance, in order to minimise on-state conduction power loss: with a slight positive temperature co-efficient at high current densities, to allow reliable parallel device connection.
- High on-state current density capability so as to avoid need for and problems associated with parallel device current sharing and differential thermal coefficients.
- Safe controlled switch off from a short circuit current condition.

Switching *(ideally instantaneous):*
- Low control power to produce switching between states, with no 'Miller' interaction.
- Short, temperature independent, turn-on and turn-off times to result in low switching losses which will allow high frequency switching.
- High initial *di/dt* capability at turn-on to allow rapid low loss build-up of the turn-on principal current.
- High surge current capability to withstand transient over current fault conditions, resulting in better fault tolerance and nuisance tripping ride through.
- Large switching safe operating area, being able to simultaneously, but briefly, support rated voltage and rated current, without the need for switch snubber circuits.

Thermal/mechanical:
- Easy to electrically connect and mechanically mount, with low thermal resistance and impedance for efficient heat removal.
- Mechanically, electrically, and thermally robust, with the ability to operate at high (and low) junction temperatures in high (and low) ambient, pressure, humidity conditions.
- Matching substrate structure and thermal properties to minimise stressing due to thermal, mechanical, and power stressing.

6.3 Switching classification

There are four principal *I-V* switching conditions during the commutation (turn-on or turn-off) of a switch, viz.:

- Hard switching;
- Soft switching;
- Resonant switching; and
- Naturally-commutated switching.

These four possibilities are classified in terms of the *switching time* t_s and the *commutation time* t_q, where $t_q \le t_s$. Figure 6.11 shows the four electrical cases and specifies the switching and commutation times for each.

- *Switching time* t_s is the time for a switch to change from fully on ($v = 0$, $i = I_L$) to fully off ($v = V_s$, $i = 0$), such that no further change occurs in the switch voltage or current due to the change of state.
- *Commutation time* t_q is associated with the external circuitry and is defined as the time the switch takes to reach zero current at turn-off or to reach zero volts at turn-on. Alternatively, commutation time is the period of switch power loss at turn-on or turn-off, due to the switch changing states.

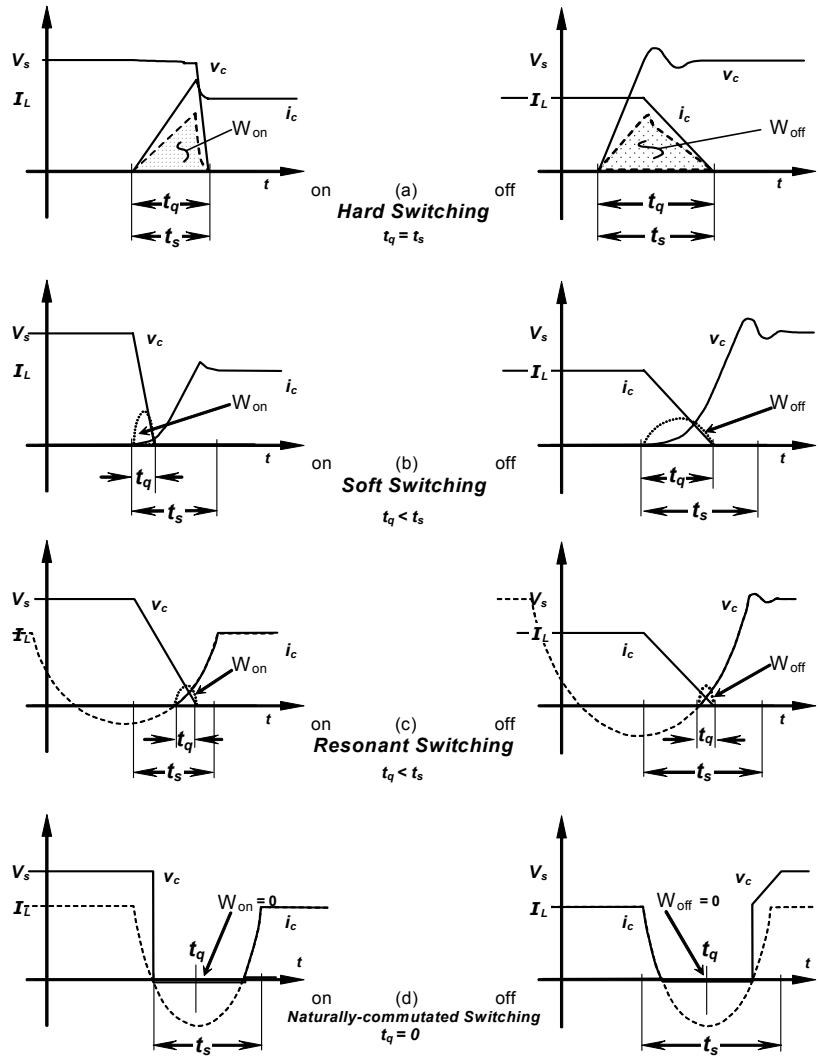


Figure 6.11. *Switch voltage (v_c), current (i_c), and power loss (W_{on} and W_{off}) of four switching classifications: (a) hard switching; (b) soft switching; (c) resonant switching; and (d) naturally-commutated switching.*

Generally, the switch loss magnitude (stress) for a given set of electrical and thermal operating conditions, decreases when progressing from severe hard switching through to virtually lossless naturally-commutated switching.

6.3.1 Hard switching: $t_q = t_s$

The turn-on and turn-off switching waveforms in figure 6.11a for an inductive load show that hard switching is characterised by $t_q = t_s$. The resistive and inductive switching considered in sections 6.1.1 and 6.1.2 are examples of hard switching. In figure 6.4 for a resistive load, the switching periods t_{on} and t_{off} (t_s) correspond to the period of switch losses (t_q) during each state transition. In figure 6.9 for the inductive load, the t_q periods correspond to the power loss periods at switching ($t_{rv} + t_{fi}$ and $t_{fv} + t_{ri}$).

6.3.2 Soft switching: $t_q < t_s$

Figure 6.11b shows typical soft-switching waveforms for turn-on and turn-off. The switching losses are complete before the switch has reached its final steady-state condition. That is, $t_s > t_q$ such that the periods t_s and t_q both commence at the same time.

At turn-on, the switch voltage reaches zero before the switch current reaches the steady-state full-load value I_L. Once the switch voltage reaches zero, the rising current no longer results in a power loss. This I-V characteristic at turn-on (usually involving inductance in series with the switch) is a form of quasi *zero current switching*, ZCS.

The inverse occurs at turn-off. The switch current reaches zero before the switch voltage has settled at the supply voltage level V_s. (Usually involving capacitance in parallel with the switch.) This is a form of quasi *zero voltage switching*, ZVS.

Soft-switching results when auxiliary stress diverting circuits, called snubber circuits, are used, as will be considered in chapters eight and nine.

6.3.3 Resonant switching: $t_q << t_s$

Resonant-switching waveforms at turn-on and turn-off are shown in figure 6.11c, with switching periods t_s shown. Resonant switching occurs if the switching period is associated with either the switch voltage or current being zero, due to external load circuit conditions which have diverted the load current or voltage. That is $t_s > t_q$.

Switching of the voltage when the switch current is zero, usually at turn-on, is called *zero current resonant switching*, ZCRS, while commutating the current while the switch voltage is zero, usually at turn-off, is called *zero voltage resonant switching,* ZVRS. Because the exact instant of zero may vary, being load circuit dependant, some control restriction is inevitable. Zero voltage or current switching can be readily attained with ac mains converter circuits since switching can be synchronised with supply zero voltage crossing, or zero current when the load current reverses due to the supply voltage reversal.

6.3.4 Naturally-commutated switching: $W = 0$, $t_q = 0$

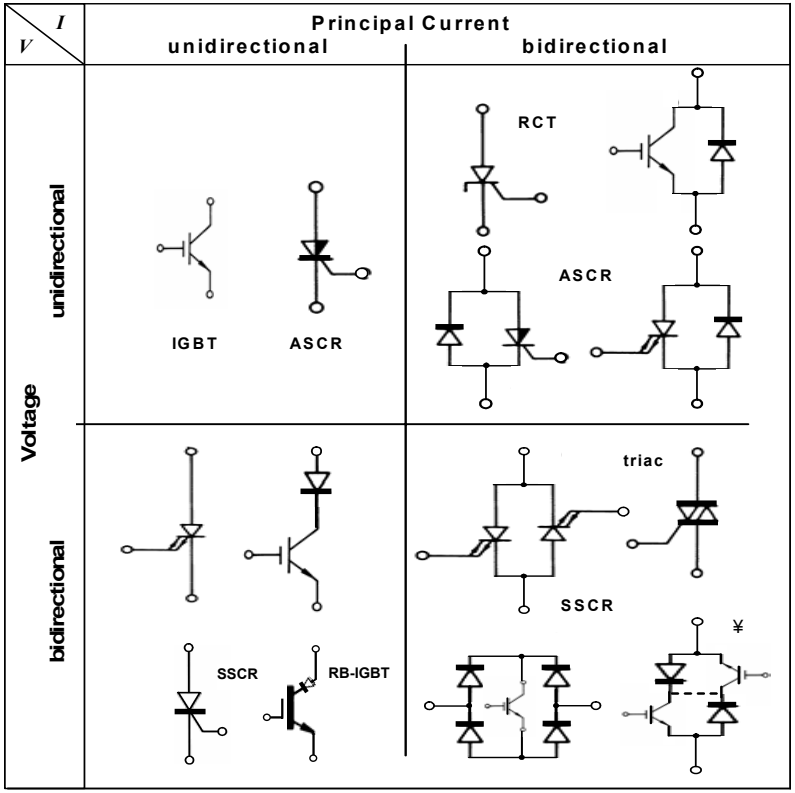
Figure 6.11d shows switching when the voltage and current are both zero, called *naturally-commutated switching*. This was a commonly used technique for force turn-off of thyristors before the exploitation of the GTO thyristor. Current from an auxiliary commutation circuit displaces (exceeds) the device principal current and reverse biases the device, at turn-off. The method was not used at turn-on. Commutated turn-on and turn-off occurs in inverter circuits where the switch has an anti-parallel connected diode. When the diode conducts and the switch is on but not conducting, if the load power factor causes the current to reverse, then the main switch automatically starts conducting with the switch voltage at zero because the diode was previously conducting, clamping the switch voltage slightly negative.

Naturally-commutated switching occurs for ac mains zero crossing switching, with a purely resistive load such that the load V and I are in phase. Switching losses are virtually zero.

6.4 Switch configurations

Most semiconductor switches are unipolar, that is, allow current and/or voltage to be supported in one direction. The MOSFET allows uncontrolled reverse current flow; hence can not support reverse voltage because of its parasitic body diode. Some structures, like the RCT considered in chapter 3.3.3, integrate an anti-parallel diode with a thyristor. Generally, such integrated approaches sacrifice some electrical characteristics. Many applications require a bi-directional current and/or bi-directional supporting voltage switches, so the basic switches can be configured as shown in figure 6.12, to give the necessary I-V characteristics. The net effect of the bi-directional voltage arrangements is good dynamic electrical characteristics but poor static characteristics. Specifically, the switching performance is as for the principal switch but the on-state loss is that of two series connected devices. In the case of the bidirectional blocking thyristor, the on-state voltage is increased slightly because an n-buffer can not be used in its fabrication. The bi-directional conducting thyristor discussed in chapter 3.3.4 attempts to minimise the sacrificed on-state voltage limitation. A reverse blocking IGBT can also be realised. Die edge passivation of the diode region by a through the die p+ diffusion, plus guard rings, increase processing complexity, and hamper voltage ratings. A punch through IGBT version with reverse voltage blocking properties, is therefore problematic. On-state voltages are increased for a given switching speed and, as with the MOSFET body diode, the non-optimal diode recovery characteristics are a compromise because of the overriding n-substrate low resistivity requirements. See chapter 3.2.4.

- Controllable switching devices with reverse blocking capability are usually required for ac to ac converters, half-wave resonant converters, and current fed inverters.
- Voltage source inverters, full-wave resonant converters, and dc to dc converters usually do not require switching devices with reverse blocking properties, but may use an antiparallel connected diode.



¥ Can be arranged so that emitters are at the same potential. Switches may be reverse blocking IGBTs.

Figure 6.12. *Switch configurations for uni-directional and bi-directional I-V characteristics.*

Reading list

Peter, J. M., *The Power Transistor in its Environment,*
Thomson-CSF, Sescosem, 1978.

Problems

6.1. During turn-on and turn-off of a power transistor the current-voltage relationships are as shown in figure 6.13. Calculate the energy loss during both turn-on and turn-off periods and the mean power loss if the transistor is being switched at a frequency of 10 kHz. What is the maximum instantaneous power dissipated?
[1.66 mJ, 16.6 mJ, 183 W, 5kW]

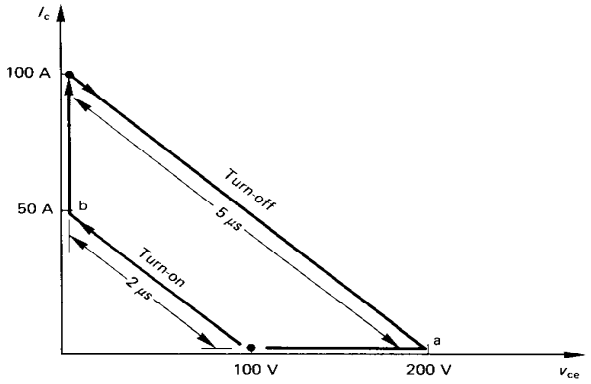


Figure 6.13.

6.2. The equivalent circuit in figure 2.4a involving parameters E_o and R_o can be extended to model a thyristor by replacing the ideal diode by an ideal thyristor. Derive general expressions for the thyristor mean power loss P_d and rms current i_o with a constant load current I_o and switch on-time duty cycle δ.
If E_o = 1 V and R_o = 0.01 Ohms, for I_o = 50 A and a 25 per cent on-time duty cycle, calculate the thyristor:
i. On-state voltage, V_F
ii. Mean power, P_d
iii. rms current, i_o.
[See example 2.1: 1.5 V, 18.75 W, 25 A]

6.3. If the collector voltage at turn-on falls according to $v_c = \frac{1}{2}V_s(1 + \cos \pi t / T_o)$ for $0 \le t \le T_o$
i. For a resistive load, R_L, calculate transistor loss at turn-off.
ii. Show that the switching trajectory across the SOA is as for the linear current fall case.
iii. Calculate the peak power dissipation and when it occurs.

6.4. A transistor is used to switch an inductive load with a current of I_m.
At transistor turn-off, the collector voltage rises to the supply rail V_s according to
$v_{ce} = \frac{1}{2}V_s (1 - \cos \pi t/T_{ov})$ for $t \le T_{ov}$,
then the collector current falls according to
$i_c = \frac{1}{2}I_m(1 + \cos \pi t/T_{oi})$ for $t \le T_{oi}$.
Using the same integration form as in equation (6.10), show that the turn-off loss is
$P = \frac{1}{2}V_sI_m T_o$ where $T_o = T_{ov} + T_{oi}$.

7

Driving Transistors and Thyristors

The thyristor, being a multiple (three) bipolar junction device, is essentially a current-controlled device. As illustrated in figure 7.1a, a current must be supplied between the gate and cathode terminals to produce cathode injection, hence anode current flow, provided the anode is forward biased. The magnitude of gate drive current determines the delay time and the anode current rise time. In gate commutated thyristors, a negative gate current must be produced, the magnitude determining the turn-off delay time and anode current fall time.

The power MOSFET and IGBT are voltage controlled devices with turn-on and turn-off requirements fundamentally different to bipolar devices. With the n-channel enhancement-mode power MOSFET and IGBT, a positive voltage must be applied between the gate and source terminals to enhance a channel which allows a drain current, if the drain is positively biased with respect to the source, as shown in figure 7.1b. Generally the MOSFET and IGBT are easier to drive than the bipolar thyristor, and only a few basic considerations are required for MOSFET and IGBT gate circuit implementation.

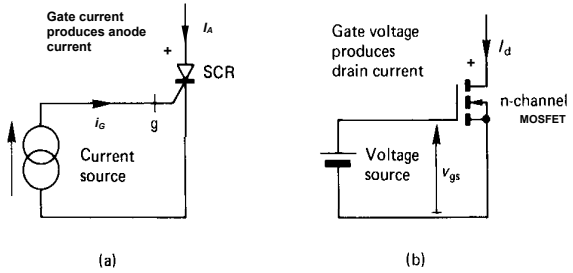


Figure 7.1. *Thyristor and transistor drive requirements: (a) current drive for the bipolar junction thyristor and (b) voltage drive for the MOSFET and IGBT.*

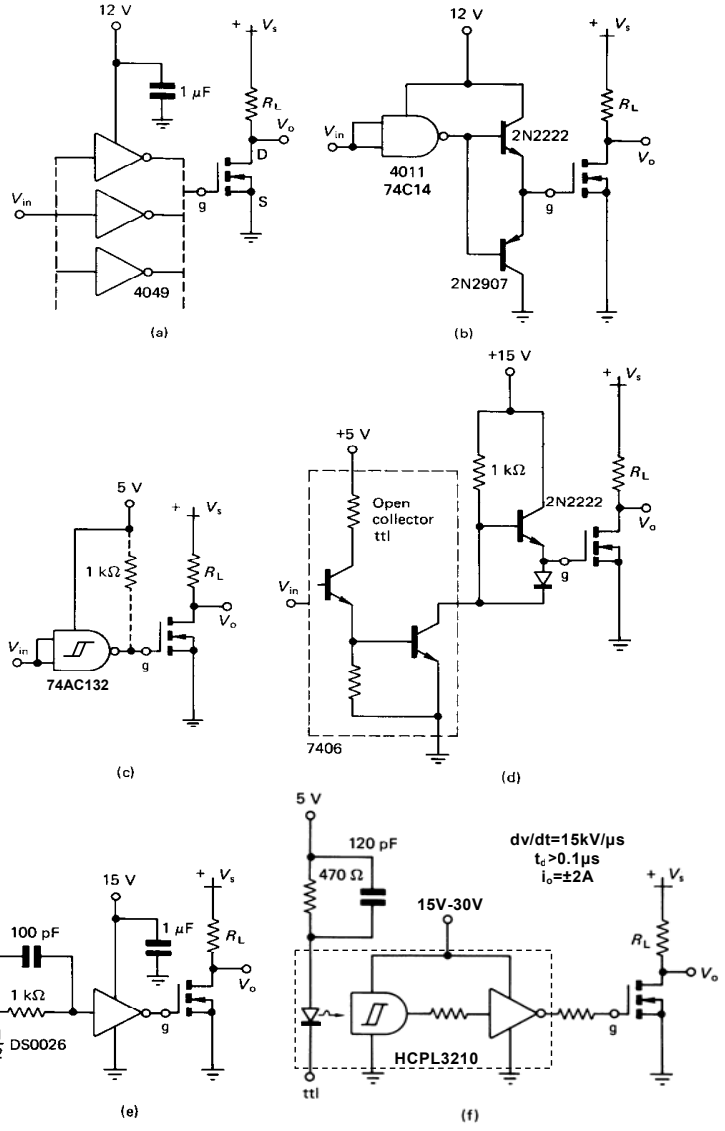
7.1 Application of the power MOSFET and IGBT

The MOSFET gate is isolated electrically from the source by a dielectric layer of silicon dioxide. Theoretically no current flows into the gate when a dc voltage is applied to it. In practice, gate current is required to charge device capacitances and a small leakage current of the order of nano-amps does flow in order to maintain the gate voltage.

When no voltage is applied between the gate and source terminals (but with zero impedance), the drain-to-source impedance is very high and only a small leakage current of less than a milli-amp flows in the drain, until the applied voltage exceeds the drain-to-source avalanche voltage, V_{DSS}.

When a positive gate voltage is applied, an electric field is produced which modulates the drain-to-source resistance. When a gate voltage exceeds the threshold voltage level the channel resistance reduces to a low resistance and drain current flows. The maximum drain current depends on the gate voltage magnitude, assuming that the impedance of the external drain circuit is not current-limiting.

Turn off - reducing the drain current to the leakage current level - is achieved by reducing the gate voltage to below the gate threshold voltage level. The drain switching speeds are essentially determined by that speed at which the gate voltage can reach a level above the threshold voltage (for turn-on) or below the threshold voltage (for turn-off). Although the gate-to-source capacitance is an important parameter, the gate-to-drain capacitance is more significant because of the Miller effect, as considered in section 4.4.2. During switching, the dynamic gate-to-drain capacitance can be effectively much larger than the gate-to-source capacitance. The Miller capacitance typically requires more charge for switching than the gate-to-source capacitance.



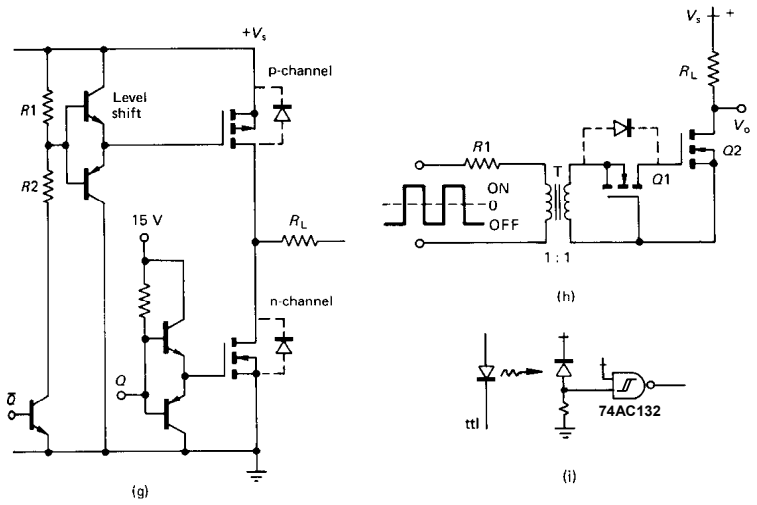


Figure 7.2. *Gate drive circuits for the MOSFET and IGBT: (a) driven from cmos; (b) driven from cmos and an emitter follower; (c) driven from ttl with pull-up resistor which increases sourcing capability; (d) driven from open collector ttl with an external current source; (e) driven from a high-current cmos clock driver; (f) opto-isolated driver circuit; (g) drive circuits for a totem pole connected p and n-channel MOSFET leg; (h) driven from a pulse transformer; and (i) fibre optic translation stage.*

7.1.1 Gate drive circuits

The trench gate n-channel enhancement-mode power MOSFET (or IGBT) with a low threshold voltage interfaces easily with logic level integrated circuits. This allows low-power digital logic circuits to control directly high-power levels. Figure 7.2 shows a series of ttl and cmos circuits driving power MOSFETs, each circuit offering different levels of switching speed and performance.

When driving a MOSFET directly from a cmos gate output, as shown in figure 7.2a, only modest rise and fall times can be expected because of the limited source and sink current available from a cmos gate. Figure 7.3a illustrates the output configuration of a typical cmos output stage, which consists of a series-connected p and n-channel MOSFET with the gates connected together. The cmos totem pole output stage is driven by a common signal, hence the name complementary mos - cmos - and when the input is high the n-channel device is on and the p-channel off, while when the input is low, the p-channel turns on and the n-channel turns off. However, cmos has a limited current output capability as shown in the 4049 source-to-sink output characteristics in figure 7.3b and c. The cmos gate output has to drive as a load the power MOSFET capacitive gate. In this configuration, the turn-on current is supplied from the p-channel fet, which has the poorer characteristics of the cmos pair. The turn-off current is sunk by the n-channel fet. Table 7.1 shows cmos typical current source and sink capabilities, switching speeds, and output impedance. It will be seen that the best performance, by far, is achieved from the 4049 and 4050 buffers.
If shorter delays and faster drain rise and fall times are required there are several ways to obtain them. The simplest is to parallel a number of identical cmos inputs and outputs as shown dotted in figure 7.2a. The additional current capability, with the six parallel connected gates of the 4049, will significantly improve MOSFET switching performance.

In figure 7.2b the gate drive current is the output current of the cmos gate multiplied by the gain β of the bipolar transistors. No bipolar saturation times are incurred since the transistors are operating as emitter followers, which cannot saturate. The operating frequency is no longer restricted by the cmos output current limitations.

MOSFETs can also be driven directly from ttl gates. Table 7.2 shows ttl typical current source and sink capabilities and switching speeds. Low supply voltage, typically 5V, and high internal sourcing impedance characteristics, restrict MOSFET switch-on speed and gate voltage level. The ttl sink capability is significantly higher than source capability, hence a pull-up resistor as shown in figure 7.2c enables the sinking capability to be exploited at turn-on, as well as at turn off. A limitation of using ttl for driving MOSFETs is that the gate voltage is restricted to less than 5V, hence if the drain current is not to be restricted, low gate threshold voltage trench gate MOSFETs and IGBTs are used. An open collector ttl drive technique as shown in figure 7.2d overcomes the gate voltage limitation as well as improving the current source limit.
Very fast switching speeds are attained with the capacitive driver shown in figure 7.2e. Such drivers can both source and sink typically 1.5 A in tens of nanoseconds. An isolated gate-to-source drive version is shown in figure 7.2f, where a floating 15 V rail is used and the gate control signal is optically transmitted with high *dv/dt* capability. The driver incorporates high current output, with modest propagation delays.

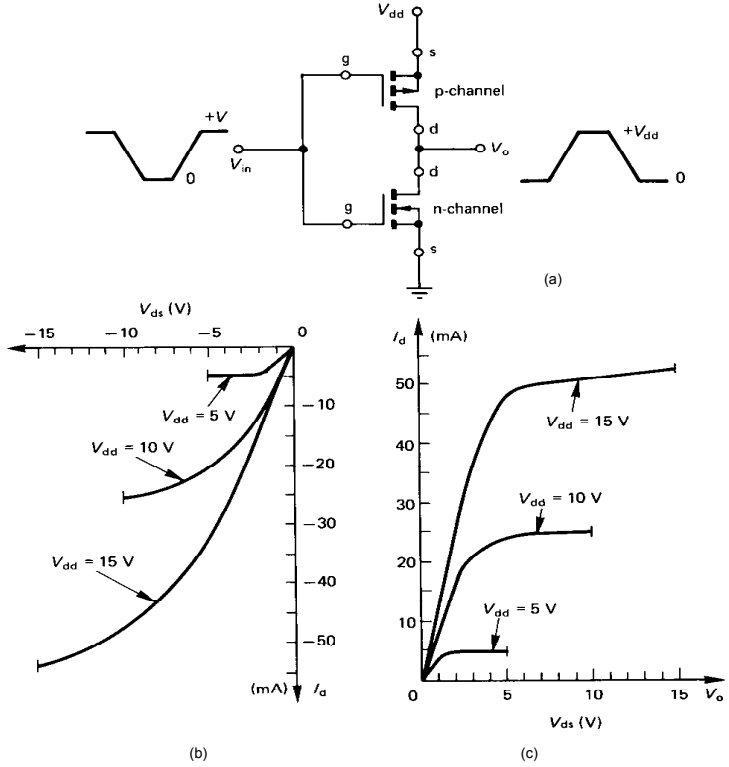


Figure 7.3. *CMOS 4049 inverter output: (a) output cmos totem pole; (b) p-channel drain sourcing; and (c) n-channel drain sinking, both at 25°C.*

Table 7.1 Driving mosfets from cmos (buffered)

Logic conditions		Standard buffered outputs at logic supply voltage (V_{dd}) of			4049/4050 drivers at logic supply voltage (V_{dd}) of		
		5 V	10 V	15 V	5 V	10 V	15 V
Logic zero							
Approximate sink current I_{OL} (mA) for $V_{OL} \leqslant 1.5$ V		1.5	3.5	4.0	20	40	40
Logic one							
Minimum source current I_{OH} (mA) for V_{OH}		−0.51 ⩾4.6 V	−1.3 ⩾9.5 V	−3.4 ⩾13.5 V	−1.25 ⩾2.5 V	−1.25 ⩾9.5 V	−3.75 ⩾13.5 V
Typical switching times (ns) of logic drive signals:							
Rise		100	50	40	100	50	40
Fall		100	50	40	40	20	15
	Typ.	1.7k	500	430			
$R_{ds(on)}$ (ohms)	Max.	12.5k	2.5k	–			
(calculated) Source	Typ.	500	420	190			
Sink	Max.	2k	1k	–			

Table 7.2 Driving mosfets from TTL

Logic conditions	Logic type									
	74	74AC	74L	74LS	74S	74ALS	74HC	DS0026	Open collector 74, 30 V
Logic zero									
Min. sink current I_{OL} for V_{OL} = 0.4 V (mA)	16	24	3.6	8	20	18	4	1.5A	40
Logic one									
Max. source current I_{OH} for V_{OH} = 2.4 V (mA)	−0.4	−24	−0.2	−0.4	−1.0	−0.4	−4	−1.5A	22
Typical gate propagation delay (ns)	10	7	50	12	4	4	10	15	12
V_{iL} max. (V)	0.8	1.35	0.7	0.8	0.8	0.8	0.9	—	0.8
V_{iH} min. (V)	2.0	3.15	2.0	2.0	2.0	2.0	3.15	—	2.0

Drive circuits for p-channel MOSFETs may be complicated by the reference signal voltage level, as shown in the series n and p-channel totem pole in figure 7.2g. This figure illustrates how the p-channel drive may be derived by means of a level shifter. The emitter follower, pnp transistor used for turn-on must have a breakdown voltage rating in excess of the totem pole rail voltage. Above 300 V the pnp transistor can be replaced by a diode as shown in figure 7.2d, or a low current high voltage MOSFET. Restricted charging of the translation MOSFET output capacitance can lead to increased delay times. The resistor divider, *R1-R2,* ensures that the p-channel gate voltage limit is not exceeded. In order to increase gate drive capability *R2* can be decreased provided a 15 V Zener is used across the p-channel MOSFET gate to source. The low-voltage npn transistor in the p-channel driver stage is used for fast turn-off, shorting the p-channel source to its gate.
A simple method of driving an n-channel MOSFET, with its source not referenced to ground, is shown in figure 7.2h. Electrical (galvanic) isolation is achieved by means of a pulse transformer. The internal parasitic diode in Q1 provides the path for the n-channel MOSFET gate to charge. When the pulse transformer saturates, Q1 blocks any discharge of the gate until turn-off, when a negative transformer pulse turns on Q1, thereby discharging the n-channel gate charge.
An alternative translation method using a fibre optic stage is shown in figure 7.2i. The temperature-independent, high threshold characteristics of 74AC technology is used for a simple detector comparator. A Schmitt input (hysteresis) gate (74AC132) improves noise immunity.
In general, translation from ttl levels can be achieved with Zener diode bias circuits.

From the circuits in figure 7.2 it is seen that there are two basic types of gate drives.

- Low-side
- High-side

Essentially a low-side driver is one where the control signal and the power device gate are at almost the same potential. The lower switches in bridge legs normally use low-side drivers, while the upper switches require high-side drivers which translate the control signal and gate power to a different potential. The gate drive circuits 7.2a to 7.2e are basic low-side gate drive circuits. The high-side drivers in figures 7.2f to 7.2i translate the control signal to the gate level.

Although the gate drive circuits in figures 7.2a to 7.2i translate the control signal to the device gate, these circuits do not address two important gate drive issues.

- The derivation of the gate drive supply, particularly for floating gate drives as encountered in inverters.
- The derivation of negative gate bias at turn-off for better immunity to false turn-on due to noise and induced Miller charging effects.

7.1.1i - *Negative gate drive*

The gate drive circuits shown in figure 7.2 only clamp the gate to near zero volts during the off period.
The lower bridge leg switch in figure 7.4 uses ±15V gate voltages. The complementary buffers drive the gate-source of the shown device in an H-bridge configuration. The buffers require an isolated 15V dc supply. Since the 15V dc supply is isolated, the complementary buffers can be used for high side gate drives, provided the control signal is isolated, as in figure 7.2i. Practically a negative gate bias of -5V is sufficient for noise immunity while any voltage in excess of this unnecessarily increases turn-on delay and increases gate power requirements. Manufacturers are continually improving power device properties and characteristics. Gate threshold voltage levels are constantly being decreased, and coupled with the fact that the threshold voltage decreases with temperature, negative voltage gate drive is necessary for high noise immunity to prevent false turn-on with high power devices. Gate capacitance improves noise immunity.

7.1.1ii - *Floating power supplies*

There are three basic methods for deriving floating power supplies for gate drives.

- A low inter-winding capacitance, high-frequency transformer
- A capacitive coupled charge pump
- A diode bootstrap

The upper bridge leg switch T_u in figure 7.4 uses both a diode bootstrap via D_{bs} and a single ended capacitor charge pump via C_{cp}, in order to derive gate power.

1 *- capacitive coupled charge pump*

By switching T_{cp} at high frequency the low-capacitance, high-voltage capacitor C_{cp} is successively charged through D_{cp1} and discharged through D_{cp}. Discharge through D_{cp} involves charging C_{gs}, the gate voltage supply capacitor. The shown charge and discharge paths both rely on either the upper switch T_u or diode D_u being in a conducting state.

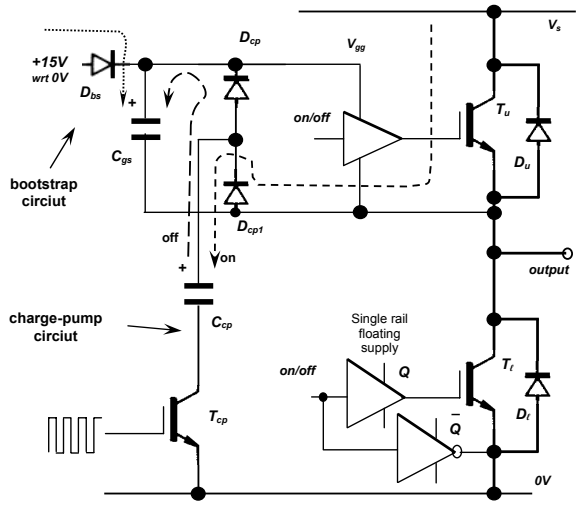


Figure 7.4. *Typical IGBT bridge leg showing ±15V gate drive on the leg lower switch and charge pump plus boot strap gate supply circuits for the leg upper switch.*

2 *- diode bootstrap*

When the lower switch T_l or diode D_l conduct, high voltage diode D_{bs} allows the upper gate supply capacitor C_{gs} to charge from a 15V dc supply which is referenced to the 0V dc rail. When the upper switch or diode conduct, the bootstrap diode is reverse bias and supports V_s + V_{gg}. Start-up is a problem since the gate of the upper switch T_u is in a high impedance state while its supply is being charged after the lower switch is turned on. For this reason, the boot strap is usually used in conjunction with a capacitor charge pump.

The only foolproof method to ensure gate power at all times, particularly at start-up and during prolong on-state periods, is to use a high-frequency (power and/or signal) transformer approach.

7.1.2 Gate drive design procedure

The effective gate to source capacitance, C_{in}, can be calculated from

$$C_{in} \triangleq \delta Q_g / \delta V_{gs} \tag{7.1}$$

The initial slope of the charge in figure 7.5a, 740 pF, is due to the gate source capacitance charging below the gate threshold level. The next charge section between Q_{g1} and Q_{g2} in figure 7.5c is due to the Miller effect. The horizontal charge portion is due to the very high drain-source depletion field capacitance as the drain falls below the gate voltage level.

The drain switching times, similar to those derived in 4.4.2, can be calculated from the charge transfer characteristics in figure 7.5, using the following equations.

(i) From figure 7.5c, for turn-on

$$t_{d\,on} = \frac{Q_{g1}}{V_{g1}} R_g \ \ell n \left(\frac{V_{gg}}{V_{gg} - V_{g1}}\right) \quad \text{(s)} \tag{7.2}$$

$$t_r = \frac{Q_{g2} - Q_{g1}}{V_{g2} - V_{g1}} R_g \ \ell n \left(\frac{V_{gg} - V_{g1}}{V_{gg} - V_{g2}}\right) \quad \text{(s)} \tag{7.3}$$

(ii) From figure 7.5d, for turn-off

$$t_{d\,off} = \frac{Q_{g2} - Q_{g2}}{V_{g2} - V_{g2}} R_g \ \ell n \frac{V_{gg}}{V_{g2}} \quad \text{(s)} \tag{7.4}$$

$$t_f = \frac{Q_{g2} - Q_{g1}}{V_{g2} - V_{g1}} R_g \ \ell n \frac{V_{g2}}{V_{g1}} \quad \text{(s)} \tag{7.5}$$

where R_g is the gate equivalent series resistance and $V_{g1} = V_{TH}$.

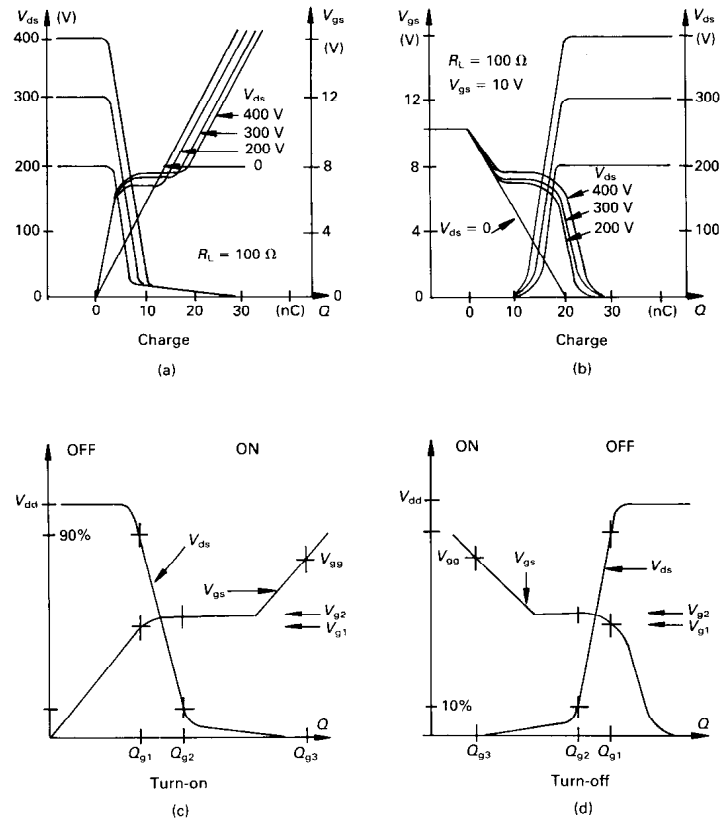


Figure 7.5. *Typical MOSFET charge transfer characteristics at: (a) turn-on; (b) turn-off; (c) turn-on showing switching parameters; and (d) turn-off showing switching parameters.*

The energy required for switching is given by

$$W = \tfrac{1}{2} Q_{g3} V_{gg} \quad \text{(J)} \tag{7.6}$$

which will be dependent on the drain current and voltage. The gate drive power requirements are given by

$$P = Q_{gs} V_{g3} f_s \quad \text{(W)} \tag{7.7}$$

Obviously the faster the switching speed requirement, the higher and faster the gate drive current delivery necessary.

If only 15 mA is available for gate drive then, based on figure 7.5, switching occurs in about 1 µs (from $Q = I \times t$). This level of performance could be expected with circuit 7.2a, and slower switching for the circuit in figure 7.2c. By employing the gate drive in figure 7.2c, the gate voltage is limited to 5 V, hence the MOSFET represented by figure 7.5 could not be switched.

The circuits in figures 7.2b and 7.2d are capable of delivering about 100 mA, which yields switching speeds of the order of 150 ns, with only 50 mW of drive power dissipation at 100 kHz. The drive circuit in figure 7.2e is capable of delivering ± 1.5 A. Hence the device characterised by figure 7.5 can be switched in only 10 ns.

Switching times deteriorate slightly if reverse gate-to-source biasing is used for higher noise immunity in the off-state. Analysis of the increase in turn-on delay as a result of the use of negative gate drive is presented in Appendix 4.8.

Example 7.1: ***MOSFET input capacitance and switching times***

A MOSFET switching a resistive load has the following circuit parameters:

$$R_g = 47\Omega, \qquad R_L = 100\Omega$$
$$V_{gg} = 10\text{ V}, \qquad V_{ds} = 400\text{ V}$$

Based on the charge transfer characteristics in figure 7.5, calculate the gate input capacitance and switching times for MOSFET turn-on and turn-off.

Solution

The charge transfer characteristics shown in figure 7.5 are valid for a 100 Ω resistive load and a 0-10 V gate voltage. A 400 V drain switching characteristic is shown.

At turn-on, from figure 7.5a and using equations (7.2) and (7.3)

(i) $C_{in} = C_{gs} = Q_{g1} / V_{g1} = 4.4\text{ nC} / 6\text{V} = 740\text{ pF}$
$t_{d\,on} = 740\text{ pF} \times 47\Omega\ \ell n\ (10\text{V}/10\text{V}-6\text{V}) = 31.9\text{ ns}$

(ii) $C_{in} = (Q_{g2} - Q_{g1}) / (V_{g2} - V_{g1}) = 5.6\text{ nC} / 1.5\text{V} = 3.7\text{ nF}$
$t_r = 3.7\text{ nF} \times 47\Omega\ \ell n\ 5.6\text{V}/2.5\text{V} = 141.3\text{ ns}$

At turn-off, from figure 7.5b and using equations (7.4) and (7.5)

(i) $C_{in} = (Q_{g3} - Q_{g2}) / (V_{gg} - V_{g2}) = 7.5\text{ nC} / 2.5\text{V} = 3\text{ nF}$
$t_{d\,off} = 3\text{ nF} \times 47\Omega\ \ell n\ 10\text{V}/7.5\text{V} = 40\text{ ns}$

(ii) $C_{in} = (Q_{g2} - Q_{g1}) / (V_{g2} - V_{g1}) = 7.5\text{ nC} / 0.9\text{V} = 8.3\text{ nF}$
$t_f = 8.33\text{ nF} \times 47\Omega\ \ell n\ 7.5\text{V}/6.6\text{V} = 50\text{ ns}$

An underestimate of the fall time results if figure 7.5a is used for both turn-on and turn-off calculations ($C_{in} = 3.7$ nF and $t_f = 39.1$ ns).

♣

7.2 Application of the Thyristor

The basic gate requirements to trigger a thyristor into the conduction state are that the current supplied to the gate is

- of adequate amplitude and sufficiently short rise time
- of sufficient duration.

The gate conditions are subject to the anode being forward-biased with respect to the cathode. Figure 7.6 illustrates a typical thyristor gate current waveform for turn-on.

The initial high and rapid current quickly turns on the device so as to increase the anode initial *di/dt* capability. After a few microseconds the gate current can be decreased to a value in excess of the minimum gate requirement. After the thyristor has latched on, the gate drive may be removed in order to reduce gate power consumption, namely the losses. In some inductive load applications, where the load current lags, a continuous train of gate pulses is usually applied to ensure turn-on.

Gate drives can be divided broadly into two types, either electrically isolated or non-isolated. To obtain electrical isolation usually involves the use of a pulse-transformer or an opto-coupler but above a few kilovolts fibre-optic techniques are applicable.

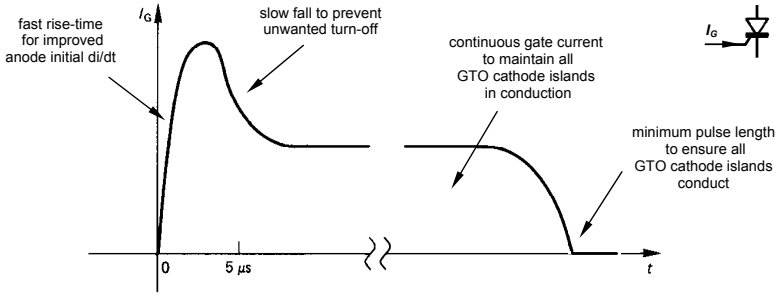


Figure 7.6. *Ideal thyristor gate current waveform for turn-on.*

7.2.1 Thyristor gate drive circuits

Only low-power thyristors with amplifying gates can be triggered directly from ttl or cmos. Usually a power interface stage is employed to convert ttl current sink and source levels of a few milliamps up to the required gate power levels.

Figure 7.7a and b shows two power interface circuits for triggering a triac. The triac could equally be another thyristor device. An important safety default feature of both these circuits is that no active device exists between the gate and Ml. During the off-state the gate is passively clamped by the resistor R_g to a voltage well below the minimum voltage level for turn-on.

Bidirectional gate current can bring the triac into conduction. Figures 7.7c and d show how negative gate turn-on current can be derived.

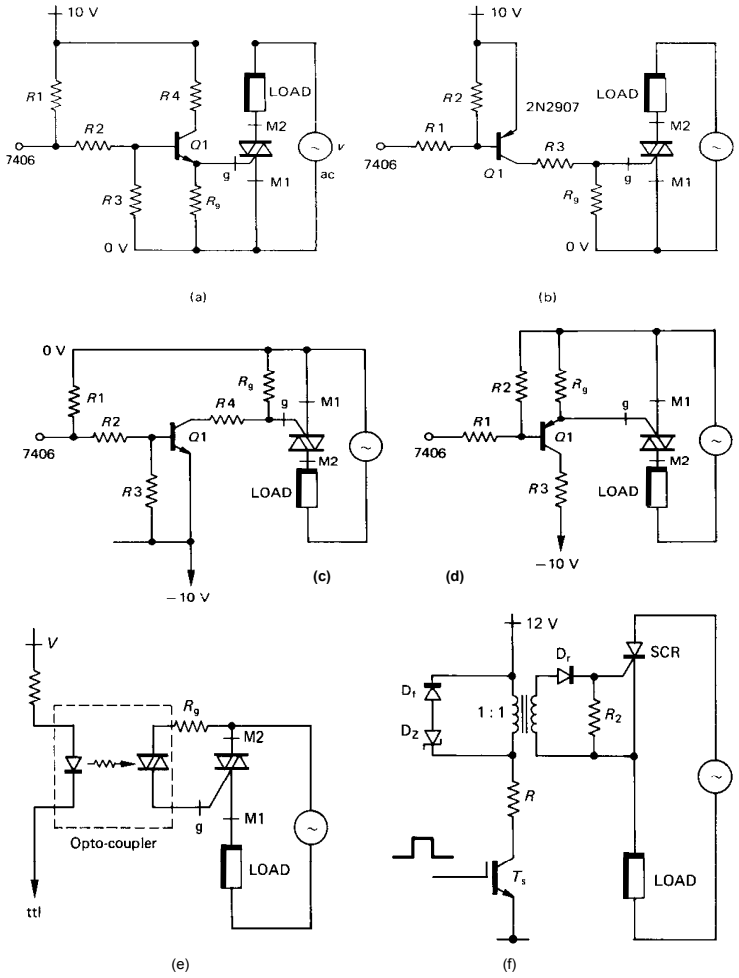


Figure 7.7. *Integrated circuit compatible triac gate drive circuits: (a) high level ttl activation; (b) low level ttl activation using an interfacing pnp transistor; (c) negative gate drive interface with high ttl output for triac activation; (d) negative gate drive interface with low ttl level for triac turn-on; (e) a triac opto-coupler isolated gate drive used to gate-drive a higher power triac; and (f) a pulse transformer drive isolated gate drive for a thyristor.*

If electrical isolation between the control circuitry and the power thyristor circuit is required, a simple triac opto-coupler can be employed as shown in figure 7.7e. The photo-triac is optically turned on which allows bidirectional main triac gate current to flow, the magnitude of which is controlled by the high-voltage resistor R_g. If the main device is an SCR, an opto-coupled SCR can be used for isolation and uni-direction gate triggering current.
When suitable voltage rails are not available or isolation is required, a pulse transformer drive circuit can be employed as shown in figure 7.7f. The diode/Zener diode series combination across the pulse transformer primary provides a path for primary magnetising current decay at turn-off and prevents saturation. The resistor R limits the secondary current into the SCR gate. This resistor can be placed in the pulse transformer primary or secondary by transforming the resistance in the turns ratio squared. If R is in the primary circuit and transformer saturation inadvertently occurs, the resistor R limits the current and protects the switching transistor T_s. The transformer secondary resistor R_2 is employed to decrease the gate to cathode impedance, thereby improving *dv/dt* capability, while the gate diode D_r prevents possible reverse gate voltage breakdown after T_s is turned off and the output voltage reverses during core reset. The transformer duty cycle must satisfy $t_{off} V_z \geq t_{on} V_s$, neglecting R.

i. Vacuum cleaner suction control circuit

In the vacuum cleaner suction control circuit in figure 7.8aX the triac is the power control element, itself controlled by a diac which is switched on by charging of C_1 through potentiometer R_2. The resistance of the diac is virtually infinite as long as its voltage it remains within the breakover voltage limits, $\pm V_{BO}$.
During each half cycle of the mains sinewave, C_1 charges until the voltage across it exceeds the diac breakover voltage. The diac then switches on and C_1 discharges itself into the gate of the triac which then switches on. Diodes D_1 and D_2 stabilise the supply voltage to the charging circuit so that its operation is independent of mains voltage fluctuations.
If $-V_{BO}$ and $+V_{BO}$ are equal and opposite, the triac will be triggered at the same time after the start of either a positive or negative half sinusoidal cycle. The conduction angle, and therefore the speed of the motor and the cleaner suction, is determined by the adjustment of R_2. Preset potentiometer R_3 is used to set the minimum suction level. The width and amplitude of the trigger pulses are kept constant by gate resistor R_4. The zinc oxide voltage dependent resistor, U, minimises the possibility of damage to the triac due to high voltage transients that may be superimposed on the mains supply voltage.

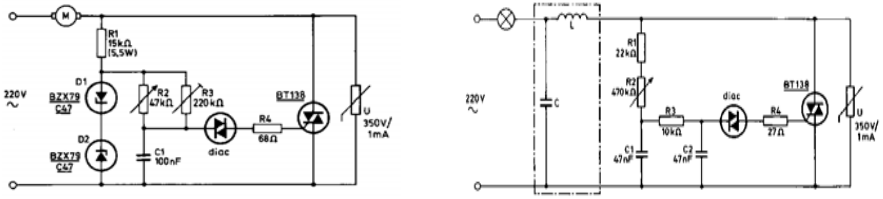


Figure 7.8. *Circuit diagram of: (a) vacuum cleaner suction controller and (b) a lamp dimmer.*

ii. Lamp dimmer circuit

A light dimmer circuit using a triac power control element, triggered via the diac, is shown in figure 7.8b. The potentiometer R_2 setting determines the phase difference between the mains sine wave and the voltage across C_2. This in turn sets the triac triggering angle and the lamp intensity.
The resistance of the diac is high as long as the voltage across it remains within its breakover voltage limits, $\pm V_{BO}$. Each half cycle of the mains charges C_2 via R_1, R_2 and R_3 until the voltage being applied to the diac reaches either of its breakover levels. The diac then conducts and C_2 discharges into the gate of the triac, switching it on. If $-V_{BO}$ and $+V_{BO}$ are equal and opposite, the triac will be triggered at the same time after the start of either a positive or negative half sinusoidal cycle. C_1 prevents the voltage across C_2 from changing abruptly after triggering, thus preventing progressively alteration of the phase relationship between the mains voltage and voltage across C_2. It thus prevents an undesirable hysteresis effect. The voltage across C_1 partially restores the voltage across C_2 after triggering and thereby minimizes the hysteresis effect. Gate resistor R4 keeps the width and amplitude of the trigger pulses constant. The VDR minimizes the possibility of the triac being damaged by high voltage transients that may be superimposed on the mains supply voltage.
Some form of filter is needed to comply with regulations concerning conducted and radiated interference. The simple LC filter shown within the dashed-lined box in figure7.8b is often adequate. The values of the filter components will vary, but a combination of 0.15mF capacitor and a low Q inductor of

2.5mH is generally sufficient for the circuit to meet the EMC limits. Circuits with intermittent loads, as with drills, sawing machines, and food mixers, are generally outwith EMC filtering regulations.

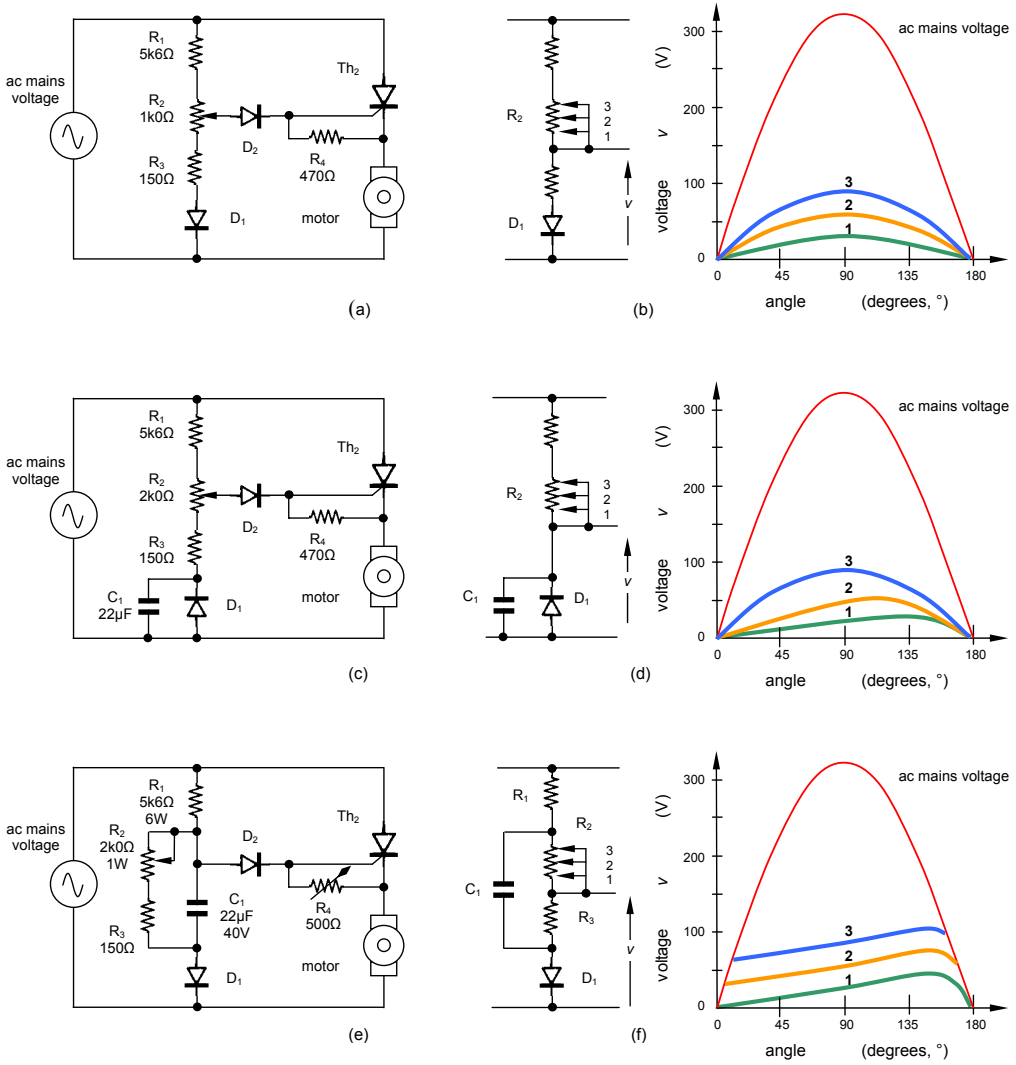


Figure 7.9. *Thyristor speed control circuit and gate waveforms using back EMF feedback: (a) basic controller, (b) improved low speed controller, and (c) improved low and high speed controller.*

iii. Back EMF feedback circuits

A motor speed control circuit, for electric drills, that employs back EMF to compensate for changes in motor load and mains voltage is shown in figure 7.9a. The series resistors R_1, R_2, R_3 and diode D_1 provide a positive going reference potential to the thyristor gate via diode D_2. Diode D_1 is used to reduce the dissipation in the series resistors and diode D_2 isolates the trigger circuit with the thyristor in the on-state. When the thyristor is not conducting the motor produces a back EMF voltage across the armature proportional to residual flux and motor speed. This appears as a positive potential at the thyristor cathode.

A thyristor fires when its gate potential is greater than cathode potential by a fixed amount. Depending on the waveform shape and amplitude at the gate, the circuit may function in several modes.
If, for example, during positive half cycles a constant dc potential was applied at the gate, figure 7.10, the thyristor would continue to fire at the beginning of each cycle until the back EMF was large enough to prevent firing. Thyristor firing would then continue intermittently at the beginning of the positive cycles to maintain some average motor speed.
Referring to figure 7.9a the waveform appearing at the thyristor gate will approximate to a half sine wave, figure 7.9b. As a result it is impossible for the firing angle to be after 90° - the most positive value of the trigger potential. At lower motor speeds the firing angle might need to be 130° for smooth operation. If the maximum firing angle is limited to 90° then intermittent firing and roughness of motor operation will result.
If, however, the waveform at the gate has a positive slope value to an angle of at least 130° then it will be possible to have a stable firing point at low speeds. Such a waveform can be produced if there is some phase shift in the trigger network.

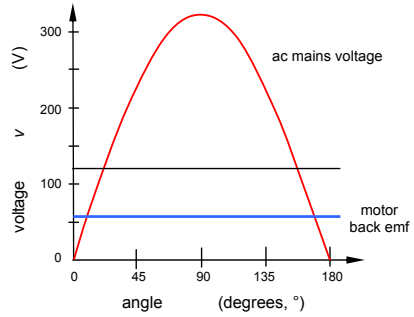


Figure 7.10. *Waveforms with a dc gate supply.*

Stable Firing at Small Conduction Angles
The trigger network of the circuit shown in figure 7.9c has been modified by the addition of a capacitor C_1 and diode D_1. The diode clamps the capacitor potential at zero during the negative going half cycles of the mains input. The waveform developed across the capacitor has a positive slope to some 140°, allowing thyristor triggering to be delayed to this point.
As R_2 is decreased, the peak of the waveform at the gate moves towards 90° as shown in figure 7.9d. As the speed increases, the no load firing angle also advances by a similar amount so stability will be maintained. This circuit will give smoother and more stable performance than the circuit of figure 7.9a. It will, however, give a marginally greater speed drop for a given motor loading at low speed settings. At the maximum speed settings the circuit of figure 7.9a approximates to that of figure 7.9c.

Improved Motor Performance with Stable Firing
The circuits in figure 7.9 a and c have gate voltage waveforms that are of near linear slope from the zero point of each positive half cycle, as seen in figures 7.9 b and d. This means that the only time that the thyristor can be fired early in the mains cycle, say at 20°, is when the back EMF and hence motor speed is low. This effect tends to prevent smooth running at high speeds and high loads.
Stable triggering, at low angles, can be achieved if the gate voltage ramp starts each cycle at a small positive level. This means that the time to reach the minimum trigger voltage is reduced. This is achieved by the circuit in figure 7.9e, where the capacitor C_1 is charged during positive half cycles via resistor R_1 and diode D_1. During negative half cycles the only discharge path for capacitor C_1 is via resistors R_2 and R3.
Diode D_1 also prevents C_1 from being discharged as the thyristor switches off by the inductively generated pulse from the motor. As the value of resistor R_2 is increased, capacitor C_1 is discharged less during negative half cycles but its charging waveform remains substantially unchanged. Hence the result of varying R_2 is to shift the dc level of the ramp waveform produced across C_1.
Diode D_2 isolates the triggering circuit when the thyristor is ON. Resistor R4 adjusts minimum speed, and by effectively bleeding a constant current, in conjunction with the gate current from the triggering circuit, it enables resistor R_2 to give consistent speed settings.

Circuit Design
If the speed controller is to be effective it must have stable thyristor firing angles at all speeds and give the best possible speed regulation with variations of motor load. The circuit in figure 7.9e gives a motor performance that satisfies both these requirements.
There are two factors that are important in the circuit operation in order to obtain the mentioned requirements.
- The value of positive slope of the waveform appearing at the thyristor gate.
- The phase angle at which the positive peak gate voltage is reached during a positive half cycle of mains input.

As described, the charging of capacitor C_1 through resistor R_1 determines the rate of rise of voltage at the thyristor gate during the positive half cycle. However, resistor R_1 must also have a resistance such that several times the maximum thyristor gate current passes through the RC network to D_1. This current will then give consistent speed settings with the spread of thyristor gate currents when the minimum speed is set by resistor R4.
The positive slope value of the thyristor gate voltage is fixed according to the motor used. A motor that gives a smooth back EMF voltage will allow a low slope value to be used, giving good torque speed characteristics.
Some motors have coarser back EMF waveforms, with voltage undulations and spikes, and a steeper slope of thyristor gate voltage must be used in order to obtain stable motor operation. The value of capacitor C_1 is chosen to provide the required positive slope of the thyristor gate voltage.
Some calculations have been made on the circuit of figure 7.9e simplified to the form of figure 7.11a, where it is assumed that current flowing to the thyristor gate is small compared with the current flowing through resistor R_1. An expression is derived later for the voltage that would appear at the anode of D_2 in terms of R_1, R_2 and C_1. Component values have been substituted into the expression to give the thyristor gate waveforms shown in figure 7.11b.

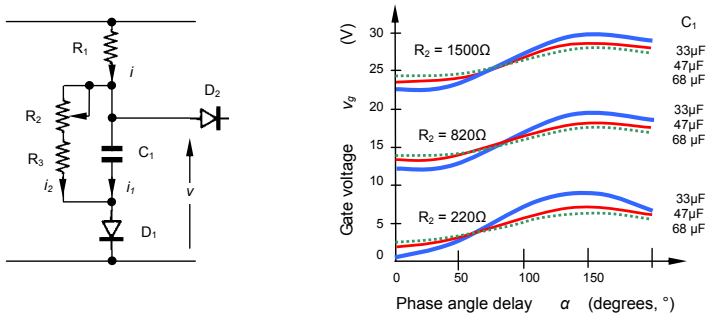


Figure 7.11. *Improved controller: (a) simplified firing circuit and (b) calculated gate waveforms.*

In order to adjust the circuit to suit a given motor, the back EMF of the motor must be known. This may be measured using the arrangement shown in figure 7.12. The voltage appearing across the motor is measured during the period when the series diode is not conducting (period A). The voltage so obtained will be the motor back EMF at its top speed on half wave operation, and corresponds to the back EMF that would be obtained from the unloaded motor at its highest speed when thyristor controlled. In practice, since the mains input is a sine wave, there is little increase in the 'no load' speed when the firing angle is reduced to less than about 70°.
The value of resistor R_2 in figure 7.9e determines the motor 'no load' speed setting. The waveforms of figure 7.11b may be used as a guide to obtaining the value of this resistor. It must be chosen so that at 70° and at its highest value, the gate voltage is higher than the measured back EMF by about 2V - the forward gate/cathode voltage of the thyristor.
The thyristor is turned ON when a trigger waveform, shown in figure 7.11b, exceeds the back EMF by the gate/cathode voltage. So, if the back EMF varies within a cycle then there will be a cycle to cycle variation in the firing angle. Normally, random variations of the firing angle by 20° are tolerable.
If, for example, there were variations in the back EMF of 1V, then with a firing angle of 70°and a capacitor of 32μF, the variation of firing angle would be about 12°. With capacitor values of 50μF and 64μF the firing angles variations would be 19° and 25° respectively. Therefore, a capacitance of 50μF would be suitable.

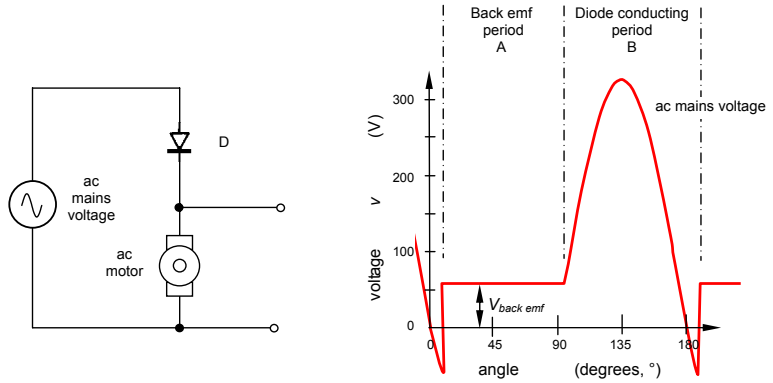


Figure 7.12. *Back EMF measurement circuit and typical voltage waveforms.*

Performance

The torque speed characteristics of the three circuits, when used to drive an electric drill, are compared in figure 7.13a. It may be seen that the circuit of figure 9c has a poorer performance than the two other circuits. That of figure 7.9e may be seen to give a similar performance to the circuit of figure 7.9a at low speeds but, at high speeds and torques, it is better. It should be noted that the circuits of figure 7.9 parts c and e provide low speed operation free from the intermittent firing and noise of the figure 7.9a circuit. Figure 7.13b compares the circuits of figure 7.9a and e when the load is a food mixer motor.

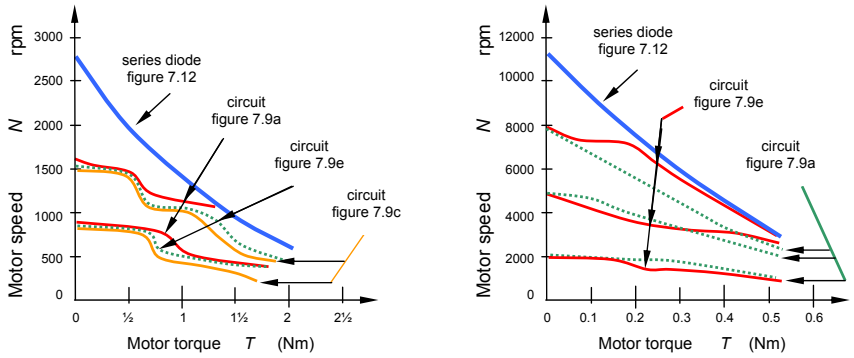


Figure 7.13. *Torque-speed load performance: (a) hand drill and (b) food mixer.*

Circuit Calculations

The following analysis derives an expression for voltage *v* at the anode of D_2. This expression can be used to produce the gate voltage waveforms shown in figure 7.11b. The analysis assumes that the current drawn by the thyristor gate is negligible in comparison with the current flowing in R_1.

The charging current i_1 for capacitor C_1 in figure 7.11a, is given by:

$$i_1 = \frac{dq}{dt} = C_1 \frac{dv}{dt}$$

and

$$i_1 = \frac{v}{R_2}$$

Representing a mains half sine wave by *f(E)* , where *E* is the peak mains voltage.

$$i = \frac{f(E) - v}{R_1} = i_1 + i_2$$

Therefore

$$\frac{f(E) - v}{R_1} = C_1 \frac{dv}{dt} + \frac{v}{R_2}$$

where *i*, i_1 and i_2 are instantaneous currents.

Simplifying

$$C_1 \frac{dv}{dt} + v\left(\frac{1}{R_1} + \frac{1}{R_2}\right) = \frac{f(E)}{R_1}$$

Fourier analysis of a half sinewave gives:

$$f(E) = E\left\{\frac{1}{\pi} + \tfrac{1}{2}\sin\theta - \frac{2}{\pi}\sum_{n=0,2,4,6..}\frac{\cos n\theta}{n^2 - 1}\right\}$$

Neglecting the Fourier terms for $n > 2$, then

$$C_1 \frac{dv}{dt} + v\left(\frac{1}{R_1} + \frac{1}{R_2}\right) = \frac{E}{R_1}\left\{\frac{1}{\pi} + \tfrac{1}{2}\sin\omega t - \frac{2}{3\pi}\cos 2\omega t\right\}$$

Then

$$C_1 \frac{dv}{dt} + v\left(\frac{1}{R_1} + \frac{1}{R_2}\right) - \frac{E}{\pi R_1} = \frac{E}{R_1}\left\{\tfrac{1}{2}\sin\omega t - \frac{2}{3\pi}\cos 2\omega t\right\}$$

Solving for v, the voltage that the trigger circuit would apply to the gate (assuming the gate draws no current), assuming $R_1+R_2 >> \sqrt{2\omega C_1 R_1 R_2}$:

$$v = \frac{R_2 E}{\pi(R_1 + R_2)} + \frac{R_2 E}{2\omega^2 C_1^2 R_1^2 R_2^2}\left\{(R_1 + R_2)\sin\omega t - \omega C_1 R_1 R_2 \cos\omega t - \frac{2}{3\pi}\omega C_1 R_1 R_2 \sin 2\omega t - \frac{1}{3\pi}(R_1 + R_2)\cos 2\omega t\right\}$$

Solving this equation for different values of C_1 and resistances for R_2 gives the curves shown in figure 7.11b.

7.2.2 *Thyristor gate drive design*

In order to design a thyristor gate interface circuit, both the logic and thyristor gate requirements must be specified.

Consider interfacing a typical ttl-compatible microprocessor peripheral which offers the following specification

$$I_{OH} = 0.3\text{mA} \ @ \ V_{OH} = 2.4\text{V}$$
$$I_{OL} = 1.8\text{mA} \ @ \ V_{OL} = 0.4\text{V}$$
$$V_{cc} = 5\text{V}$$

These specifications are inadequate for turning on a power thyristor or an optical interfacing device. If the power thyristor gate, worst case requirements are

$$I_{GT} = 75\text{ mA}, V_{GT} = 3\text{ V} \ @ \ -65°\text{C}$$

then a power interfacing circuit is necessary. Figure 7.14 shows an interfacing circuit utilising a p-channel MOSFET with the following characteristics

$$C_{gs} = 400\text{ pf} \qquad V_{TH} = 3.0\text{V}$$
$$R_{ds(on)} = 10\text{ ohms} \qquad I_d = 0.5\text{A}$$

The resistor R_1 limits the MOSFET C_{gs} capacitance-charging current and also specifies the MOSFET turn-on time. If the charging current is to be limited to 1.8 mA when V_{OL} = 0.4 V, then

$$R_1 = (V_{cc} - V_{OL})/I_{OL} \qquad \text{(ohms)}$$
$$= (5\text{V} - 0.4\text{V})/1.8\text{mA} = 2.7\text{ kilohms}$$

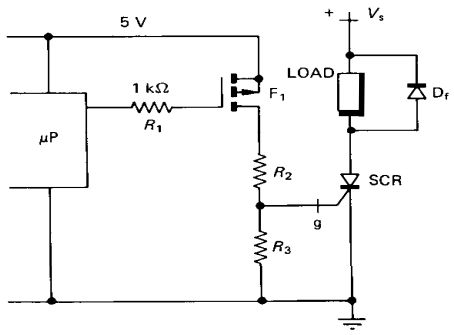


Figure 7.14. *Interfacing a microprocessor to a power thyristor.*

A smaller resistance could be used but this would not preserve the microprocessor low-voltage output level integrity if it were also being used as input to ttl logic. The MOSFET will not turn on until C_{gs} has charged to 3 V or, with a 5 V rail, approximately one *R*-*C* time constant. That is

$$t_{delay} = R_1 C_{gs} \quad \text{(s)}$$
$$= 2.7 \text{ kilohms} \times 400 \text{ pF} = 1 \text{ μs}$$

The MOSFET must provide the thyristor gate current and the current through resistor R_3 when the gate is at 3 V.
The maximum value of resistor R_2 is when $R_3 = \infty$ and is given by

$$R_2 = \frac{V_{cc} - V_{GT} - I_{GT} \times R_{ds(on)}}{I_{GT}}$$
$$= \frac{5\text{V} - 3\text{V} - 75 \text{ mA} \times 10\Omega}{75 \text{ mA}} = 16.6 \text{ ohms}$$

Use $R_2 = 10$ ohms.
The resistor R_3 provides a low cathode-to-cathode impedance in the off-state, thus improving SCR noise immunity. When $V_{GT} = 3$ V

$$I_d = \frac{V_{cc} - V_{GT}}{R_{ds(on)} + R_2} \quad \text{(A)}$$
$$= \frac{5\text{V} - 3\text{V}}{10\Omega + 10\Omega} = 100 \text{ mA}$$

of which 75 mA must flow into the gate, while 25 mA can flow through R_3. That is

$$R_3 = V_{GT} / (I_d - I_{GT}) \quad \text{(ohms)}$$
$$= 3 \text{ V} / 25 \text{ mA} = 120 \text{ ohms}$$

After turn-on the gate voltage will be about 1 V, hence the MOSFET current will be 200mA. Assuming 100 per cent on-state duty cycle, the I^2R power loss in the MOSFET and resistor R_2 will each be 0.4 W. A 1 W power dissipation 10 ohm resistor should be used for R_2.

Example 7.2: ***A light dimmer***

A diac with a breakdown voltage of ±30V is used in a light dimming circuit as shown in figure 7.15. If *R* is variable from 1kΩ to 22kΩ and *C* = 47nF, what are the maximum and minimum firing delays? What is the controllable output power range with a 10Ω load resistor?

Solution

The capacitor voltage v_c is given by

$$v_c = \frac{-j/\omega C}{R - j/\omega C} \times 240\angle 0°$$
$$= \frac{1}{1 + j\omega CR} \times 240\angle 0°$$

i. For $R = 1$ kΩ
$v_c = 237.36 \angle -8.4°$
that is, $v_c = 335.8 \sin(\omega t - 8.4°)$
The diac conducts when $v_c = 30$V, that is
minimum delay = $\omega t = 8.4° + \sin^{-1}(30\text{V}/335.8\text{V}) = 13.5°$

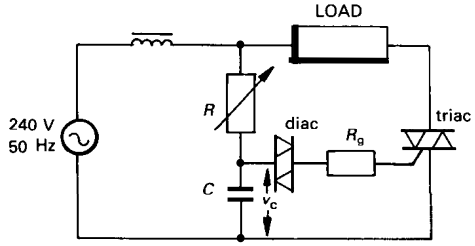


Figure 7.15. *Light dimmer.*

ii. For $R = 22$ kΩ
$v_c = 70.6 \angle -72.8°$
that is, $v_c = 99.8 \sin(\omega t - 72.8°)$
The diac conducts when $v_c = 30$V, that is
minimum delay = $\omega t = 72.8° + \sin^{-1}(30\text{V}/99.8\text{V}) = 92°$

The maximum power output, if continuous conduction were possible, is $\hat{P}_o = 240\text{V}^2 / 10\Omega = 5760\text{W}$.
From equation (12.16), the output power for a resistive load is given by

$$P_o = \frac{V_{rms}^2}{R} = \frac{V^2}{R}\left\{1 - \frac{2\alpha - \sin 2\alpha}{2\pi}\right\} \quad \text{(W)}$$

Minimum power at $\alpha = 92°$ (1.6 rad) is $P_o = \frac{240^2}{10\Omega} \times \left\{1 - \frac{2\times 92° - \sin 2\times 92°}{2\pi}\right\} = 2862\text{W}$

Maximum power at $\alpha = 13½°$ (0.24 rad) is $P_o = \frac{240^2}{10\Omega} \times \left\{1 - \frac{2\times 13½° - \sin 2\times 13½°}{2\pi}\right\} = 5536\text{W}$

♣

7.3 Drive design for GCT and GTO thyristors

The gate turn-off thyristor is not only turned on from the gate but, as its name implies, is turned off from its gate with negative gate current. Basic GTO thyristor gate current requirements are very similar to those for the power bipolar transistor (now virtually obsolete) when reverse base current is used for fast BJT turn-off.
Figure 7.16 shows a gate drive circuit for a GTO thyristor which is similar to that historically used for power bipolar junction transistor base drives. The inductor *L*, in figure 7.16, is the key turn-off component since it controls the *di/dt* of the reverse gate current. The smaller the value of *L*, the larger the reverse *di/dt* and the shorter the turn-off time. But with a shorter turn-off time the turn-off gain decreases, eventually to unity. That is, if the GTO thyristor is switched off rapidly, the reverse gate current must be of the same magnitude as the anode current to be extinguished. A slowly applied reverse gate current *di/dt* can produce a turn-off gain of over 20 but at the expense of increased turn-off saturation delay and switching losses. For the GTO thyristor *L* is finite to get a turn-off gain of more than one, while to achieve unity gain turn-off for the GCT, *L* is minimised.

The GTO thyristor cathode-to-gate breakdown voltage rating V_{RGM} specifies the maximum negative rail voltage. A level of -15 to -20V is common, and for supply rail simplicity a ± 15 V rail may be selected.
Resistor R_4 limits the base current of T_t. If an open collector ttl driver is employed, the current through R_4 is given by

$$I_{OL} = (V_{cc} - V_{beT_t} - V_{Db} - V_{OL}) / R_4 \quad \text{(A)}$$

For the open collector 74 ttl series, $I_{OL} = 40$ mA when $V_{OL} = 0.5$ V whence R_4 can be specified. The resistor R_3 speeds up turn-off of T_t. It is as large as possible to ensure that minimal base current is diverted from T_t. Diodes D_b and D_{as} form a Baker's clamp, preventing T_t from saturating thereby minimising its turn-off delay time.

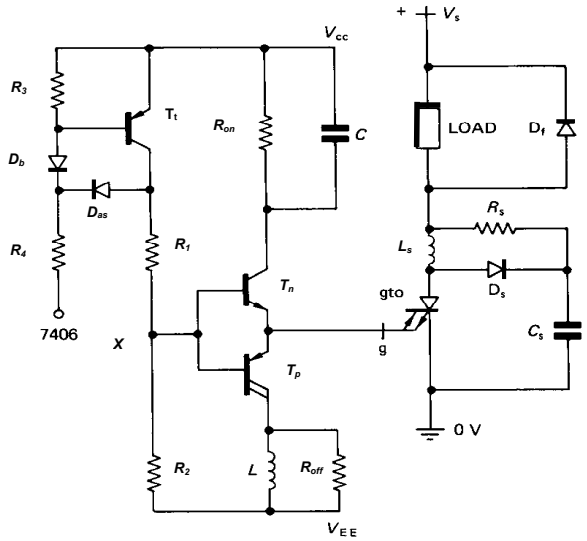


Figure 7.16. *Gate drive circuit and anode snubber circuits for a GTO thyristor.*

The two driver transistor T_n and T_p should

- have high gains
- be fast switching
- have collector voltage ratings in excess of $V_{cc} + V_{EE}$.

The GTO thyristor gate turn-on current is determined by resistor R_{on}, which is specified by

$$R_{on} = \frac{V_{cc} - V_{ceT_n} - V_{GC}}{I_G} \qquad \text{(Ohms)}$$

The power rating of R_{on} is given by

$$P_{R_{on}} = \delta(V_{cc} - V_{ceT_n} - V_{GC})I_G \qquad \text{(W)}$$

where δ is the maximum on-state duty cycle. The capacitor C_{on}, in parallel with R_{on}, provides a short current boost at turn-on, as shown in figure 7.6, thereby speeding up turn-on, increasing turn-on initial *di/dt* capability, and reducing turn-on losses.
The series resistors R_1 and R_2 bias the bases of the totem pole level shift driver and, for an on-condition, the potential of point X in figure 7.16 is given by

$$V_X = V_{beT_n} + V_{GC} \qquad \text{(V)}$$

The total current flow through R_1 is made up of the transistor T_n base current and that current flowing through R_2, that is

$$I_{R1} = \frac{I_G}{\beta_{T_n}} + \frac{V_X + V_{EE}}{R_2} \qquad \text{(A)}$$

from which

$$R_1 = (V_{cc} - V_X)/I_{R1} \qquad \text{(ohms)}$$

The power rating of R_1 is

$$P_{R_1} = \delta(V_{cc} - V_X)I_{R_1} \qquad \text{(W)}$$

For fast turn-off, if the reverse gate current at turn-off is to be of the same magnitude as the maximum anode current, then R_2 must allow sufficient base current to drive T_p. That is

$$R_2 = \frac{V_X + V_{beT_p}}{I_c/\beta\beta_p} \qquad \text{(ohms)}$$

Once the gate-to-cathode junction of the GTO has recovered, the reverse gate current decays to the leakage level. The power rating of R_2 can be low at lower switching frequencies.
The small inductor L in the turn-off circuit is of the order of microhenrys and it limits the rate of rise of reverse gate current, while R_{off} damps any inductor current oscillation.

The turn-on and turn-off BJT output totem pole in figure 7.16 can be replaced by suitable n-channel MOSFET circuitry in high power GCT and GTO thyristor applications. In high power IGCT applications, MOSFETs and rail decoupling electrolytic capacitors are extensively parallel connected. Typically 21 capacitors and 42 MOSFETs are parallel connected to provide a low impedance path for unity anode current extraction from the GCT gate. The gate inductance (including the GCT internal package inductance) is minimised, whence L is zero. Typically, the IGCT gate drive, gate connection, and internal package inductance are each about 2nH. This is achieved by minimising lengths, capacitive decoupling, and using parallel go and return paths. As a result, gate reverse *di/dt*'s of over 5kA/µs are attainable with a -15V dc negative gate supply.

Table 7.3: Gate drive isolation techniques summary

Technique	data transfer	power transfer	comments
Transformer	direct signal coupling	direct magnetic transfer	duty cycle limited corona breakdown limit
opto-coupler	slow, with capacitive effects	n/a	voltage and *dv/dt* limit
fibre optics	fast, virtually no voltage limit	n/a	best signal transmission at MV and HV
charge couple	n/a	requires switching	induced effects between ground level and gate level, LV application
bootstrap	n/a	requires switching	

Reading list

International Rectifier, *HEXFET Data Book,*
HDB-5, 1987.

Peter, J. M., *The Power Transistor in its Environment,*
Thomson-CSF, Sescosem, 1978.

Siliconix Inc., *Mospower Design Catalog,*
January 1983.

Grafham, D. R. *et al., SCR Manual,*
General Electric Company, 6th Edition, 1979.

Problems

7.1. Calculate suitable resistor values for the triac gate drive circuit in figure 7.7a, assuming a minimum gate current requirement of 50 mA and the gain of Q1 is 50 at 50 mA.

7.2. Repeat problem 7.1 for the circuits in figures
- 7.7b
- 7.7c
- 7.7d.

7.3. Repeat example 7.2 assuming a 2V triac gate threshold voltage for turn-on.

8

Protecting Diodes, Transistors, and Thyristors

All power switching devices attain better switching performance if some form of switching aid circuit, called snubber, is employed. Snubber activation may be either passive or active which involves extra power switches. Only passive snubbers, which are based on passive electrical components, are considered in this chapter, while active snubbers are considered in Chapter 9. Fundamentally, the MOSFET and IGBT do not require switching aid circuits, but circuit imperfections, such as stray inductance and diode recovery, can necessitate the need for some form of switch snubber protection.
Protection in the form of switching aid circuits performs three main functions:

- Divert switching losses from the switch thereby allowing a lower operating temperature, or higher electrical operating conditions for a given junction temperature.
- Prevent transient electrical stressing that may exceed *I-V* ratings thereby causing device failure.
- Reduce conducted and radiated electromagnetic interference

Every semiconductor switching device can benefit from switching protection circuits, but extra circuit component costs and physical constraints may dictate otherwise.
The bipolar diode suffers from reverse recovery current and voltage snap which induces high but short duration circuit voltages. These voltage transients may cause interference to the associated circuit and to nearby equipment. A simple series non-polarised *R-C* circuit connected in parallel to the stressed or offending device is often used to help suppress the voltage oscillation at diode turn-off. Such a suppression circuit can be effectively used on simple mains rectifying circuits when rectification causes conducted and radiated interference.
Although the MOSFET and IGBT can usually be reliably and safely operated without external protection circuitry, stringent EMC application emission restrictions may dictate the use of snubbers. In specific applications, the IGBT is extensively current derated as its operating frequency increases. In order to attain better device current utilization, at higher frequencies, various forms of switching aid circuits can be used to divert switching losses from the stressed semiconductor switch.
Generally, all thyristor devices benefit from a polarised turn-on switching aid circuit, which is based on a series connected inductor that is active at thyristor turn-on. Such an inductive turn-on snubber is obligatory for the high-power GCT and GTO thyristor. In order to fully utilise the GTO thyristor, it is usually used in conjunction with a parallel-connected capacitive turn-off snubber, which decreases device stressing during the turn-off transient. Triacs and rectifier grade SCRs and diodes tend to use a simple *R-C* snubber connected in parallel to the switch to reduce interference. The design procedure of the *R-C* snubber for a diode is different to that for the *R-C* snubber design for a thyristor device, because the protection objectives and initial conditions are different. In the case of a thyristor or rectifier diode, the objective is to control both the voltage rise at turn-off and recovery overshoot effects. For the fast recovery diode or any high-speed switch, the principal objectives are to control the voltage overshoot magnitude at diode snap recovery or at turn-off respectively, which are both exacerbated because of stray circuit inductance carrying current.

8.1 The non-polarised *R-C* snubber

The series *R-C* snubber is the simplest switching aid circuit and is connected in parallel to the device being aided. It is characterized by having low series inductance and a high transient current rating. These requirements necessitate carbon type resistors for low inductance, below a few watts, and metal film resistors at higher powers. The high current and low inductance requirements are also provided by using metallised, polypropylene capacitors with high *dv/dt* ratings of typically hundreds of V/µs.
Theoretically a purely capacitive snubber would achieve the required protection objectives, but series resistance is added to decrease the current magnitude when the capacitor is discharging and to damp any voltage oscillation by dissipating the oscillatory energy generated at turn-off when an over-voltage tends to occur.

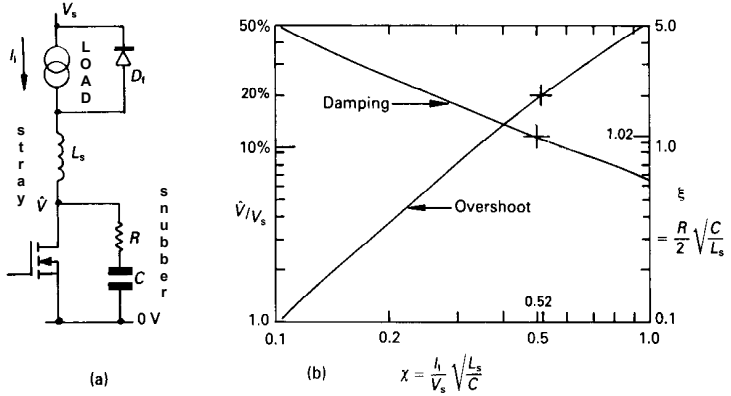


Figure 8.1. *MOSFET drain to source R-C snubber protection:*
(a) MOSFET circuit showing stray inductance, L_s, and R-C protection circuit and
(b) R-C snubber optimal design curves.

8.1.1 R-C switching aid circuit for the GCT, the MOSFET, and the diode

In figure 8.1a, at switch turn-off, stray inductance L_s unclamped by the load freewheel diode, D_f, produces an over voltage $\hat{V}$ on the MOSFET or IGBT. The energy associated with the inductor can be absorbed in the shown drain to source connected *R-C* circuit, thereby containing the voltage overshoot to a controlled safe level. Such an *R-C* snubber circuit is used extensively in thyristor circuits, 8.1.2, for *dv/dt* protection, but in such cases the initial current in the stray inductance is assumed zero. Here the initial inductor current is equal to the maximum load current magnitude, I_f. The design curves in figure 8.1b allow selection of *R* and *C* values based on the maximum voltage overshoot $\hat{V}$ and an initial current factor χ, defined in figure 8.1b. The *C* and *R* values are given by

$$C = L_s\,(I_f/\chi V_s)^2 \qquad \text{(F)} \tag{8.1}$$

$$R = 2\xi\, V_s\,\chi/\, I_f \qquad (\Omega) \tag{8.2}$$

If the *R-C* circuit time constant, $\tau = RC$, is significantly less than the MOSFET voltage rise and fall times, t_{rv} and t_{fv}, at reset (when the capacitor is discharged through the resistor and switch at turned on), a portion of the capacitor energy $\frac{1}{2}CV_s^2$, is dissipated in the switch, as well as in *R*. The switch appears as a variable resistor in series with the *R-C* snubber. Under these conditions (t_{fv} and $t_{rv} > RC$) the resistor power loss is approximately by

$$P_R = P_{Ron} + P_{Roff}$$

$$= \frac{\tau}{\tau + t_{fv}} P_{C0} + \frac{\tau}{\tau + t_{rv}}(P_{C0} + P_{L0}) \qquad \text{(W)} \tag{8.3}$$

where $P_{C0} = \frac{1}{2}CV_s^2 f_s$ and $P_{L0} = \frac{1}{2}L_s I_f^2 f_s$

otherwise (t_{fv} and $t_{rv} < RC$) the resistor losses are the energy to charge and discharge the snubber capacitor, plus the energy stored in the stray inductance, that is $2P_{C0} + P_{L0}$.
Note the total losses are independent of snubber resistance. The snubber resistor determines the time over which the energy is dissipated, not the amount of energy dissipated.
When the *R-C* snubber is employed across a fast recovery diode, the peak reverse recovery current is used for I_f in the design procedure.

Example 8.1: ***R-C snubber design for MOSFETs***

A MOSFET switches a 40 A inductive load on a 200 V dc rail, at 10 kHz. The unclamped drain circuit inductance is 20 nH and the MOSFET voltage rise and fall times are both 100 ns. Design a suitable *R-C* snubber if the MOSFET voltage overshoot is to be restricted to 240 V (that is, 40V overshoot, viz. 20%).

Solution

From figure 8.1b, for 20 per cent voltage overshoot

$$\xi = 1.02,\ \chi = 0.52$$

Using equations (8.1) and (8.2) for evaluating *C* and *R* respectively,

$$C = L_s\,(I_t/\chi V_s)^2 = 20\text{nH}\left(\frac{40\text{A}}{0.52\times 200\text{V}}\right)^2 = 3\text{nF}$$

$$R = 2\xi V_s \chi / I_t = 2\times 1.02\times\frac{0.52\times 200\text{V}}{40\text{A}} = 5.3\Omega$$

Use *C* = 3.3 nF, 450V dc, metallised polypropylene capacitor and *R* = 5.6 Ω.

Since the *RC* time constant, 18.5ns, is short compared with the MOSFET voltage transient times, 100ns, the resistor power rating is given by equation (8.3).

$$P_{C0} = \tfrac{1}{2}CV_s^2 f_s = \tfrac{1}{2}\times 3.3\text{nF}\times 200^2\times 10\text{kHz} = 2.64\text{W}$$

$$P_{L0} = \tfrac{1}{2}L_s I_t^2 f_s = \tfrac{1}{2}\times 20\text{nH}\times 40^2\times 10\text{kHz} = 0.16\text{W}$$

$$P_R = \frac{18.5\text{ns}}{100\text{ns}+18.5\text{ns}}\times 2.64\text{W} + \frac{18.5\text{ns}}{100\text{ns}+18.5\text{ns}}\times(2.64\text{W}+0.16\text{W}) = 0.85\text{W}$$

Use a 5.6 Ω, 1 W carbon composition resistor for low self inductance, with a working voltage of at least 250V dc. Parallel connection of two 12Ω ½W, carbon composition resistors may be necessary since resistance values below 10Ω are uncommon.
The MOSFET switching losses are $2W_{C0} + P_{L0} - 0.85\text{W} = 4.95\text{W}$ higher than those incurred by switching un-aided at 200V and 40A. From equations 6.9 and 6.10, the switching losses would be at least 8W, (4W+4W).

♣

8.1.2 Non-polarised R-C snubber circuit for a converter grade thyristor and a triac

The snubber circuit for a low switching frequency thyristor is an anode-to-cathode parallel connected *R-C* series circuit for off-state voltage transient suppression. Thyristor series inductance may be necessary (forming a turn-on snubber) to control anode *di/dt* at turn-on. This inductive snubber is essential for the GCT and the GTO thyristor, and will be considered in section 8.3.3.

Off-state dv/dt suppression snubber

Thyristors, other than the GCT and the GTO thyristor, normally employ a simple *R-C* snubber circuit as shown in figure 8.2. The purpose of the *R-C* snubber circuit is not primarily to reduce turn-off switching loss but rather to prevent false triggering (turn on) from applied or reapplied anode *dv/dt*, when the switch is in a forward voltage blocking off-state.
Any thyristor rate of rise of forward-voltage anode *dv/dt* produces a central junction charging current which may cause the thyristor to inadvertently turn on. The critical *dv/dt* is defined as the minimum value of *dv/dt* which will cause switching from the off-state to the on-state. In applications as shown in figure 8.2, an occasional false turn-on is generally not harmful to the triac or the load, since the device and the load only have to survive the surge associated with a half-a-cycle of the ac mains voltage supply.
In other applications, such as reversible converters, a false *dv/dt* turn-on may prove catastrophic. A correctly designed snubber circuit is therefore essential to control the rate of rise of anode voltage.
The action of this *R-C* snubber circuit relies on the presence of inductance in the main current path. The inductance may be stray, from transformer leakage or a supply, or deliberately introduced. Zero inductor current is the initial condition, since the device is in the off-state when experiencing the anode positive *dv/dt*. Analysis is based on the response of the *R-C* portion of an *L-C-R* circuit with a step input voltage and zero initial inductor current. Figure 8.3 shows an *L-C-R* circuit with a step input voltage and the typical resultant voltage across the SCR or *R-C* components. The circuit resistor *R* damps (by dissipating power) any oscillation and limits the capacitor discharge current through the SCR at subsequent SCR device turn-on initiated from the gate. The snubber resistor dissipates power even if the triac is not switching, since the snubber capacitor voltage alternates, tracking the ac voltage supply.

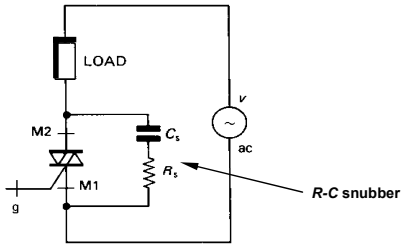


Figure 8.2. *Thyristor (triac) ac circuit with an R-C snubber circuit.*

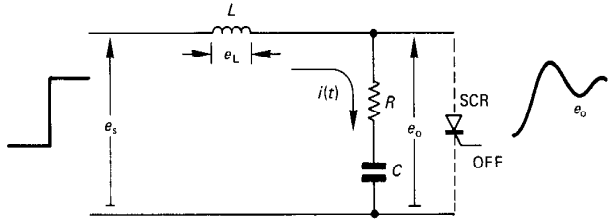


Figure 8.3. *Non-polarised R-C snubber equivalent circuit showing the second-order output response e_o to a step input voltage e_s.*

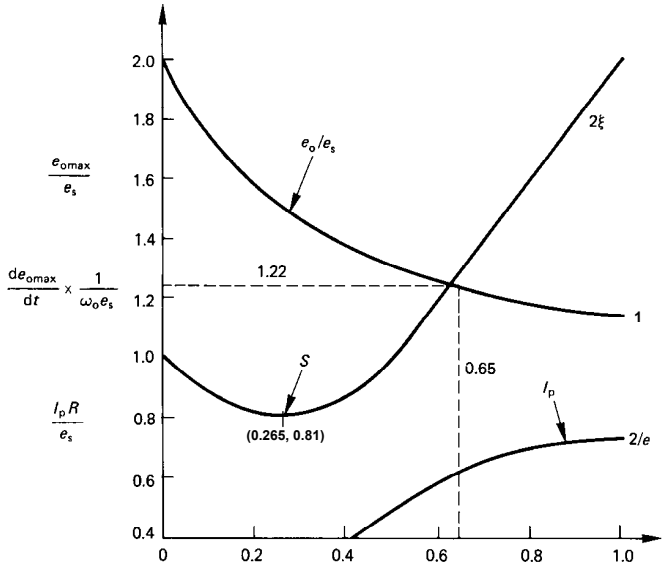


Figure 8.4. *Variation of snubber peak voltage, e_o, maximum de_o/dt, $\hat{S}$; and peak current, I_p; with L-C-R damping factor ξ.*

Based on the snubber circuit analysis presented in the appendix in section 8.5 at the end of this chapter, the maximum *dv/dt*, $\hat{S}$, which is usually specified for a given device, seen by the SCR for a step input of magnitude e_s, is given by

$$\hat{S} = e_s R / L \qquad \text{(V/s)} \tag{8.4}$$

for a damping factor of $\xi > ½$. That is, after rearranging, the snubber resistance is given by

$$R = L\hat{S}/e_s \qquad \text{(ohms)} \tag{8.5}$$

while the snubber capacitance is given by

$$C = \frac{4\xi^2 e_s}{R\hat{S}} \qquad \text{(F)} \tag{8.6}$$

and the peak snubber current is approximated by

$$\hat{I} = \frac{e_s}{R}\frac{2\xi}{\sqrt{1-\xi^2}} \qquad \text{(A)} \qquad \text{for } \xi < 1. \tag{8.7}$$

Figure 8.4 shows the variation of the different normalised design factors, with damping factor ξ.

Example 8.2: ***Non-polarised R-C snubber design for a converter grade thyristor***

Design an *R-C* snubber for the SCRs in a circuit where the SCRs experience an induced *dv/dt* due to a complementary SCR turning on, given

- peak switching voltage, e_s = 200 V
- operating frequency, f_s = 1 kHz
- *dv/dt* limit, *S* = 200 V/µs.

Assume

- stray circuit *L* = 10 µH
- 22 per cent voltage overshoot across the SCR
- an *L-C-R* snubber is appropriate.

Solution

From equation (8.5) the snubber resistance is given by

$$R = L\hat{S}/e_s = \frac{10\mu\text{H}\times 200\text{V}/\mu\text{s}}{200\text{V}} = 10\Omega$$

At turn-on the additional anode current from the snubber capacitor will be 200V/10Ω = 20A, which decays exponentially to zero, with a 1.8µs (10Ω×180nF) *RC* time constant.
Figure 8.4 shows the *R-C* snubber circuit overshoot magnitude, $\hat{e}_0/e_s$ for a range of damping factors ξ. The normal range of damping factors is between 0.5 and 1. Thus from figure 8.4, allowing 22 per cent overshoot, implies ξ = 0.65. From equation (8.6)

$$C = \frac{4\xi^2 e_s}{R\hat{S}} = \frac{4\times(0.65)^2\times 200\text{V}}{10\Omega\times 200\times 10^6}$$

$$= 180\text{ nF (preferred value) rated at 244 V peak.}$$

From equation (8.7) the peak snubber current during the applied *dv/dt* is

$$\hat{I} = \frac{e_s}{R}\frac{2\xi}{\sqrt{1-\xi^2}}$$

$$= \frac{200\text{V}}{10\Omega}\frac{2\times 0.65}{\sqrt{1\text{-}0.65^2}} = 34\text{ A}$$

The 10 ohm snubber resistor losses are given by

$$P_{10\Omega} = C e_0^2 f_s$$

$$= 180\times 10^{-9}\times 244^2\times 1\times 10^3 = 11\text{W}$$

Resistor current flows to both charge (maximum 34A) and discharge (initially 20A) the capacitor. The necessary 10Ω, 11W resistor must have lower inductance, hence two 22 Ω, 7W, 500V dc working voltage, metal oxide film resistors can be parallel connected to achieve the necessary ratings.

♣

Variations of the basic *R-C* snubber circuit are shown in figure 8.5. These circuits use extra components in an attempt to control SCR initial *di/dt* arising from snubber discharge through R_L at thyristor turn-on. Figure 8.5a has the disadvantage that three series devices (*C*-R_s-*D*) provide turn-off protection. The parasitic series inductance can be decreased by using a turn-off snubber with two series components (*C*-*D*), as shown in figure 8.5b.
An *R-C* snubber can be used across a diode in order to control voltage overshoot at diode snap-off during reverse recovery, as a result of stray circuit inductance, as considered in 8.1.1.
The *R-C* snubber can provide decoupling and transient overvoltage protection on both ac and dc supply rails, although other forms of *R-C* snubber circuit may be more applicable, specifically the soft voltage clamp.

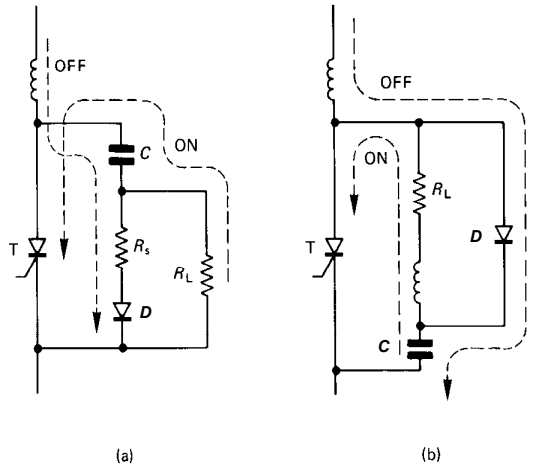


Figure 8.5. *Polarised variations of the basic thyristor R-C snubber: (a) $R_s << R_L$ and (b) transistor-type R-C-D snubber, R_s = 0.*

8.2 The soft voltage clamp

A primary function of the basic *R-C* snubber is to suppress voltage overshoot levels. The *R-C* snubber commences its clamping action from zero volts even though the objective is to clamp any switch over-voltage to the supply voltage level, V_s. Any clamping action below V_s involves the unnecessary transfer of energy. The soft voltage clamp reduces energy involvement since it commences clamping action once the switch voltage has reached the supply voltage V_s, and the voltage overshoot commences.
The basic polarised *R-C-D* soft voltage clamp is shown in figure 8.6a, with resistor *R* parasitic inductance, L_R, and stray or deliberately introduced unclamped inductance *L*, shown.
The voltage clamp functions at switch turn-off once the switch voltage exceeds V_s. The capacitor voltage never falls below the supply rail voltage V_s. Due to the stored energy in *L*, the capacitor *C* charges above the rail voltage and *R* limits current magnitudes as the excess capacitor charge discharges through *R* in to V_s. All the energy stored in *L*, $½LI_m^2$, is dissipated in *R*. The inductor current i_L and capacitor voltage V_c waveforms are shown in figure 8.6b.
At switch turn-on, the diode D blocks, preventing discharge of C which remains charged to V_s.
The energy drawn from the supply V_s as the capacitor overcharges, is returned to the supply as the capacitor discharges through *R* into the supply. The net effect is that only the energy in *L*, $½LI_m^2$, is dissipated in *R*.

Analysis is simplified if the resistor inductance L_R is assumed zero. The inductor current decreases from I_m to 0 according to

$$i_L(\omega t) = I_m \omega_0/\omega e^{-\alpha t}\cos(\omega t - \phi) \qquad \text{(A)} \tag{8.8}$$

where

$$\alpha = ½RC \quad \text{(s)} \qquad \omega_o = 1/\sqrt{LC} \quad \text{(rad/s)}$$

$$\omega = \sqrt{\omega_o^2 - \alpha^2} \quad \text{(rad/s)} \qquad \phi = \tan^{-1}\alpha/\omega \quad \text{(rad)}$$

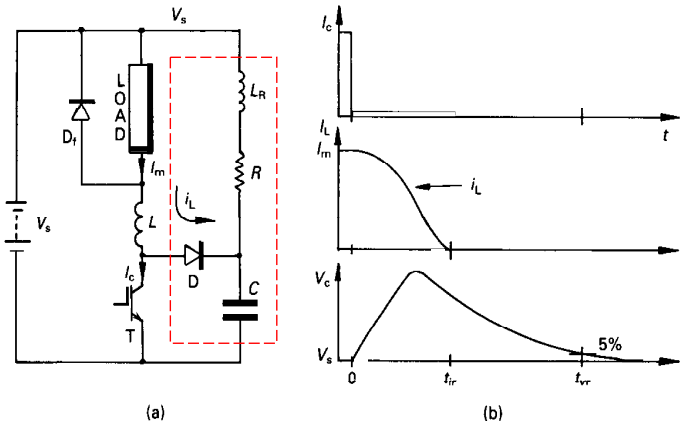


Figure 8.6. *Soft voltage polarised clamp:*
(a) circuit diagram and (b) turn-on inductor current, I_L, and capacitor voltage, V_c, at switch turn-off.

The inductor current reaches zero, termed the current reset time, t_{ir}, in time

$$t_{ir} = (\tfrac{1}{2}\pi + \phi)/\omega \quad \text{(s)} \qquad (8.9)$$

which must be shorter than the switch minimum off-time, t_{off}. The capacitor charges from V_s according to

$$V_c(\omega t) = V_s + \frac{I_m}{\omega C} e^{-\alpha t} \sin \omega t \quad \text{(V)} \qquad (8.10)$$

The maximum capacitor voltage, hence maximum switch voltage, occurs for large R

$$\hat{V}_c = V_s + I_m \sqrt{\frac{L}{C}} \quad \text{(V)} \qquad (8.11)$$

Once the current in L has reduced to zero the capacitor discharges to V_s exponentially, with a time constant RC.
The practical R-C circuit, which includes the stray inductance L_R, must be over-damped, that is

$$R > 2\sqrt{\frac{L_R}{C}} \quad (\Omega) \qquad (8.12)$$

The capacitor voltage reset time t_{vr} is the time for the capacitor to discharge to within 5 per cent of V_s, as shown in figure 8.6b.
The stray inductance L_R increases the peak capacitor voltage and increases the voltage reset time. Design of the voltage clamp, including the effects of L_R, is possible with the aid of figure 8.7. Design is based on specifying the maximum voltage overshoot, V_{cp} and minimizing the voltage reset time, t_{vr}, which limits the upper switching frequency, f_s, where $f_s \le 1/t_{vr}$ such that $t_{off} \ge t_{ir}$.

Example 8.3: ***Soft voltage clamp design***

A 5 μH inductor turn-on snubber is used to control diode reverse recovery current and switch turn-on loss, as shown in figure 8.6a. The maximum collector current is 25 A, while the switch minimum off-time is 5 μs and the maximum operating frequency is 50 kHz.

i. Assuming an independent L-C resonant transfer from L to C and a subsequent R-C discharge cycle, calculate soft voltage clamp R and C requirements.
ii. Use figure 8.7 to determine the voltage clamp requirements if the discharge (reset) resistor inductance L_R is
(a) 0
(b) 1.0μH.

In each case, the maximum switch overshoot is to be restricted to 50 V.

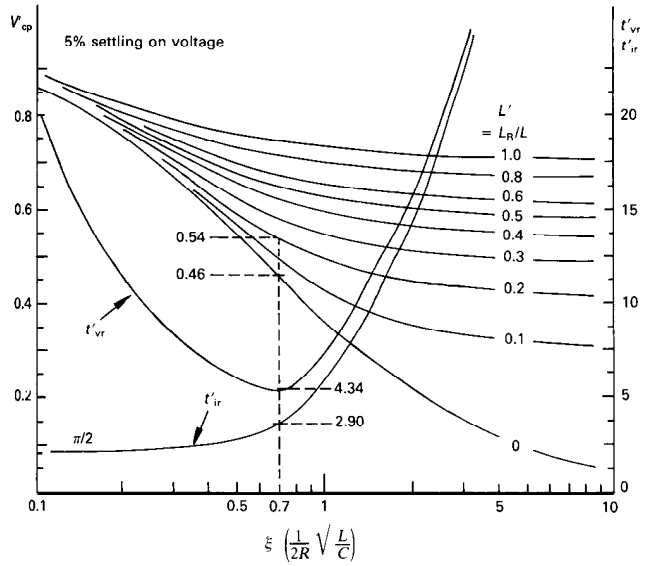


Figure 8.7. *Voltage clamp capacitor normalised peak over-voltage, V'_{cp}, versus damping factor, ξ, for different resistor normalised inductances, L', and voltage and current normalised settling times, t'_{vr}, $t'_{ir} = V_{cp}/\{I_m\sqrt{L/C}\}$, $t'_{vr} = t_{rv}/\omega_0$, $t'_{ir} = t_{ir}/\omega_0$.*

Solution

i. Assuming all the inductor energy is transferred to the clamp capacitor, before any discharge through R occurs, then from equation (8.11), for a 50 V capacitor voltage rise

$$50 = I_m\sqrt{L/C}$$

that is, C = 5 μH/(50V/25A)^2 = 1.25 μF (use 1.2 μF, rated at, at least 50V above the dc supply V_s).

From equation (8.9), for R = 0, the energy transfer time (from L to C) is

$$t_{ir} = \tfrac{1}{2}\pi\sqrt{LC} = \tfrac{1}{2}\pi\sqrt{5\mu\text{H}\times 1.25\mu\text{F}} = 4\mu\text{s}$$

which, as required, is less than the switch minimum off-time of 5 μs.
If the maximum operating frequency is 50 kHz, the capacitor must discharge in 20 - 4 = 16 μs.
Assuming five RC time constants for capacitor discharge

$$5\times RC = 16\mu\text{s}$$
$$R = 16\mu\text{s}/(5\times 1.2\mu\text{F}) = 2\tfrac{2}{3}\,\Omega \quad \text{(use } 2.4\Omega\text{)}$$

The resistor power rating is

$$P_R = \tfrac{1}{2}LI_m^2 f_s = \tfrac{1}{2}\times 5\mu\text{H}\times 25^2\times 50\text{kHz} = 78\text{W}$$

Obviously with a 2.4 Ω discharge resistor and 50V overshoot, discharge current would flow as the capacitor charges above the voltage rail. A smaller value of C could be used. A more accurate estimate of C and R values is possible, as follows.

ii. (a) L_R = 0, that is L' = L_R/L = 0
From figure 8.7, for the minimum voltage reset time, as indicated

$$V'_{cp} = 0.46, \quad t'_{ir} = 2.90, \quad t'_{rv} = 4.34, \quad \text{and} \quad \xi = 0.70$$

From $V'_{cp} = V_{cp}/I_m\sqrt{L/C}$

$0.46 = 50\text{V}/25\text{A}\sqrt{5\mu\text{H}/C}$ gives $C = 0.27\mu\text{F}$

From $\xi = 1/2R\sqrt{L/C}$, $R = 1/2\xi\sqrt{L/C} = \dfrac{1}{2\times 0.7}\sqrt{5\mu\text{H}/0.27\mu\text{F}} = 3.2\Omega$

(Use 3.3 Ω, 78 W)

The reset times are given by

$$t_{vr} = t_{vr}'\sqrt{LC} = 4.34 \times 1.16 = 5\mu s \quad (<20\mu s)$$

$$t_{ir} = t_{ir}'\sqrt{LC} = 2.9 \times 1.16 = 3.4\mu s \quad (<5\mu s)$$

It is seen that smaller capacitance (0.27 μF vs 1.2 μF) can be employed if simultaneous *L-C* transfer and *R-C* discharge are accounted for. The stray inductance of the resistor discharge path has been neglected. Any inductance decreases the effectiveness of the *R-C* discharge. Larger *C* than 0.27 μF and *R* < 3.3Ω are needed, as is now shown.

ii. (b) L_R = 1μH, that is, $L' = L_R/L = 0.2$
In figure 8.7, for a minimum voltage reset time, ξ = 0.7, V_{cp}' = 0.54 when the L' = 0.2 curve is used. The normalised reset times are unchanged, that is t_{ir}' = 2.9 and t_{vr}' = 4.34.
Using the same procedure as in part ii b

$$0.54 = 50\text{V}/25\text{A}\sqrt{5\mu\text{H}/C} \text{ gives } C = 0.37\mu\text{F} \text{ (use } 0.39\mu\text{F)}$$

$$R = 1/2\xi\sqrt{L/C} = \frac{1}{2\times 0.7}\sqrt{5\mu\text{H}/0.39\mu\text{F}} = 2.6\Omega \quad \text{(use } 2.7\Omega, 78\text{W)}$$

Since resistor inductance has been accounted for, parallel connection of four 10Ω, 25W wire-wound aluminium clad resistors can be used.

$$t_{vr} = 4.34\times 1.4 = 6\mu s \quad (< 20\mu s)$$

$$t_{ir} = 2.90\times 1.4 = 4\mu s \quad (< 5\mu s)$$

Note that circuit supply voltage V_s is not a necessary design parameter, other than to specify the capacitor absolute dc voltage rating. This supply independence is expected since in ac circuit analysis, as is applicable here, dc voltage sources are shorted.

♣

8.3 Polarised switching-aid circuits

Optimal gate drive electrical conditions minimize collector (or drain or anode) switching times, thus minimizing switch electrical stresses and power losses. Proper gate drive techniques greatly enhance the switching robustness and reliability of a power switching device. Switching-aid circuits, commonly called *snubber circuits,* can be employed to further reduce device switching stresses and losses. Optimal gate drive conditions minimise the amount of snubbering needed.

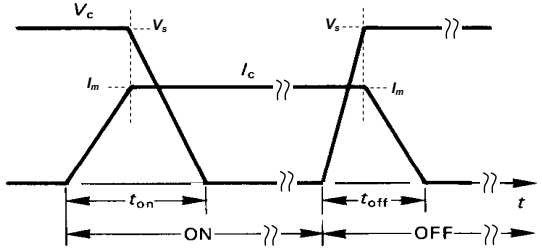


Figure 8.8. *Idealised collector (anode) switching waveforms for an inductive load.*

During both the switch-on and the switch-off transition intervals, for an inductive load as considered in chapter 6.2, an instant exists when the switch simultaneously supports the supply voltage V_s and conducts the full load current I_m, as shown in figure 8.8. The gate drive conditions cannot alter this peak power loss but can vary the duration of the switching periods (t_{on} and t_{off}). From chapter 6, for an inductive load, the switching losses, *W*, dissipated as heat in the switch, are given by

for turn-on: $W_{on} = \frac{1}{2}V_s I_m t_{on}$ (J) (8.13)

for turn-off: $W_{off} = \frac{1}{2}V_s I_m t_{off}$ (J) (8.14)

In order to reduce switching losses, two snubber circuits can be employed on a power switching device, one operational during switch turn-on, the other effective during turn-off. In the case of the turn-off snubber, energy (current) is diverted from the switch turning off into a parallel capacitor as shown in figure 8.9a. The turn-on snubber utilises an inductor in series with the collector as shown in figure 8.9b in order to support part of the dc voltage supply as the collector (anode) voltage falls. The inductor therefore controls the rate of rise of collector (anode) current during the collector voltage fall time. For both snubbers, the *I-V* SOA trajectory is modified to be within that area shown in figure 6.8.

- *A series* ***inductive turn-on snubber*** is essential for the GCT and the GTO thyristor in order to control the anode initial *di/dt* current to safe levels at switch turn-on. In large area thyristor devices, the inductor controlled current increase at turn-on allows sufficient time for the silicon active area to spread uniformly so as to conduct safely the prospective load current. Special thyristor gate structures such as the amplifying gate, as shown in figure 3.24, allow initial anode *di/dt* values of up to 1000 A/us. Use of an inductive turn-on snubber with the MOSFET and the IGBT is limited but may be used because of freewheel diode imposed limitations rather than an intrinsic need by the switch.
- *The shunt* ***capacitive turn-off snubber*** is used extensively on the GTO thyristor. The *R-D-C* circuit is necessary to ensure that GTO turn-off occurs at a low anode-to-cathode voltage, preventing excessive power loss at the central GTO junction during reverse recovery. Larger area GTOs employ 1 to 8 μF in an *R-D-C* turn-off snubber and at high voltages and frequencies the associated losses, $\frac{1}{2}C_s V_s^2 f_s$, tend to be high. To reduce this loss, GTOs with an increased SOA, namely GCTs, for use without a turn-off snubber are available. These devices under utilise their voltage and current density capabilities as compared with when used with a turn-off snubber.

While the switching performance of IGBTs and MOSFETs can be enhanced by using the turn-off snubber, it is not a prerequisite for safe, reliable switch operation.

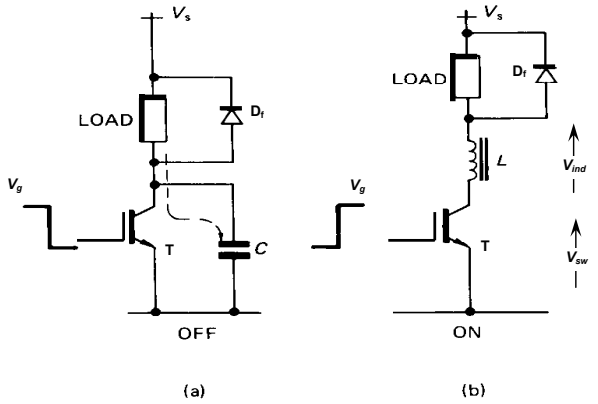


Figure 8.9. *Basic switching-aid circuits comprising:*
(a) a parallel capacitor for current shunting at switch turn-off and (b) a series inductor for supporting voltage, thus limiting the rate of rise of principal current at turn-on.

8.3.1 The polarised turn-off snubber circuit - assuming a linear current fall

Figure 8.10 shows a complete turn-off snubber circuit comprising a capacitor-diode plus resistor combination across the anode-to-cathode/collector-to-emitter terminals of the switching device. At switch turn-off, load current is diverted into the snubber capacitor *C* via the diode D, while the switch principal current decreases. The anode/collector voltage is clamped to the capacitor voltage, which is initially zero. The larger the capacitor, the slower the anode/collector voltage rises for a given load current and, most importantly, turn-off occurs without a condition of simultaneous supply voltage and maximum load current (V_s, I_m). Figure 8.11 shows the anode/collector turn-off waveforms for different magnitudes of snubber capacitance. The GTO/IGBT tail current has been neglected, thus the switching device is analysed without any tail current. For clarity, the terminology to be henceforth used, refers to an IGBT, viz., collector, emitter, and gate. Circuit operational explanations equally apply to thyristors.

Figure 8.11a shows turn-off waveforms for a switch without a snubber, where it has been assumed that the collector voltage rise time is short compared with the collector current fall time, which is given by $i_c(t) = I_m(1 - t/t_f)$. For low capacitance values, the snubber capacitor (whence collector voltage) may charge to the rail voltage before the collector current has fallen to zero, as seen in figure 8.11b. For larger capacitance, the collector current reaches zero before the capacitor (whence collector voltage) has charged to the rail voltage level, as shown in figure 8.11c.

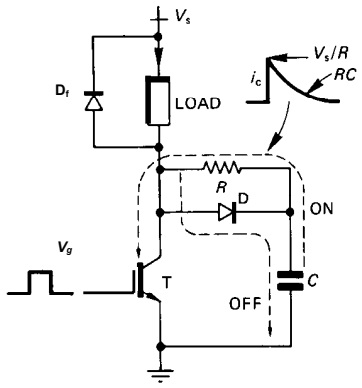


Figure 8.10. *Practical capacitive turn-off snubber showing capacitor charging and discharging paths during device switching.*

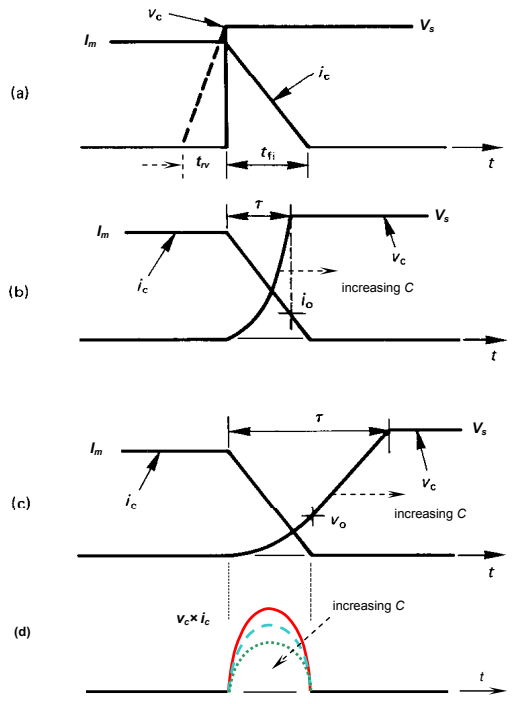


Figure 8.11. *Switch turn-off waveforms: (a) unaided turn-off; (b) turn off with small snubber capacitance; (c) turn-off with large snubber capacitance; (d) and switch power losses.*

For analysis, the collector voltage rise time for an unaided switch is assumed zero. The device switch-off energy losses without a snubber, as shown in figure 8.11a, are given by

$$W = \tfrac{1}{2} V_s I_m t_{fi} \quad \text{(J)} \qquad (8.15)$$

With a snubber circuit, switch losses are decreased as shown in figure 8.11d, but snubber (resistor) losses are incurred. After turn-off the capacitor is charged to the rail voltage. This stored energy, $\tfrac{1}{2} C_s V_s^2$, is subsequently dissipated as heat in the snubber circuit resistor at subsequent switch turn-on, when an *R-C* discharge current flows. If the snubber *RC* time constant is significantly shorter than the switch voltage fall time at turn-on, the capacitor energy dissipated in the resistor is less than $\tfrac{1}{2} C_s V_s^2$ and switch losses are increased as considered in 8.1.1. A range of capacitance values exists where the total losses - snubber plus switch - are less than those losses incurred if the same device is switched unaided, when losses as given by equation (8.14) result. Two distinct snubber design cases exist, depending on capacitance magnitude, as indicated by figures 8.11b and 8.11c. The two possibilities and the associated circuit voltage and current waveforms in each case are shown in detail in figure 8.12. The waveforms are based on satisfying Kirchhoff's voltage and current laws for each case.

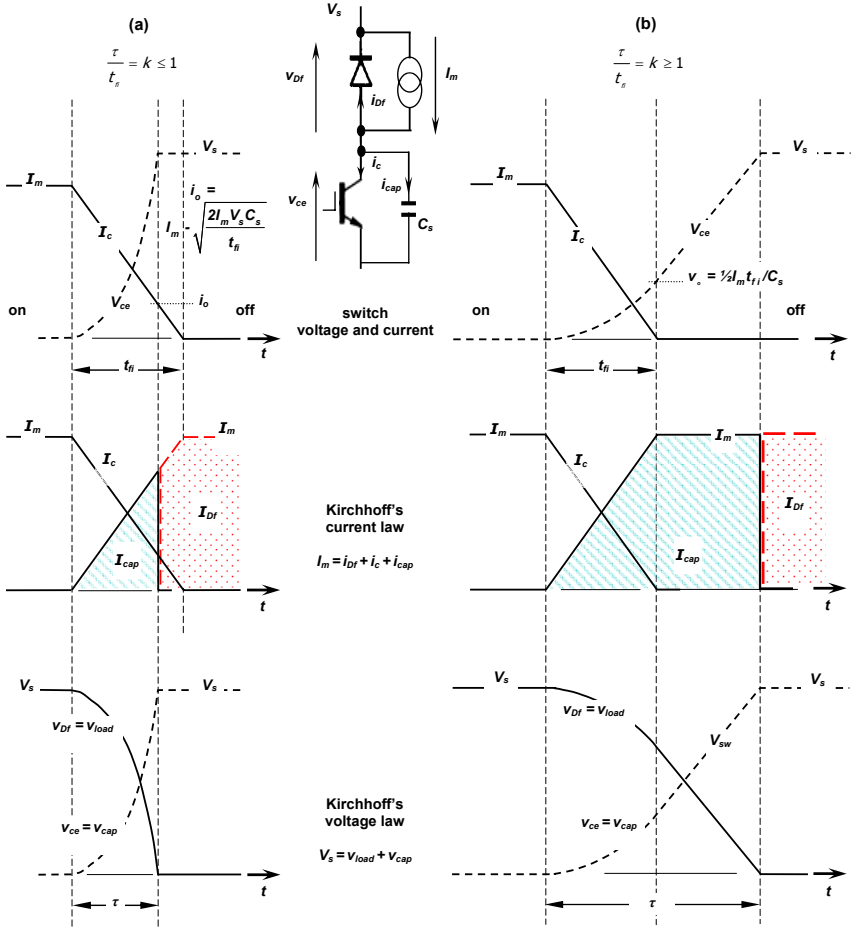


Figure 8.12. *Switch turn-off waveforms satisfying Kirchhoff's laws: (a) turn-off with small snubber capacitance and (b) turn-off with large snubber capacitance.*

From $i = C\,dv/dt$, the snubber capacitor charges according to $v_c(t) = V_s\,(t/\tau)^2$, to V_s before the collector current has reached zero, thus the switch losses are given by

$$W_t = \tfrac{1}{2}V_s I_m t_{fi}\left(1 - \tfrac{1}{3}k + \tfrac{1}{2}k^2\right) \quad \text{(J)} \tag{8.16}$$

for $k \le 1$, where $k = \tau/t_{fi}$, as defined in figures 8.12a and 8.13.
Alternatively, with larger capacitance, if the snubber capacitor charges to $v_o < V_s$, according to $v_c(t) = v_o\,(t/t_{fi})^2$, thus not charging to V_s until after the collector current reaches zero, that is $k \ge 1$, then the switch losses are given by

$$W_t = \tfrac{1}{2}V_s I_m t_{fi} \Big/ 6(2k-1) \quad \text{(J)} \tag{8.17}$$

for $k \ge 1$ as defined in figures 8.12b and 8.13. Initially the capacitor voltage increase is quadratic, then when the collector current reaches the load current level, the capacitor voltage increase becomes linear.

These losses, normalised with respect to the unaided switch losses given by equation (8.15), are plotted in figure 8.13. The switch and capacitor (subsequently resistor) components contributing to the total losses are also shown. A number of points arise concerning turn-off snubbers and snubber losses.

(a) Because of current tailing, voltage overshoot, and the assumption that the voltage rise time t_{rv} is insignificantly short, practical unaided switch losses, equation (8.14), are approximately twice those indicated by equation (8.15).
(b) As the snubber capacitance increases, that is, k increases, the switch loss is progressively reduced but at the expense of increased snubber associated loss.
(c) If $k \le 1.41$ the total losses (switch and reset resistor) are less than those for an unaided switch. In the practical case $k \le 2.70$ would yield the same condition.

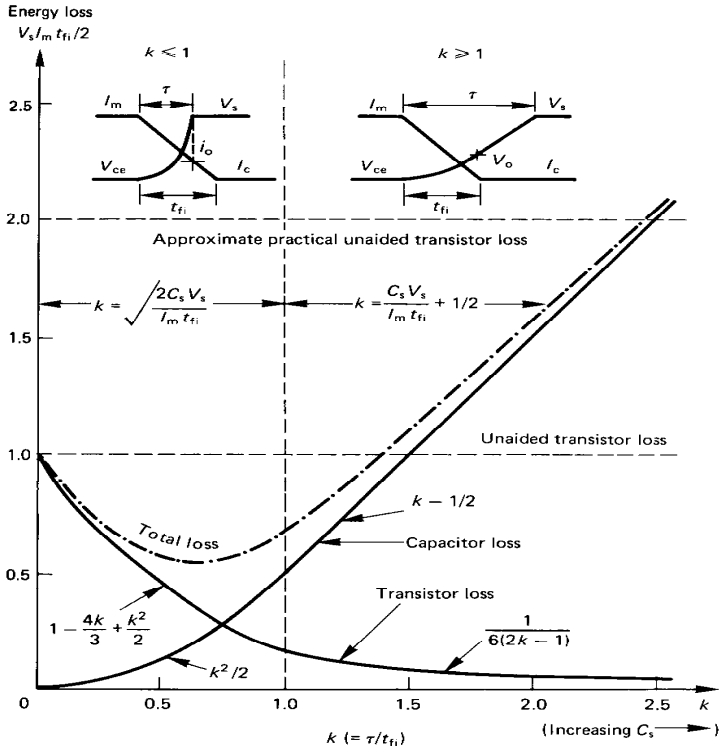


Figure 8.13. *Loss components for a switch at turn-off when employing a capacitance-type snubber and assuming the collector current falls according to* $i_c = I_m(1 - t/t_{fi})$.

(d) A minimum total loss (switch plus reset resistor) condition exists. When $k = \frac{2}{3}$ the total losses are only *5/9* those of an unaided switch. The snubber capacitance for this optimal case is given by

$$C_s = \frac{2}{9}\frac{I_m t_{fi}}{V_s} \quad \text{(F)} \tag{8.18}$$

(e) Losses are usually minimised at the maximum loss condition, that is maximum load current I_m. At lower currents, the capacitor charging time is increased, increasing the output voltage distortion.
(f) Snubbers not only reduce total losses, but because the loss is distributed between the switch and resistor, more effective heat dispersion can be achieved.
(g) High switch current occurs at turn-on and incorporates the load current I_m, the snubber capacitor exponential discharge $V_s/R\left(1 - e^{-t/CR}\right)$, and any freewheel diode reverse recovery current.

The capacitor energy $\frac{1}{2}C_s V_s^2$ is removed at turn-on and is exponentially dissipated mainly in the snubber circuit resistor R. The power rating of this resistor is independent of resistance but dependent on the maximum switching frequency. The reset resistor power rating is given by

$$P_{R_r} = \tfrac{1}{2}C_s V_s^2 f_s \quad \text{(W)} \tag{8.19}$$

Two factors specify the snubber circuit resistance value.

- The snubber circuit RC time constant period must ensure that after turn-on the capacitor discharges before the next switch turn-off is required. If $\hat{t}_{on}$ is the minimum switch on-time, then $\hat{t}_{on} = 5R_sC_s$, is sufficient to ensure the correct snubber circuit initial conditions, namely, zero capacitor volts.
- The resistor initial current at capacitor discharge is V_s/R_s. This component is added to the load current at switch turn-on, hence adding to the turn-on stresses. The maximum collector current rating must not be exceeded. In order to reduce the initial discharge current, a low valued inductor can be added in series with the resistor, (or a wire-wound resistor used), thus producing an overdamped *L-C-R* discharge current oscillation at turn-on. Note that the resistor power loss in equation (8.19) is independent of resistance value. The resistance determines the period of time over which the capacitor stored energy is dissipated at switch turn-on.

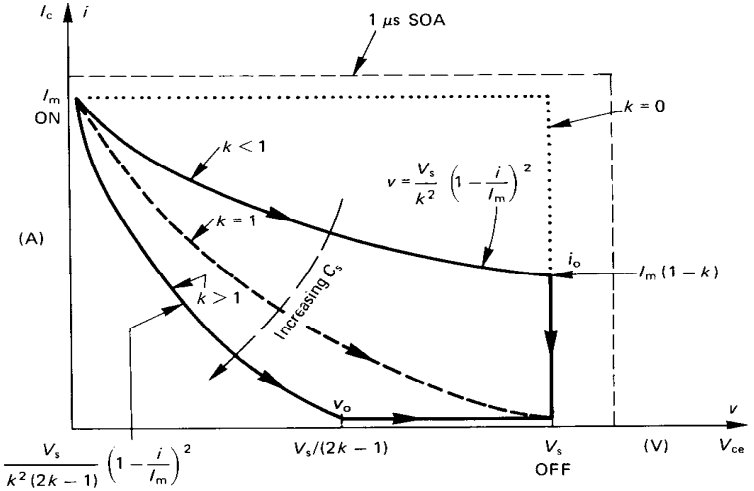


Figure 8.14. *The collector I-V trajectory at turn-off with a capacitive switching-aid circuit.*

As a result of utilising a capacitive turn-off snubber, the collector trajectory across the SOA is modified as shown in figure 8.14. It is seen that the undesired unaided condition of simultaneous supply voltage V_s and load current I_m is avoided. Typical trajectory conditions for a turn-off snubbered device are shown for three situations, depending on the relative magnitudes of t_{fi} and τ (the magnitude of C_s). A brief mathematical derivation describing the turn-off switching-aid circuit action is presented in the appendix in section 8.6 at the end of this chapter.

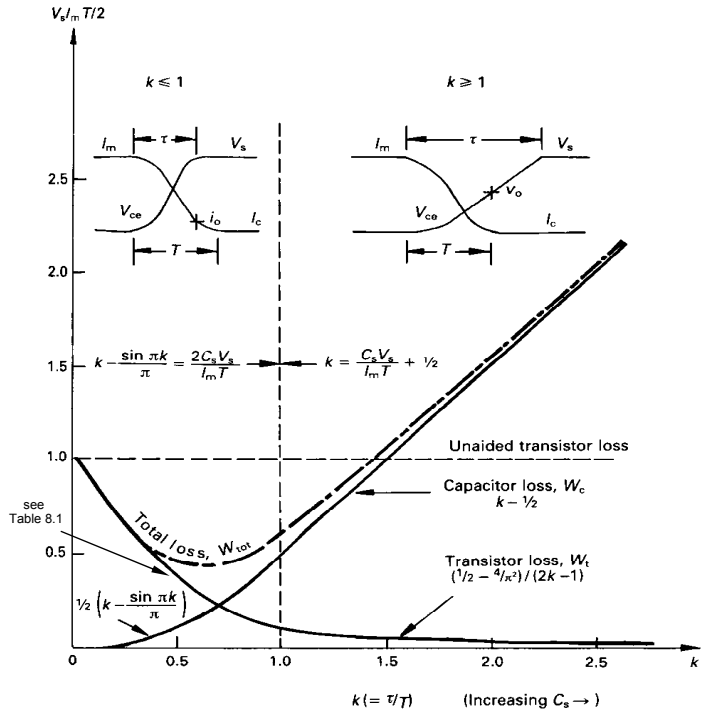


Figure 8.15. *Loss components for a switch at switch-off when employing a capacitance-type snubber and assuming a collector fall current according to* $i_c = \frac{1}{2}I_m\{1 + \cos(\pi t/T)\}$.

8.3.2 *The turn-off snubber circuit - assuming a cosinusoidal current fall*

As an alternative to a linear current fall at turn-off, it may be more realistic to assume that the current falls cosinusoidally according to

$$i_c(t) = \tfrac{1}{2}I_m(1+\cos \pi t/T) \qquad \text{(A)} \qquad (8.20)$$

for $0 \le t \le T$, as shown in figure 8.15.
As with a linear current fall, two cases exist.

(i) $\tau \le T$ ($k \le 1$), that is the snubber capacitor charges to V_s in time τ, before the switch current reaches zero, at time T.
(ii) $\tau \ge T$ ($k \ge 1$), that is the snubber capacitor charges to the supply V_s after the switch current has fallen to zero.

These two cases are shown in figure 8.15 where k is defined as τ/T. Using a similar analysis as presented in the appendix (section 8.6), expressions can be derived for switch and snubber resistor losses. These and the total losses for each case are summarised in table 8.1.
Figure 8.15 shows that a minimum total loss occurs, namely

$$W_{total} = 0.41 \times \tfrac{1}{2} V_s I_m T \quad \text{at} \quad k = 0.62$$

when

$$C_s = 0.16 \frac{I_m T}{V_s} \qquad \text{(F)} \qquad (8.21)$$

For $t_{fi} < 0.85T$, a cosinusoidal fall current predicts lower total losses than a linear fall current, with losses shown in figure 8.13.

Table 8.1: Normalised switching loss components at turn-off with a cosinusoidal current fall of half period *T*

× $V_s I_m T/2$ loss	$k = \tau/T \le 1$	$k = \tau/T \ge 1$
Switch W_t	$\dfrac{\frac{k^2}{2} + \frac{2}{\pi^2}(\cos \pi k - 1) + \frac{k}{\pi}\sin \pi k - \frac{\sin^2 \pi k}{2\pi^2}}{\left(k - \frac{\sin \pi k}{\pi}\right) + 1 - k - \frac{\sin \pi k}{\pi}}$	$\dfrac{\frac{1}{2} - \frac{4}{\pi^2}}{2k - 1}$
Resistor W_c	$\left(k - \frac{\sin \pi k}{\pi}\right)\Big/ 2$	$k - \frac{1}{2}$
W_{tot}	$W_t + W_c$	$\left(k - \frac{1}{2}\right) + \left(\frac{1}{4} - \frac{2}{\pi^2}\right)\Big/\left(k - \frac{1}{2}\right)$

Example 8.4: ***Capacitive turn-off snubber design***

A 600V, 100A machine field winding is switched at 10kHz. In maintaining a constant field current, the switch operates with an on-state duty cycle ranging between 5% and 95% ($5\% \le \delta \le 95\%$) and has a turn-off linear current fall time of 100ns, that is, $i_c(t) = 100 \times (1 - t/100\text{ns})$.

i. Estimate the turn-off loss in the switch.
ii. Design a capacitive turn-off snubber using the dimensionally correct identity $i = Cdv/dt$. What is the capacitor voltage when the collector current reaches zero.
iii. Design a capacitive turn-off snubber such that the switch voltage reaches 600V at the same time the conducting current reaches zero.

In each snubber case calculate the percentage decrease in un-aided switch turn-off power dissipation.

Solution

i. The switch un-aided turn-off losses are given by equation (8.14). The turn-off time is greater than the current fall time (since the voltage rise time t_{rv} has been neglected), thus the turn-off switching losses will be greater than

$$W_{off} = \tfrac{1}{2} V_s I_m t_{off} = \tfrac{1}{2} \times 600\text{V} \times 100\text{A} \times 100\text{ns} = 3\text{mJ}$$
$$P_{off} = W_{off} \times f_s = 3\text{mJ} \times 10\text{kHz} = 30\text{W}$$

ii. Use of the equation $i = Cdv/dt$ results in a switch voltage that reaches the rail voltage after the collector current has fallen to zero. From $k = \tfrac{1}{2} + C V_s / I_m t_{fi}$ in figure 8.13, $k = 3/2$ satisfies the dimensionally correct capacitor charging equation. Substitution into $i = Cdv/dt$ gives the necessary snubber capacitance

$$100\text{A} = C \frac{600\text{V}}{100\text{ns}}$$
$$\text{that is} \quad C = 16\tfrac{2}{3}\text{nF}$$

Use an 18nF, 1000V dc, metallised polypropylene, high *dv/dt* capacitor.
The snubber capacitor discharges at switch turn-on, and must discharge during the switch minimum on-time. That is

$$\overset{\vee}{t}_{on} = 5\,CR$$
$$5\% \text{ of } 1/10\text{kHz} = 5 \times R \times 18\text{nF}$$
$$\text{that is} \quad R = 55.5\Omega \quad \text{Use } 56\Omega$$

The discharge resistor power rating is independent of resistance and is given by

$$P_{56\Omega} = \tfrac{1}{2} C V_s^2 f_s$$
$$= \tfrac{1}{2} \times 18\text{nF} \times 600\text{V}^2 \times 10\text{kHz} = 32.4\text{W} \quad \text{Use } 50\text{W}.$$

The resistor can be wire-wound, the internal inductance of which reduces the initial peak current when the capacitor discharges at switch turn-on. The maximum discharge current into the switch during reset, which is added to the 100A load current and any diode reverse recover current, is

$$I_{56\Omega} = V_s / R = 600\text{V} / 56\Omega = 10.7\text{A}$$

which decays exponential to zero in five time constants, 5µs. The peak switch current (neglecting freewheel diode recovery) is 100A+10.7A=110.7A, at turn-on.

At switch turn-off, when the switch current reduces to zero, the snubber capacitor has charged to a voltage less than the 600V rail voltage, specifically

$$v_0 = \frac{1}{C}\int i_{cap}\,dt$$

$$= \frac{1}{16\tfrac{2}{3}\,\text{nF}} \int_0^{100\text{ns}} 100\text{A} \times \left(\frac{t}{100\text{ns}}\right) dt = 300\text{V} \quad (277\text{V with } 18\text{nF})$$

The switch turn-off losses are reduced from 30W to

$$P_{off} = f_s \int_0^{100\text{ns}} i_c v_{ce}\,dt = f_s \int_0^{100\text{ns}} I_m \left(1-\frac{t}{100\,\text{ns}}\right) \times v_0 \left(\frac{t}{100\text{ns}}\right)^2 dt$$

$$= f_s \int_0^{100\text{ns}} 100\text{A}\left(1-\frac{t}{100\,\text{ns}}\right) \times 300\text{V}\left(\frac{t}{100\text{ns}}\right)^2 dt = 2.5\text{W} \quad (2.3\text{W with } 18\text{nF})$$

The total turn-off losses (switch plus snubber resistor) are 2.5W+32.4W=34.9W, which is more than the 30W for the unaided switch. Since the voltage rise time has been neglected in calculating the un-aided losses, 34.9W would be expected to be less than the practical un-aided switch losses. The switch losses have been reduced by 91⅔%, from 30W to 2.5W.

iii. As the current in the switch falls linearly to zero, the capacitor current increases linearly to 100A (*k* = 1), such that the load current remains constant, 100A. The capacitor voltage increases in a quadratic function according to

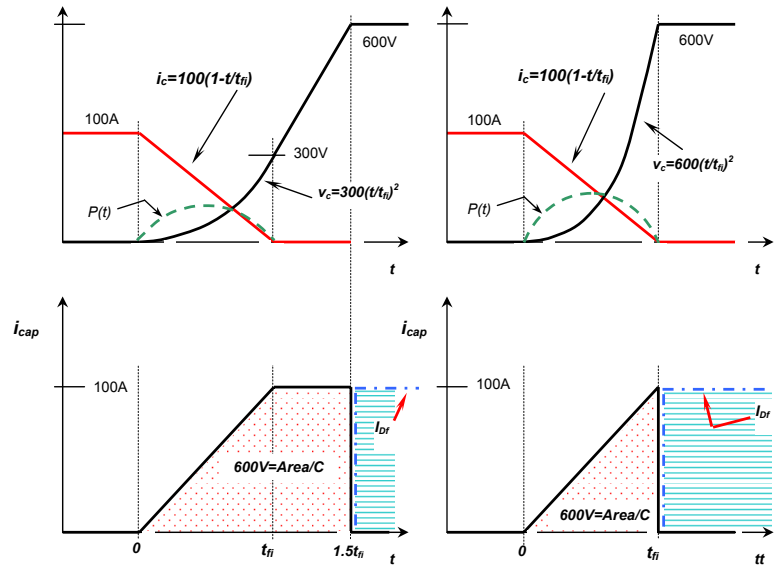
$$v_{cap}(t) = \frac{1}{C}\int i_{cap}\,dt$$

The capacitor charges quadratically to 600V in 100ns, as its current increases linearly from zero to 100A, that is

$$600\text{V} = \frac{1}{C}\int_0^{100\text{ns}} 100\text{A}\,\frac{t}{100\text{ns}}\,dt$$

that is $C = 8\tfrac{1}{3}\,\text{nF}$

Use a 10nF, 1000V dc, metallised polypropylene, high *dv/dt* capacitor.



Example 8.4. part (b) *Example 8.4. part (c)*

The necessary reset resistance to discharge the 10nF capacitor in 5µs is

$$5\mu\text{s} = 5 \times R \times 10\text{nF}$$

that is $R = 100\Omega$

The power dissipated in the reset resistor is

$$P_{100\Omega} = \tfrac{1}{2}CV_s^2 f_s$$

$$= \tfrac{1}{2} \times 10\text{nF} \times 600\text{V}^2 \times 10\text{kHz} = 18\text{W}$$

Use a 100Ω, 25W, wire-wound, 600V dc withstand voltage, metal clad resistor.
The resistance determines the current magnitude and the period over which the capacitor energy is dissipated. The resistance does not determine the amount of energy dissipated. The capacitor exponentially discharges with an initial current of 600V/100Ω = 6A, which adds to the 100A load current at switch turn-on. The peak switch current is therefore 100A+6A = 106A, at turn-on.
The energy dissipated in the switch at turn-off is reduced from 30W when un-aided to

$$P_{off} = f_s \int_0^{100\text{ns}} i_c v_c\,dt = f_s \int_0^{100\text{ns}} I_m \left(1-\frac{t}{100\,\text{ns}}\right) \times V_s \left(\frac{t}{100\text{ns}}\right)^2 dt$$

$$= f_s \int_0^{100\text{ns}} 100\text{A}\left(1-\frac{t}{100\,\text{ns}}\right) \times 600\text{V}\left(\frac{t}{100\text{ns}}\right)^2 dt = 5\text{W} \quad (\text{using } 8\tfrac{1}{3}\,\text{nF})$$

The total losses (switch plus snubber resistor) with a turn-off snubber are 5W+18W =23W, which is less than the 30W for the unaided switch. The switch loss has been decreased by 83⅓%, (30W to 5W).

Note that the losses predicted by the equations in figure 8.13 amount to 5W + 15W = 20W. The discrepancy is due to the fact that the preferred value of 10nF with *k* = 1.2 giving 5W + 18W = 23W (rather that the calculated 8⅓nF, *k* =1) has been used for the resistor loss calculation.

♣

8.3.3 *The polarised turn-on snubber circuit – with air-core (non-saturable) inductance*

A series turn-on snubber comprises an inductor-diode combination in the collector circuit as shown in figure 8.16. At turn-on the inductor controls the rate of rise (from zero) of collector current and supports a portion of the supply voltage while the collector voltage falls. At switch turn-off the energy stored in the inductor, $\tfrac{1}{2}L_s I_m^2$, is transferred in the form of current through the diode and dissipated in the diode D_s and any added series resistance *R*, and in the resistance of the inductor.

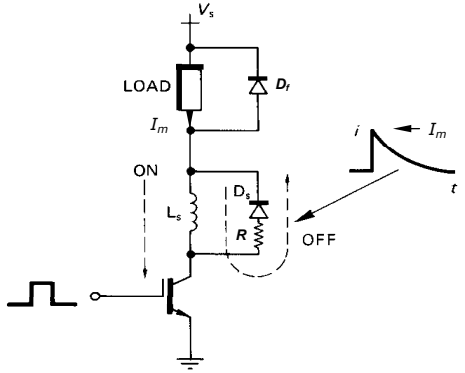


Figure 8.16. *Turn-on switching-aid circuit incorporating series inductance, L_s.*

Figure 8.17 shows collector turn-on waveforms with and without a turn-on snubber circuit. The turn-on loss associated with an unaided switch, figure 8.17a, neglecting the current rise time, is given by

$$W = \tfrac{1}{2}V_s I_m t_{fv} \quad \text{(J)} \qquad (8.22)$$

where it is assumed that the collector current rise time is zero and that the collector voltage falls linearly, according to $v_c(t) = V_s(1 - t/t_{fv})$.
When an inductive turn-on snubber circuit is employed, collector waveforms as in figure 8.17b or 8.17c result.
The two possibilities and the associated circuit voltage and current waveforms in each case are shown in detail in figure 8.18. The waveforms are based on satisfying Kirchhoff's voltage and current laws for each case.

For low inductance the collector current reaches its maximum value I_m (the load current) before the collector voltage has reached zero. As shown in figure 8.17b, from $v = L\ di/dt$, the collector current increases quadratically $i_c(t) = I_m (t/\tau)^2$ and the switch turn-on loss is given by

$$W_t = \tfrac{1}{2} V_s I_m t_{fv} \left(\tfrac{1}{2}k^2 - \tfrac{4}{3}k + 1\right) \quad \text{(J)} \qquad (8.23)$$

for $k \le 1$, where $k = \tau / t_{fv}$ as defined in figure 8.17.

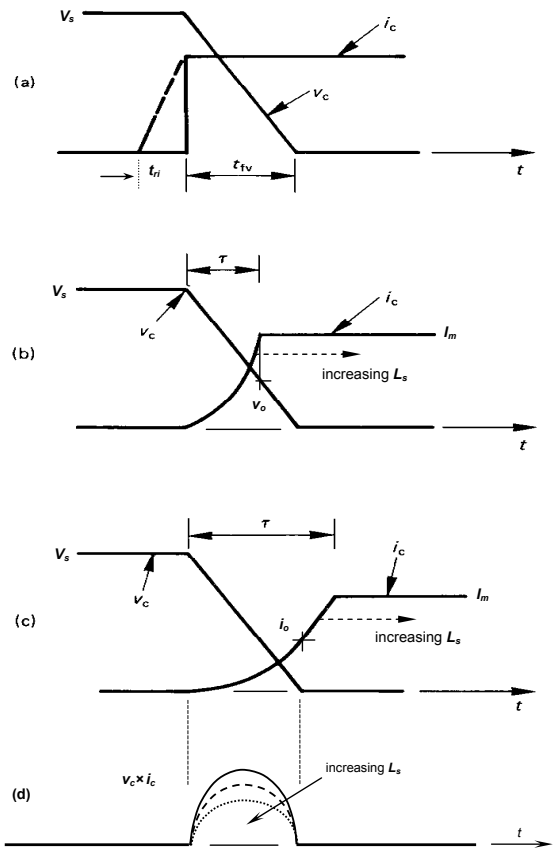


Figure 8.17. *Switch voltage and current collector waveforms at turn-on: (a) without a snubber; (b) and (c) with an inductive snubber; and (d) switch power losses.*

These losses include both switch losses and stored inductor energy subsequently dissipated. For higher snubber inductance, the collector voltage reaches zero before the collector current reaches the load current level. Initially the inductor current increases quadratically $i_{Ls}(t) = i_o (t/t_{fv})^2$, then when the collector voltage has reached zero, the current increases linearly. The switch loss is given by

$$W_t = \tfrac{1}{2} V_s I_m t_{fv} \frac{1}{6 \times (2k - 1)} \quad \text{(J)} \qquad (8.24)$$

Note that these equations are similar to those for the turn-off snubber, except that the current fall time t_{fi} is replace by the voltage fall time, t_{fv}. The normalised loss components for the capacitive snubber in figure 8.13 are valid for the inductive turn-on snubber.

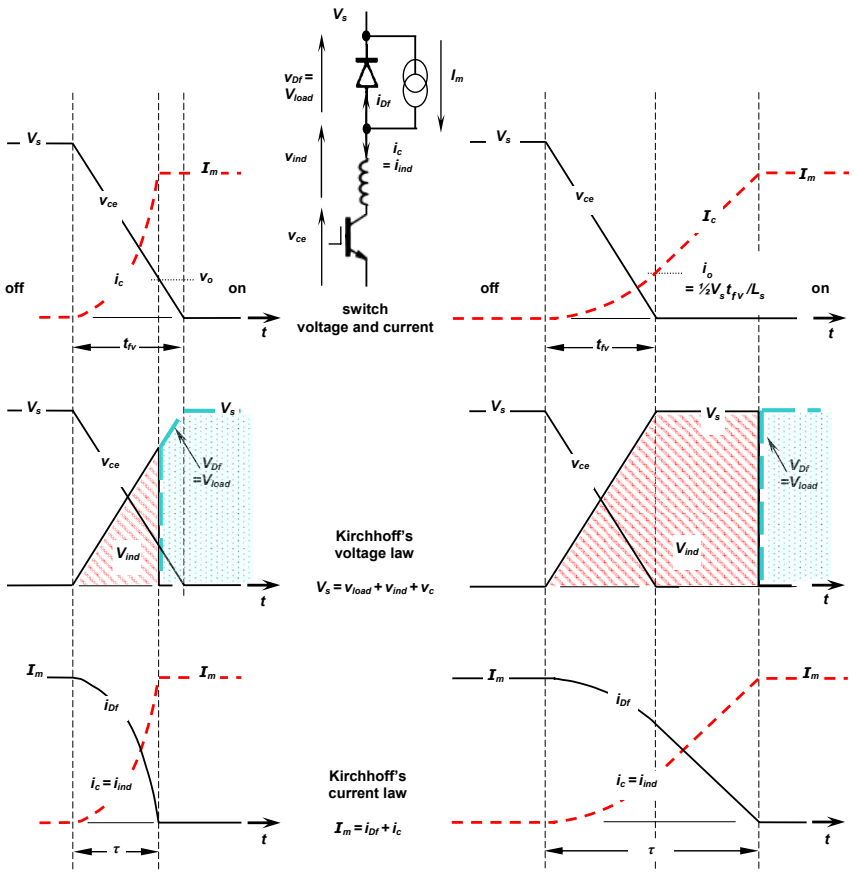


Figure 8.18. *Turn-on snubber waveforms satisfying Kirchhoff's laws: (a) turn-on with small snubber inductance and (b) turn-on with large snubber inductance.*

Minimum total turn-on losses of 5/9 those of the un-aided case, occur at $k = ⅔$ when

$$L_s = \frac{2}{9} \frac{V_s t_{fv}}{I_m} \quad \text{(H)} \qquad (8.25)$$

At switch turn-off, the snubber inductance stored energy is dissipated as heat in the snubber freewheeling diode path. The maximum power loss magnitude is dependent on the operating frequency and is given by

$$P_{L_s} = \tfrac{1}{2} L_s I_m^2 f_s \quad \text{(W)} \qquad (8.26)$$

This power is dissipated in the inductor winding resistance, resistance R, and freewheeling diode D_s. The time constant is designed such that $t_{off} = 5L_s / R$ where t_{off} is the minimum device off-time, where R is the effective total series resistance. The time constant can be reduced either by increasing the series resistance or by inserting a Zener diode as shown in figure 8.19.

A disadvantage of series resistance R as in figure 8.19a is that the switch collector voltage at turn-off is increased from V_s to $V_s + I_m R$. The resistor must also have low self-inductance in order to allow the collector current to rapidly transfer from the switch to the resistor/diode reset circuit. The advantage of using a Zener diode as in figure 8.19b is that the maximum overvoltage is fixed, independent of the load

current magnitude. For a given maximum overvoltage, the Zener diode absorbs the inductor-stored energy quicker than would a resistor (see example 6.3 and problem 8.9). The advantages of using resistive dissipation are lower costs and more robust heat dissipation properties.

Alternatively the Zener diode can be placed across the switch as shown in figure 8.19c. The power dissipated is increased because of the energy drawn from the supply, through the inductor, during reset. At higher power, the soft voltage clamp shown in figure 8.19d, and considered in section 8.2, can be used. At switch turn-off, the energy stored in L_s, along with energy from the supply, is transferred and stored in a clamp capacitor. Simultaneously energy is dissipated in R and returned to the supply as the capacitor voltage rises. The advantage of this circuit is that the capacitor affords protection directly across the switch, but with lower loss than a Zener diode as in figure 8.19c. The energy loss equation for each circuit is also shown in figure 8.19. In high-voltage applications, the combined features of the soft clamp in figure 8.19d and the low loss Zener clamp in figure 8.19b can be realised by inserting a series Zener as shown in the figure 8.19d insert. This avoids the need to series connect Zener diodes, which would be necessary if the circuit in figure 8.19c were used at voltages above a few hundred volts.

Figure 8.20 shows how a switch turn-on snubber circuit modifies the SOA trajectory during switch-on, avoiding a condition of simultaneous maximum voltage V_s and current I_m.

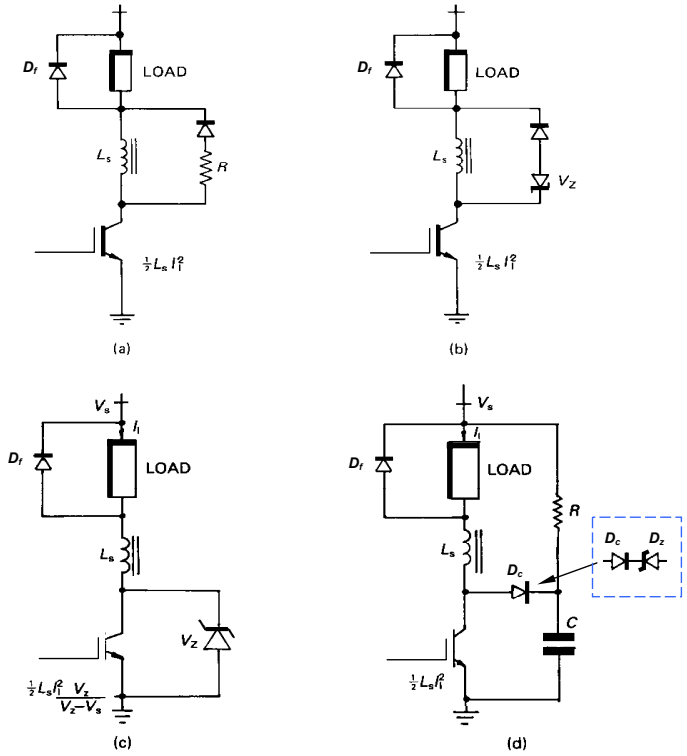


Figure 8.19. *Four turn-on snubber modifications for increasing the rate of release of inductor L_s stored energy: (a) using a power resistor; (b) using a power Zener diode; (c) parallel switch Zener diode, $V_Z > V_s$; and (d) using a soft voltage clamp.*

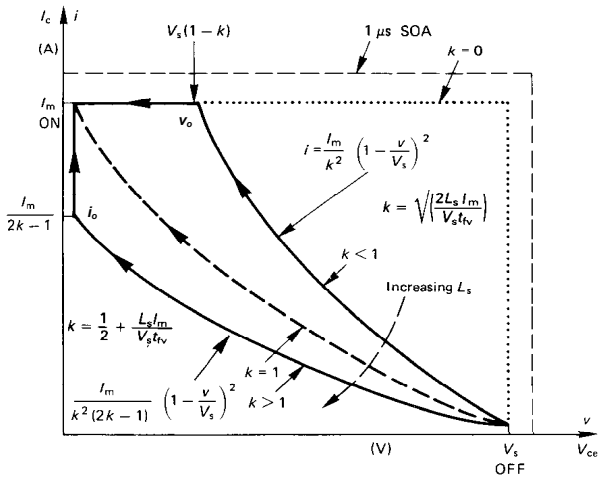


Figure 8.20. *The collector I-V trajectory at turn-on with a switching-aid circuit.*

Example 8.5: ***Turn-on air-core inductor snubber design***

A 600V, 100A machine field winding is switched at 10kHz. In maintaining a constant field current, the switch operates with an on-state duty cycle between 5% and 95% (5% ≤ δ ≤ 95%) and has a turn-on voltage fall time of 100ns, that is, $v_c(t) = 600\text{V}(1 - t/100\text{ns})$.

i. Estimate the turn-on loss of the switch.
ii. Design an inductive turn-on snubber using the dimensionally correct identity *v = Ldi/dt.* What is the current magnitude in the turn-on inductor when the switch voltage reaches zero.
iii. Design an inductive turn-on snubber such that the switch current reaches 100A at the same time the switch collector voltage reaches zero.

In each snubber case, using first a resistor and second a Zener diode for inductor reset, calculate the percentage decrease in switch power dissipation at turn-on, compared to the un-aided case.

Solution

i. The switch un-aided turn-on losses are given by equation (8.13). The turn-on time is greater than the voltage fall time (since the current rise time t_{ri} has been neglected), thus the turn-on switching losses will be greater than

$$W_{on} = \tfrac{1}{2} V_s I_m t_{on} = \tfrac{1}{2} \times 600\text{V} \times 100\text{A} \times 100\text{ns} = 3\text{mJ}$$
$$P_{on} = W_{on} \times f_s = 3\text{mJ} \times 10\text{kHz} = 30\text{W}$$

ii. Use of the equation *v = Ldi/dt* results in a switch current that reaches the load current magnitude after the collector voltage has fallen to zero. From $k = \tfrac{1}{2} + L_s I_m / V_s t_{fv}$ in figure 8.20, k = 3/2 satisfies the dimensionally correct inductor equation. Substitution into *v = Ldi/dt* gives the necessary snubber inductance

$$600\text{V} = L\frac{100\text{A}}{100\text{ns}}$$
$$\text{that is } L = 600\text{ nH}$$

The snubber inductor releases its stored energy at switch turn-off, and must discharge (demagnetise) during the switch minimum off-time, t_{off}. That is

$$t_{off} = 5L/R$$
$$5\% \text{ of } 1/10\text{kHz} = 5 \times 0.6\mu\text{H}/R$$
$$\text{that is } R = 0.6\,\Omega$$

Use the preferred value 0.68Ω (nearest higher preferred value), which reduces the *L/R* time constant.

The discharge resistor power rating is independent of resistance and is given by

$$P_{0.68\Omega} = \tfrac{1}{2} L I_m^2 f_s$$
$$= \tfrac{1}{2} \times 600\text{nH} \times 100\text{A}^2 \times 10\text{kHz} = 30\text{W}$$

The resistor in the circuit in figure 8.19a must have low inductance to minimise voltage overshoot at switch turn-off. Parallel connection of metal oxide resistors may be necessary to fulfil both resistance and power rating requirements. The maximum switch over-voltage at turn-off, (assuming zero resistor inductance), at the commencement of core reset, which is added to the supply voltage, 600V, is

$$V_{0.68\Omega} = I_m R = 100\text{A} \times 0.68\Omega = 68\text{V}$$

which decays exponential to zero volts in five time constants, 5µs. The maximum switch voltage is 600V + 68V = 668V, at turn-off. The reset resistor should be rated at 0.68Ω, 30W, metal film, 750V dc working voltage.

A Zener diode, as in figure 8.19b, of $V_z = L I_m / \breve{t}_{off} = 0.6\mu\text{H} \times 100\text{A}/5\mu\text{s} = 12\text{V}$, will reset the inductor in the same time as 5 *L/R* time constants. The switch voltage is clamped to 612V during the 5µs inductor reset time at switch turn-off.

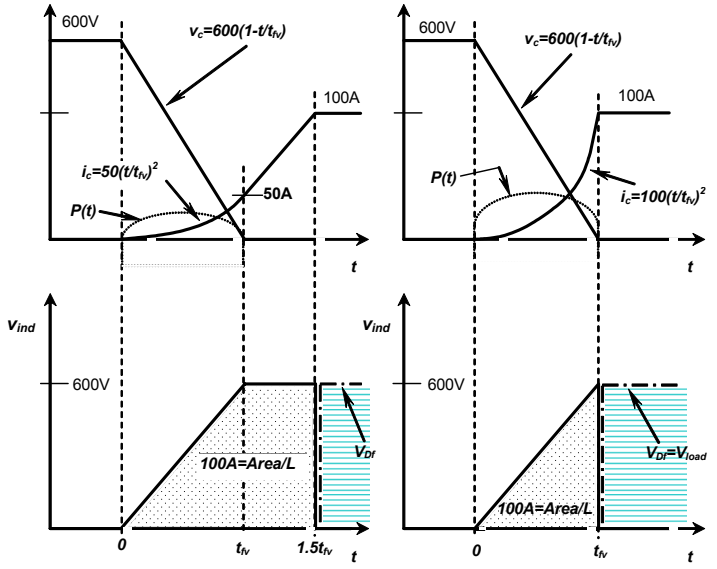
At turn-on when the switch voltage reduces to zero, the snubber inductor current (hence switch current) is less than the load current, 100A, specifically

$$i_0 = \frac{1}{L}\int v_{ind}\,dt$$
$$= \frac{1}{600\text{nH}} \int_0^{100\text{ns}} 600\text{V} \times \left(\frac{t}{100\text{ns}}\right) dt = 50\text{A}$$

The switch turn-on loss is reduced from 30W to

$$P_{on} = f_s \int_0^{100\text{ns}} i_c v_c\,dt = f_s \int_0^{100\text{ns}} V_s \left(1 - \frac{t}{100\,\text{ns}}\right) \times i_0 \left(\frac{t}{100\text{ns}}\right)^2 dt$$
$$= f_s \int_0^{100\text{ns}} 600\text{V}\left(1 - \frac{t}{100\,\text{ns}}\right) \times 50\text{A}\left(\frac{t}{100\text{ns}}\right)^2 dt = 2.5\text{W}$$

The total turn-on losses (switch plus snubber resistor) are 2.5W + 30W = 32.5W, which is more than the 30W for the unaided switch. Since the current rise time t_{ri} has been neglected in calculating the 30W un-aided turn-on losses, it would be expected that 32.5W would be less than the practical un-aided case. The switch loss is decreased by 92⅔%, from 30W down to 2.5W.



Example 8.5. part (b) *Example 8.5. part (c)*

iii. As the voltage across the switch falls linearly to zero from 600V, the series inductor voltage increases linearly to 600V (*k* = 1), such that the voltage sum of each component adds to 600V. The inductor current increases in a quadratic function according to

$$i_{ind}(t) = \frac{1}{L}\int v_{ind}\,dt$$

The inductor current increases quadratically to 100A in 100ns, as its voltage increases linearly from zero to 600V, that is

$$100\text{A} = \frac{1}{L} \int_0^{100\text{ns}} 600\text{V} \;\; t/100\text{ns}\, dt$$

that is $L = 300\text{nH}$

The necessary reset resistance to reduce the 300nH inductor current to zero in 5µs is

$$\breve{t}_{off} = 5\mu\text{s} = 5 \times 0.3\mu\text{H} / R$$

that is $R = 0.3\Omega$

Use the preferred value 0.33Ω in order to reduce the time constant.

The power dissipated in the 0.33Ω reset resistor, which is independent of resistance, is

$$P_{0.33\Omega} = \tfrac{1}{2} L I_m^2 f_s = \tfrac{1}{2} \times 300\text{nH} \times 100\text{A}^2 \times 10\text{kHz} = 15\text{W}$$

The resistance determines the voltage magnitude and the period over which the inductor energy is dissipated, not the amount of inductor energy to be dissipated. The inductor peak reset voltage is 100A×0.33Ω = 33V, which is added to the supply voltage of 600V, giving 633V across the switch at turn-off. That is, use a 0.33Ω, 15W metal film (for low inductance), 750V dc working voltage resistor.

A Zener diode, as in figure 8.19b, of $V_z = L I_m / \breve{t}_{off} = 0.3\mu\text{H} \times 100\text{A}/5\mu\text{s} = 6\text{V}$ (use 6.8V), will reset the inductor in the same time as 5 *L/R* time constants. The switch voltage is clamped to 606.8V during the $\breve{t}_{off} = 5\mu\text{s}$ inductor reset time at turn-off.

The energy dissipated in the switch at turn-on is reduced from 30W to

$$P_{on} = f_s \int_0^{100\text{ns}} i_c v_c\,dt = f_s \int_0^{100\text{ns}} V_s \left(1 - \frac{t}{100\,\text{ns}}\right) \times I_m \left(\frac{t}{100\,\text{ns}}\right)^2 dt$$
$$= f_s \int_0^{100\text{ns}} 600\text{V}\left(1 - \frac{t}{100\,\text{ns}}\right) \times 100\text{A}\left(\frac{t}{100\,\text{ns}}\right)^2 dt = 5\text{W}$$

The total turn-on snubber losses (switch plus snubber resistor) are 5W+15W = 20W, which is less than the 30W for the unaided switch. The switch losses, with an inductive turn-on snubber, are decreased by 83⅓%, from 30W to 5W.

♣

8.3.4 The polarised turn-on snubber circuit - with saturable ferrite inductance

The purpose of a turn-on snubber circuit is to allow the switch collector voltage to fall to zero while the collector current is low. Device turn-on losses are thus reduced, particularly for inductive loads, where during switching the locus point (V_s, I_m) occurs in the un-aided transition case.

This turn-on loss reduction effect can be achieved with a saturable inductor in the circuit shown in figure 8.21a, rather than using a non-saturable (air core) inductor as previously considered in section 8.3.3. The saturable inductor in the snubber circuit is designed to saturate after the collector voltage has fallen to zero, at point **Y** in figure 8.21. Before saturation the saturable inductor presents high reactance and only a low magnetising current flows. Once the collector voltage has reached zero, the inductance can saturate since the switch-on loss period is finished. From Faraday's equation, assuming the collector voltage fall to be linear, $V_s\left(1 - t/t_{fv}\right)$, the saturable inductor ℓ_s must satisfy

$$v_L = N\frac{d\phi}{dt} = NA\frac{dB}{dt} \qquad (8.27)$$

Rearranging, using an inductor voltage $v_L(t) = V_s - v_c(t) = V_s\, t / t_{fv}$, and integrating gives

$$B_s = \frac{1}{NA}\int_0^{t_{fv}} v_L(t)\,dt = \frac{1}{NA}\int_0^{t_{fv}} V_s \frac{t}{t_{fv}}\,dt \qquad (8.28)$$

which yields the identity

$$V_s = \frac{2NA\,B_s}{t_{fv}} \qquad \text{(V)} \qquad (8.29)$$

where *N* is the number of turns,
A is the core area, and
B_s is the core ferro-magnetic material saturation flux density.

The inductor magnetising current I_M should be much less than the load current magnitude I_m, $I_M << I_m$, and the magnetising current at saturation is given by

$$I_M = H_s L_{eff} / N \quad \text{(A)} \qquad (8.30)$$

where L_{eff} is the core effective flux path length and H_s is the magnetic flux intensity at the onset of saturation. Before core saturation the inductance is given by

$$L = N\Phi / I = N^2 / \Re = \mu_0 \mu_r A N^2 / L_{eff} \quad \text{(H)} \qquad (8.31)$$

When the core saturates the inductance falls to that of an air core inductor ($\mu_r = 1$) of the same turns and dimensions, that is, the incremental inductance is

$$L_{sat} = \mu_0 A N^2 / L_{eff} \quad \text{(H)} \qquad (8.32)$$

The energy stored in the inductor core is related to the *B-H* area shown in figure 8.21c and magnetic volume, and is approximated by

$$W_{\ell_s} = \tfrac{1}{2} B_s H_s A L_{eff} = \tfrac{1}{2} L I_M^2 \quad \text{(J)} \qquad (8.33)$$

The collector turn-on waveforms are shown in figure 8.21b, while the corresponding *B-H* curve and SOA trajectories are illustrated in figure 8.21 parts c and d. It will be seen in figure 8.21b that little device turn-on electrical stressing occurs.

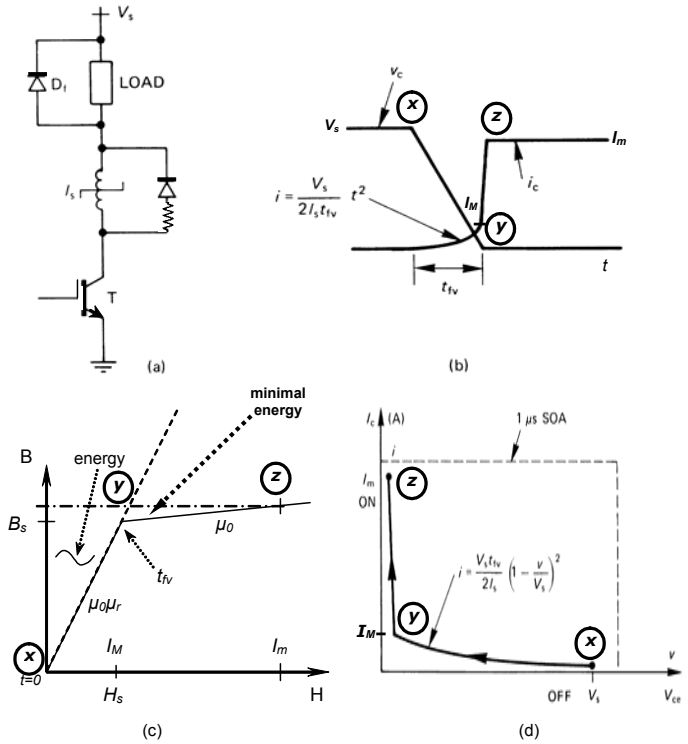


Figure 8.21. *Switch turn-on characteristics when a saturable inductor is used in the turn-on snubber: (a) circuit diagram; (b) collector voltage and current waveforms; (c) magnetic core B-H curve trajectory; and (d) safe operating area I-V turn-on trajectory.*

Example 8.6: ***Turn-on ferrite-core inductor snubber design***

A 600V, 100A machine field winding is switched at 10kHz. In maintaining the field current constant, the switch operates with an on-state duty cycle between 5% and 95% (5% ≤ δ ≤ 95%) and has a turn-on voltage fall time of t_{fv} = 100ns, that is, $v_c(t) = 600\text{V}(1 - t/100\text{ns})$.

i. Design a saturable inductor turn-on snubber that saturates as the collector voltage reaches zero, using a ferrite core with the following parameters.
- A = 0.4 sq cm
- L = 4cm
- B_s = 0.4T
- H_s = 100At/m

ii. Calculate the switch losses at turn-on when using the saturable reactor. What is the percentage reduction in switch turn-on losses?

iii. If an air cored inductor is used to give the same switch turn-on loss, what are the losses at reset?

Solution

From *example 8.5*, the unaided switch turn-on loss is 30W.

i. From equation (8.29) the number of core turns is

$$N = \tfrac{1}{2} V_s t_{fv} / A B_s$$
$$= \tfrac{1}{2} \times 600\text{V} \times 0.1\mu\text{s} / 0.4 \times 10^{-4} \times 0.4\text{T} \approx 2 \text{ turns}$$

The magnetising current I_M at saturation, that is, when the collector voltages reaches zero, is given by equation (8.30)

$$I_M = H_s L_{eff} / N$$
$$= 100\text{At/m} \times 0.04 / 2 = 2\text{A}$$

Since $I_M << I_m$, (2A<<100A), this core with 2 turns produces satisfactory turn-on snubber action, resulting in greatly reduced switch losses at turn-on.

From equation (8.31) the inductance before saturation is

$$L = NAB_s / I_M$$
$$= 2 \times 0.4 \times 10^{-4} \times 0.4\text{T} / 2\text{A} = 16\mu\text{H}$$

The incremental inductance after saturation, from equation (8.32), is given by

$$L_{sat} = N^2 A \mu_0 / L_{eff}$$
$$= 2^2 \times 0.4 \times 10^{-4} \times 4\pi \times 10^{-7} / 0.04 = 50\text{nH}$$

From equation (8.33) the energy stored in the core and released as heat in the reset resistor is

$$W_L = \tfrac{1}{2} L I_M^2 \quad \left(= \tfrac{1}{2} B_s H_s L_{eff} A\right)$$
$$= \tfrac{1}{2} \times 16\mu\text{H} \times 2^2 = 32\mu\text{J}$$
$$P_L = W_L \times f_s = 32\mu\text{J} \times 10\text{kHz} = 0.32\text{W}$$

The time $\check{t}_{off}$ for core reset via the resistor in five *L/R* time constants, is dominated by the 16μH section (the pre-saturation section) of the *B-H* curve, thus

$$\check{t}_{off} = 5\mu\text{s} = 5 \times 16\mu\text{H} / R$$

that is $R = 16\Omega$

Use a 18Ω, 1W, carbon composition resistor, for low inductance.
This resistance results in a switch voltage increase above 600V of 18Ω×100A=1800V at turn-off. This high-voltage may be impractical in terms of the switch and resistor voltage ratings.

Alternatively, the Zener diode clamps shown in figures 8.19 b or c, may be suitable to dissipate the 0.32W of stored magnetic energy. The Zener voltage is determined by assuming that a fixed Zener voltage results in a linear decrease in current from 2A to zero in 5μs. That is, assuming minimal stored energy associated with the current decrease from 100A to 2A,

$$W_L = V_Z \int_0^{\check{t}_{off}} i_{ind} dt \quad \left(= \tfrac{1}{2} L I_M^2\right)$$
$$32\mu\text{J} = \tfrac{1}{2} \times V_Z \times 2\text{A} \times 5\mu\text{s}$$

that is $V_Z = 6.4\text{V}$

Use a 6.8V, 1W Zener diode to clamp and dissipate the 0.32W of reset power.
Series connected Zener diodes in parallel with the switch, as in figure 8.19c, dissipate 30W.

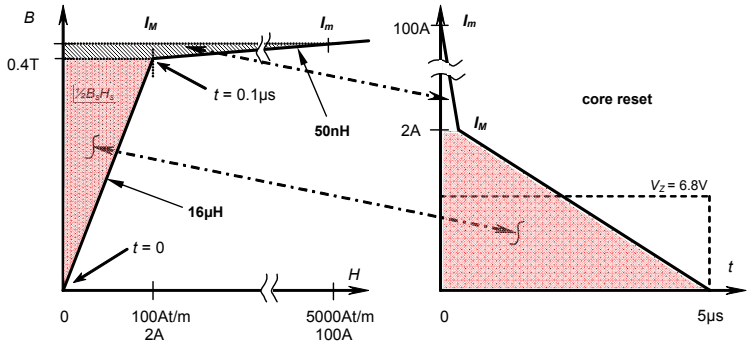
The energy associated with saturation is small and is released in an insignificant time compared to the 5μs minimum off-time. The advantage of the Zener diode clamping approach, as opposed to using a resistor, is that the maximum switch voltage is clamped to 606.8V, even during the short, low energy period when the inductor current falls from 100A to 2A.

ii. The switch turn-on losses with the saturable reactor are given by

$$P_{on} = f_s \int_0^{100\text{ns}} i_c v_c dt = f_s \int_0^{100\text{ns}} V_s \left(1 - \frac{t}{100\text{ns}}\right) \times I_M \left(\frac{t}{100\text{ns}}\right)^2 dt$$

$$= f_s \int_0^{100\text{ns}} 600\text{V} \left(1 - \frac{t}{100\text{ns}}\right) \times 2\text{A} \left(\frac{t}{100\text{ns}}\right)^2 dt = 0.1\text{W}$$

The switch losses at turn-on have been reduced from 30W to 0.1W, a 99⅔% decrease in losses. The total losses (switch plus Zener diode) are 0.1W + 0.32W = 0.42W, which is significantly less than the 30W in the un-aided case.



Example 8.6

iii. If an air core inductor of 16μH (from part i) were to replace the saturable reactor, the stored energy released would give losses

$$W = ½LI_m^2$$

$$= ½ \times 16\mu\text{H} \times 100^2 = 80\text{mJ}$$

$$P = W \times f_s = 80\text{mJ} \times 10\text{kHz} = 800\text{W}$$

Clearly the use of an air cored inductor rather than a saturable reactor, to achieve the same switch loss of 0.1W at turn-on, is impractical.

♣

8.3.5 The unified turn-on and turn-off snubber circuit

Figure 8.22 shows a switching circuit which incorporates both an inductor turn-on and a capacitor turn-off snubber circuit. Both C_s and L_s are dimensioned by the analysis outlined in sections 8.3.1 and 8.3.3, respectively. The power rating of the dissipating resistor R incorporates a contribution from both the turn-on inductor L_s and turn-off capacitor C_s, according to

$$P_{R_s} = ½\left(L_s I_m^2 + C_s V_s^2\right) f_s \qquad \text{(W)} \qquad (8.34)$$

Calculated resistance values to satisfy both minimum off and on time reset according to $\check{t}_{on} \geq 5R_sC_s$ and $\check{t}_{off} \geq 5L_s/R_s$, may result in irreconcilable resistance and/or switch voltage/current requirements. The snubber capacitor discharges at turn-on via an *L-C-R* circuit rather than the usual *R-C* circuit, hence reducing the turn-on current stressing of the switch.

In example 8.4 the resistor requirement for the 16⅔nF capacitive turn-off snubber is $R < 56\Omega$, while the 0.6μH inductive turn-on snubber in example 8.5 requires $R > 0.68\ \Omega$. Thus $0.68\Omega < R < 56\Omega$ satisfies snubber L_s and C_s reset requirements. The maximum reset current and voltage are determined by $Z = \sqrt{L/C}$. In combining the two snubber functions, the single resistor may reduce the maximum switch over-voltage at turn-off, $I_m Z$, and the maximum switch snubber current at turn-on V_s/Z. If R is too small a high switch snubber current V_s/Z flows at turn-on, while if R is too large, a large switch over-voltage, $I_m Z$, occurs at switch turn-off.
An important by-product from using a turn-on snubber circuit is that the inductor controls the reverse recovery process of the load freewheeling bipolar diode at switch turn-on.

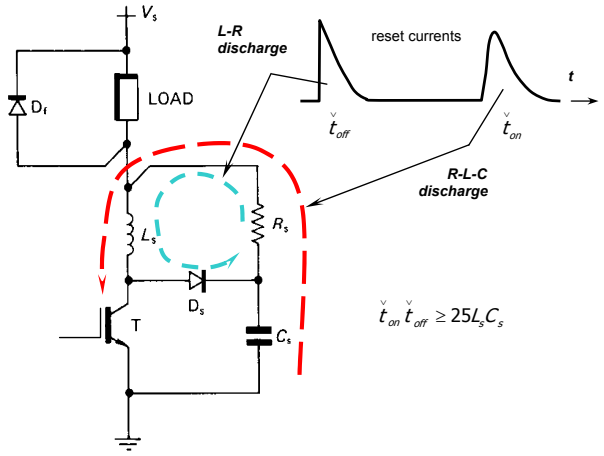


Figure 8.22. *Unified snubber incorporating both a turn-on and a turn-off circuit which share the one dissipation reset resistor.*

8.4 Snubbers for bridge legs

Figure 8.23 parts a to c show three typical switch bridge leg configurations used in inverters as shown in figures 14.1 and 14.3. The polarised inductive turn-on snubber L_s and capacitive turn-off snubber C_s are incorporated into the bridge legs as shown in each circuit in figure 8.23.
The combinational polarised snubber circuit in figure 8.23a can be used to minimise the number of snubber components. The turn-on snubber inductance L_s, reset resistor R, and snubber capacitor C_{sc}, are common to any number of bridge legs. The major disadvantage of this circuit is that turn-off snubber action associated with the lower switch is indirect, relying on low inductance decoupling through C_s and C_{sc}.

With an inductive load, unwanted turn-off snubber action occurs during the switch modulation sequence as shown in figures 8.23b and 8.23c. When the upper switch T_u is turned off as in figure 8.23b the load current I_m is diverted to the freewheel diode D_f. While D_f conducts, the capacitor C_s discharges to zero through the resistor R, as shown, dissipating energy $½C_sV_s^2$. When the switch T_u is turned on, the load current is provided via the switch T_u and the snubber capacitor C_s is charged through the series turn-on snubber inductance, as shown in figure 8.23c. A lightly damped *L-C* oscillation occurs and C_s is over charged. Advantageously, the recovery voltage of the freewheel diode D_f is controlled by the capacitor voltage rise.
The unwanted snubber action across the non-power conducting switch can be avoided in some applications by using a series blocking diode as shown in figure 8.23d. The diode D_b prevents C_s from discharging into the load as occurs with the lower switch in figure 8.23b. A blocking diode can be used to effectively disable the internal parasitic diode of the MOSFET. Adversely, the blocking diode increases the on-state losses.
In reactive load applications, bridge legs are operated with one switch on, with only a short underlap when both switches are off. Thus although the snubber capacitor cannot discharge into the load in figure 8.23d, it always discharges through the switch T, regardless of load current flow through the switch.

In IGBT and MOSFET applications, the conventional *R-C-D* turn-off snubber is not usually required. But because of diode recovery limitations, a turn-on snubber may be necessary. In low frequency applications, a single turn-on snubber inductor can be used in the dc link as shown in figure 8.24a. Snubber circuit design is based on the turn-on snubber presented in 8.3.3. The circuit in figure 8.24b is based on the conventional turn-on snubber being incorporated within the bridge leg. Figures 8.24c and d show turn-on snubbers which use the soft voltage clamp, presented in 8.2, to reset the snubber inductor current to zero at turn-off.

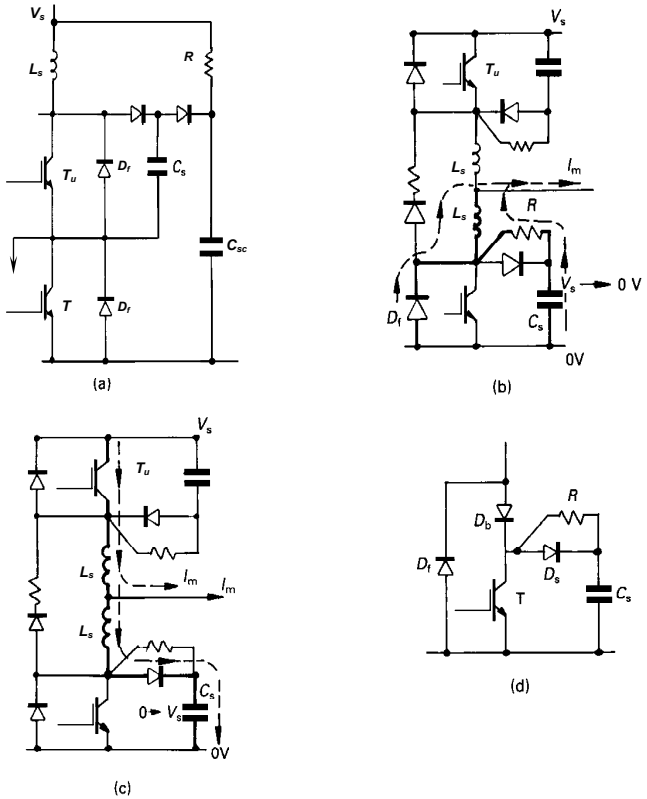


Figure 8.23. *Bridge leg configurations: (a) Undeland indirect leg snubber circuit; (b) leg with turn-on snubber and turn-off snubber C_s discharge path shown; (c) L-C oscillation at switch-on; and (d) blocking circuit to prevent snubber capacitor discharge when D_f conducts.*

In each circuit at switch turn-off, t_3, the energy $\frac{1}{2}LI_m^2$ stored in the turn-on snubber inductor is dissipated in the resistor of the discharge circuit. The energy $\frac{1}{2}LI_{rm}^2$ in *L*, due to diode recovery, is dissipated in the resistor at time t_1, in circuits (a), (b) and (c). In figure 8.24d the energy in excess of that associated with the load, $\frac{1}{2}LI_m^2$, due to diode recovery, is dissipated in the switch and its parallel connected diode. At time t_1

$$W = \tfrac{1}{2}LI_{rm}^2 + \frac{V_{ce}}{V_{ce}+V_{Df}} LI_{rm}I_m \tag{8.35}$$

is dissipated in the two semiconductor components. Since the energy is released into a low voltage v_{ce} + v_{Df}, the reset time $t_2 - t_1$ is large.
Magnetic coupling of the inductors in figures 8.24c and d does not result in any net energy savings.

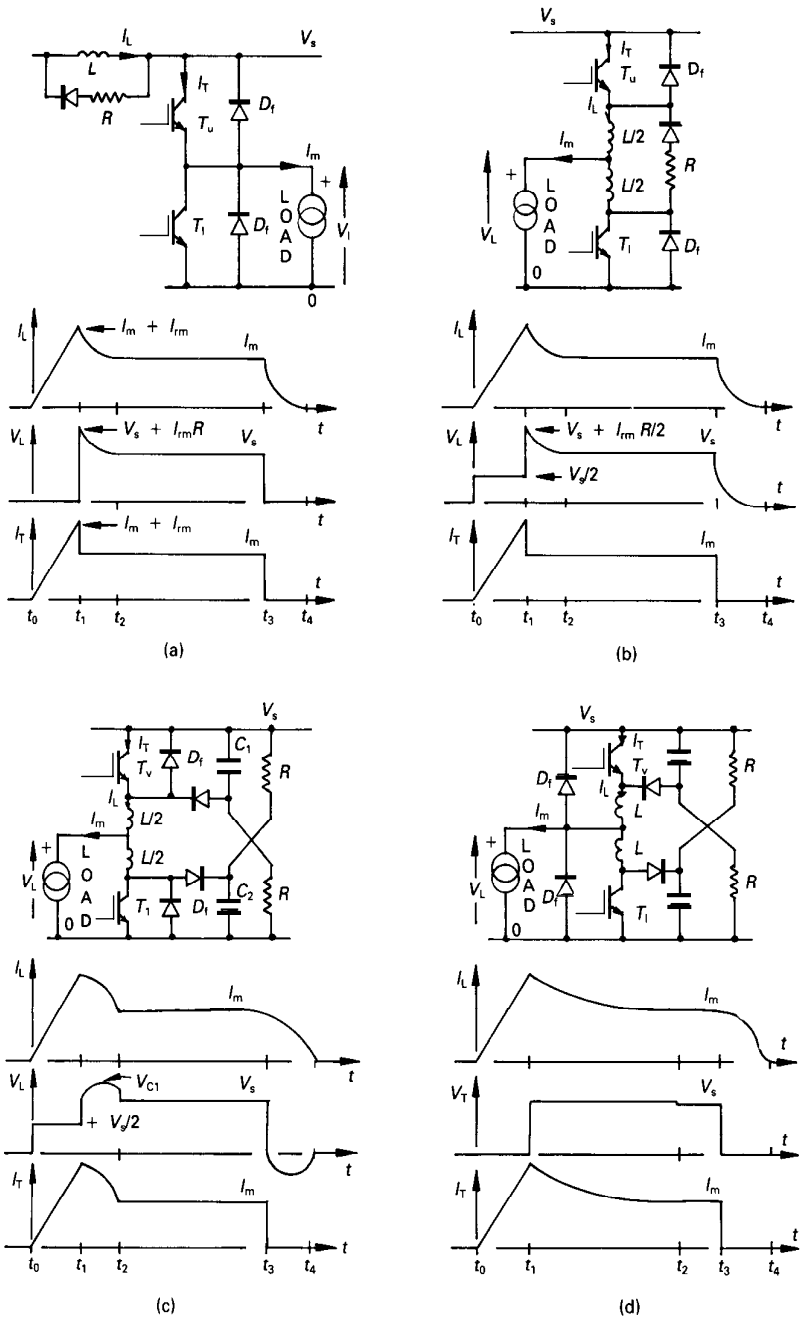


Figure 8.24. *Turn-on snubbers for bridge legs: (a) single inductor in dc link; (b) unified L-R-D snubber; (c) soft voltage clamp; and (d) soft voltage clamp with load clamped.*

8.5 Appendix: Non-polarised turn-off *R-C* snubber circuit analysis

When a step input voltage is applied to the *L-C-R* circuit in figure 8.3, a ramped voltage appears across the *R-C* part of the circuit. If this *dv/dt* is too large, a thyristor in the off-state will turn on as a result of the induced central junction displacement current, which causes injection from the outer junctions.

The differential equations describing circuit current operation are

$$(D^2 + 2\xi\omega_0 D + \omega_0^2)\, I = 0 \qquad (8.36)$$

and

$$(\tau D + 1) I = C D e_0 \qquad (8.37)$$

where D = differential operator = d/dt
and ξ = damping ratio = $\frac{1}{2}R\sqrt{C/L}$
ω_0 = natural frequency = $1/\sqrt{LC}$
ω = oscillation frequency = $\omega_0\sqrt{1-\xi^2}$

Solution of equations (8.36) and (8.37), for initial current $I_o = 0$, leads to

(a) The snubber current

$$I(t) = \frac{e_s}{R}\frac{2\xi}{\sqrt{1-\xi^2}}\, e^{-\xi\omega_0 t}\sin\omega t \quad \text{(A)} \qquad (8.38)$$

(b) The rate of change of snubber current

$$\frac{dI}{dt} = \frac{e_s}{L}\, e^{-\xi\omega_0 t}\left(\cos\omega t - \frac{\xi}{\sqrt{1-\xi^2}}\sin\omega t\right) \quad \text{(A/s)} \qquad (8.39)$$

(c) Snubber *R-C* voltage

$$e_0 = e_s\left(1 - e^{-\xi\omega_0 t}\left\{\cos\omega t - \frac{\xi}{\sqrt{1-\xi^2}}\sin\omega t\right\}\right) \quad \text{(V)} \qquad (8.40)$$

(d) The rate of change of *R-C* voltage

$$\frac{de_0}{dt} = \omega_0 e_s\, e^{-\xi\omega_0 t}\left(2\xi\cos\omega t + \frac{1-2\xi^2}{\sqrt{1-\xi^2}}\sin\omega t\right) \quad \text{(V/s)} \qquad (8.41)$$

The maximum value expressions for each equation can be found by differentiation

(a) Maximum snubber current

$$I_p = \frac{e_s}{R}\, 2\xi\, e^{\left(-\xi\cos^{-1}\xi/\sqrt{1-\xi^2}\right)} \quad \text{(A)} \qquad (8.42)$$

when cos $\omega t = \xi$

(b) The maximum snubber *di/dt* is given by

$$\frac{dI_p}{dt} = \frac{e_s - e_0}{L} \quad \text{(A/s)} \qquad (8.43)$$

(c) Maximum *R-C* voltage

$$\hat{e}_o = e_s\left(1 + e^{\left(-\xi\cos^{-1}\left(2\xi^2-1\right)/\sqrt{1-\xi^2}\right)}\right) \quad \text{(V)} \qquad (8.44)$$

when cos $\omega t = 2\xi^2 - 1$

(d) Maximum slew rate, $\frac{de_0}{dt} = \hat{S}$

for $\xi < \frac{1}{2}$

$$\hat{S} = e_0\omega_0 e^{\left\{-\xi\cos^{-1}\xi\left(3-4\xi^2\right)/\sqrt{1-\xi^2}\right\}} \quad \text{(V/s)} \qquad (8.45)$$

when cos $\omega = \xi(3-4\xi^2)$. The minimum value of the maximum slew rate is 0.81 pu at $\xi = 0.265$.

at $\xi = \frac{1}{2}$, $\hat{S} = \omega_o e_s$.

for $\xi > \frac{1}{2}$

$$\hat{S} = 2\xi e_s w_0 \quad (= e_s R/L) \quad \text{(V/s)} \qquad (8.46)$$

when $t = 0$

Equations (8.42) to (8.46), after normalisation are shown plotted in figure 8.4 as a function of the snubber circuit damping factor ξ. The power dissipated in the resistor is approximately $C e_s^2 f_s$.

8.6 Appendix: Polarised turn-off *R-C-D* switching aid circuit analysis

Switch turn-off loss for an unaided switch, assuming the collector voltage rise time is negligible compared with the collector current fall time, is

$$W = \tfrac{1}{2} V_s I_m t_{fi} \qquad (8.47)$$

If τ is the time in figure 8.11 for the snubber capacitor C_s to charge to the supply V_s, and t_{fi} is the switch collector current fall time, assumed linear such that $i_c(t) = I_m(1 - t/t_{fi})$, then two capacitor charging conditions can exist

- $\tau \le t_{fi}$
- $\tau \ge t_{fi}$

Let $k = \tau/t_{fi}$ and electrical energy $W = \int_0^t vi\, dt$

Case 1: $\tau \le t_{fi}$, $k \le 1$

Figure 8.11b shows ideal collector voltage and current waveforms during aided turn-off for the condition $t \le t_{fi}$. If, assuming constant maximum load current, I_m, during the switching interval, the collector current falls linearly, then the load deficit, $I_m t/t_{fi}$, charges the capacitor C_s. From $i = C dv/dt$, the capacitor voltage, and collector voltage, therefore increase quadratically. The collector voltage v_c and current i_c are given by

$$\begin{bmatrix} i_c(t) = I_m\left(1 - \dfrac{t}{t_{fi}}\right) \\ v_c(t) = V_s\left(\dfrac{t}{\tau}\right)^2 \end{bmatrix},\ 0 \le t \le \tau \quad \text{and} \quad \begin{bmatrix} i_c(t) = I_m\left(1 - \dfrac{t}{t_{fi}}\right) \\ v_c(t) = V_s \end{bmatrix},\ \tau \le t \le t_{fi} \qquad (8.48)$$

The final capacitor charge is given by

$$Q = C_s V_s = \int_0^{\tau} (I_m - i_c(t))dt = \tfrac{1}{2} I_m t_{fi} k^2 \quad \text{(C)} \qquad (8.49)$$

The energy stored by the capacitor, W_c, and energy dissipated in the switch, W_t, are given by

$$\begin{aligned} W_c &= \tfrac{1}{2} C_s V_s^2 \quad (= \tfrac{1}{2} Q V_s) \\ &= \tfrac{1}{2} V_s I_m t_{fi} \times \tfrac{1}{2} k^2 \quad \text{(J)} \end{aligned} \qquad (8.50)$$

$$\begin{aligned} W_t &= \int_0^{\tau} V_s I_m (t/\tau)dt + \int_0^{t_{fi}} V_s I_m (1 - t/t_{fi})dt \\ &= \tfrac{1}{2} V_s I_m t_{fi}\left(1 - \tfrac{1}{3}k + \tfrac{1}{2}k^2\right) \quad \text{(J)} \end{aligned} \qquad (8.51)$$

The total circuit losses W_{total}, are

$$W_{total} = W_t + W_c = \tfrac{1}{2} V_s I_m t_{fi} \times \left(1 - \tfrac{1}{3}k + \tfrac{1}{2}k^2\right),\ k \le 1 \quad \text{(J)} \qquad (8.52)$$

Case 2: $\tau \ge t_{fi}$, $k \ge 1$

Figure 8.11c shows the ideal collector voltage and current switch-off waveforms for the case when $k \ge 1$. When the collector current falls to zero the snubber capacitor has charged to a voltage, v_0, where

$$\begin{aligned} v_o &= \frac{1}{C_s}\int_0^{t_{fi}} i\, dt \\ &= \frac{1}{C_s} \times \tfrac{1}{2} I_m t_{fi} \quad \text{(V)} \end{aligned} \qquad (8.53)$$

The collector voltage v_c and current i_c are given by

$$\begin{bmatrix} i_c(t) = I_m\left(1 - \dfrac{t}{t_{fi}}\right) \\ v_c(t) = v_o\left(\dfrac{t}{t_{fi}}\right)^2 \end{bmatrix},\ 0 \le t \le t_{fi} \quad \text{and} \quad \begin{bmatrix} i_c(t) = 0 \\ v_c(t) = \dfrac{1}{t_{fi}}\dfrac{(V_s - v_0)t}{k-1} + \dfrac{kv_0 - V_s}{k-1} \end{bmatrix},\ t_{fi} \le t \le \tau \qquad (8.54)$$

The final capacitor charge is given by

$$Q = C_s V_s = \int_0^{t_{fi}} (I_m - i_c(t))dt + \int_{t_{fi}}^{\tau} I_m dt = I_m t_{fi}\left(k - \tfrac{1}{2}\right) \quad \text{(C)} \tag{8.55}$$

The energy stored by the capacitor W_c, and energy dissipated in the switch W_t, are given by

$$\begin{aligned} W_c &= \tfrac{1}{2} C_s V_s^2 \quad (= \tfrac{1}{2} Q V_s) \\ &= \tfrac{1}{2} V_s I_m t_{fi} \times \left(k - \tfrac{1}{2}\right) \quad \text{(J)} \\ W_t &= \int_0^{t_{fi}} v_0 I_m (1 - t/t_{fi})\left(t/t_{fi}\right)^2 dt \\ &= \tfrac{1}{12} v_0 I_m t_{fi} \quad \text{(J)} \end{aligned} \tag{8.56}$$

Using equations (8.53) and (8.55) to eliminate v_0 yields

$$W_t = \tfrac{1}{2} V_s I_m t_{fi} \times \frac{1}{6(2k-1)} \quad \text{(J)} \tag{8.57}$$

The total circuit losses W_{total}, are

$$W_{total} = W_t + W_c = \tfrac{1}{2} V_s I_m t_{fi} \times \frac{\left(k^2 - k + \tfrac{1}{3}\right)}{\left(k - \tfrac{1}{2}\right)}, \quad k \geq 1 \quad \text{(J)} \tag{8.58}$$

The equations (8.50) to (8.52), and (8.56) to (8.58) have been plotted, normalised with respect to unaided losses $\tfrac{1}{2} V_s I_m t_{fi}$, in figure 8.13.

Reading list

International Rectifier, *HEXFET Data Book,* HDB-5, 1987.

Peter, J. M., *The Power Transistor in its Environment,* Thomson-CSF, Sescosem, 1978.

Siliconix Inc., *Mospower Design Catalog,* January 1983.

Grafham, D. R. *et al., SCR Manual,* General Electric Company, 6th Edition, 1979.

Problems

8.1. The figure 8.25 shows GTO thyristor turn-off anode *I-V* characteristics. Calculate

i. turn-off power loss at 1 kHz. What percentage of the total loss does the tail current account for?

ii. losses when a capacitive turn-off snubber is used and the anode voltage rises quadratically to 600V in 0.5µs. What percentage of the total losses does the tail current account for? What is the necessary capacitance?

iii. losses when a capacitive turn-off snubber is used and the anode voltage rises quadratically to 600V in 2µs. What percentage of the total losses does the tail current account for? What is the necessary capacitance?

[10.5 W]

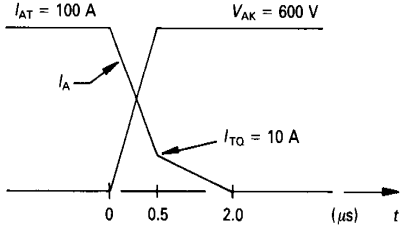


Figure 8.25. *Problem 8.1, GTO thyristor tail current characteristics.*

8.2. Prove that the minimum total losses (switch plus snubber resistor), associated with a switch which utilises a capacitive turn-off switching-aid circuit, occur if the snubber capacitor is fully charged when the collector current has fallen to ⅓ its original value. Derive an expression for this optimal snubber capacitance.

$$C_s = \tfrac{2}{9} \frac{I_m t_{fi}}{V_s} \quad \text{(F)}$$

8.3. Derive an expression for the optimal turn-on switching-aid circuit inductance, assuming the collector current rise time in the unaided circuit is very short compared with the collector voltage fall time.

$$L_s = \tfrac{2}{9} \frac{V_s t_{fv}}{I_m} \quad \text{(H)}$$

8.4. A ferrite toroid has *B-H* characteristics as shown in figure 8.26 and a cross-sectional area, *A,* of 10 mm^2 and effective length, L_{eff} of 50 mm.

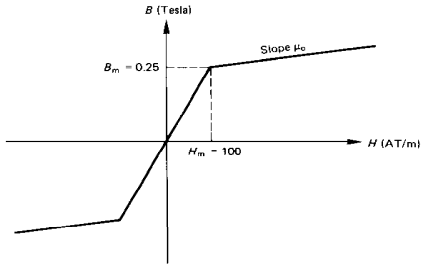


Figure 8.26. *Problem 8.4, B-H characteristics.*

A number of such toroid cores are to be stacked to form a core for a saturable inductor turn-on snubber in a switching circuit. The circuit supply voltage is *V* and the switch voltage fall time at turn-on is t_{fv}. Assume t_{fv} is independent of supply voltage and falls linearly from *V* to 0 V.

i. Using Faraday's Law, show that if the ferrite inductor is to saturate just as the switch collector voltage falls to zero at turn-on, then the number of turns *N* for *n* cores is given by

$$N = \frac{V t_{fv}}{2 B_m A n}$$

ii. Derive an expression for the inductance before saturation.

iii. It is required that the maximum magnetising current before saturation does not exceed 1 A. If only 10 turns can be accommodated through the core window, what is the minimum number of cores required if *V* = 200 V and t_{fv} = 1 µs?

iv. How many cores are required if the supply *V* is increased to the peak voltage of the three-phase rectified 415 V ac mains, and the load power requirements are the same as in part iii?

v. Calculate the percentage change in the non-saturated inductance between parts iii and iv.

vi. What are the advantages of saturable inductance over linear non-saturable inductance in turn-on snubber applications? What happens to the inductance and stored energy after saturation?

[$\ell = N^2/R$, n = 5, n = 4, 1:9]

8.5. Prove, for an inductive turn-on snubber, where the voltage fall is assumed linear with time, that

$$k = \sqrt{\frac{2 L_s I_m}{V_s t_{fv}}} \quad \text{for} \quad k \geq 1$$

$$k = \frac{L_s I_m}{V_s t_{fv}} + \tfrac{1}{2} \quad \text{for} \quad k \leq 1$$

where $k = t_{fv}/\tau$ (see figures 8.17 and 8.20).

8.6. Derive the expressions in table 8.1 for a turn-off snubber assuming a cosinusoidal current fall. Prove equation (8.21), the optimal capacitance value.

8.7 Show that when designing a capacitive turn-off snubber using the dimensionally correct equation $i = Cdv/dt$, as in example 8.4b, the capacitor charges to $½V_s$ when the switch current reaches zero.

8.8 Show that when designing an inductive turn-on snubber using the dimensionally correct equation $v = Ldi/dt$, as in example 8.5b, the inductor current reaches $½I_m$ when the switch voltage reaches zero.

8.9 Reset of inductive turn-on snubber energy $½L_s I_m^2$ can be affected through a resistor, R, as in figure 8.19a or through a Zener diode, D_z, as in figure 8.19b.
Show that for the same reset voltage, namely $V_z=I_mR$, in each case, Zener diode reset is n times faster the resistor reset when $nR_sC_s \le \breve{t}_{on}$.

8.10 In figure 8.12a show that for $\tau < t_{fi}$, the collector current is given by

$$i_o = I_m - \sqrt{\frac{2\,I_m V_s C_s}{t_{fi}}}$$

when the collector voltage reaches the supply voltage rail V_s.

8.11 In figure 8.18a show that for $\tau < t_{fv}$, the collector voltage is given by

$$v_o = V_s - \sqrt{\frac{2\,I_m V_s L_s}{t_{fv}}}$$

when the collector current reaches the load current I_m.

8.12 An RCD turn-off snubber is used across a switch in a 600V dc, 10A, 20kHz chopper application. The switch current fall time at 10A is 100ns.

i. What is the capacitor voltage when the switch current reaches zero at turn-off, if the switch turn-off loss is to be 1W?
ii. What snubber capacitance is necessary?

For the same un-aided switching conditions, the total losses (switch plus snubber resistor) are to be (a) 10/3W
(b) 9W
What is the capacitor requirement and what is the capacitor voltage when the switch current reaches zero?
[600V, 0.83nF]

8.13 For a cosinusoidal current fall at turn-off as shown in figure 8.15, derive expressions for the switch current i_o when $k < 1$ and collector voltage v_o when $k > 1$.

Blank

9

Switching-aid Circuits with Energy Recovery

Passive turn-on and turn-off snubber circuits for the IGBT transistor, the GCT and the GTO thyristor have been considered in chapter 7. These snubber circuits modify the device *I-V* switching trajectory and in so doing reduce the device transient losses. Snubber circuit action involves temporary energy stored in either an inductor or capacitor. In resetting these passive components it is usual to dissipate the stored energy in a resistor as heat. At high frequencies these losses (being proportional to frequency) may become a limiting factor because of the difficulties associated with equipment cooling. Instead of dissipating the switching-aid circuit stored energy, it may be viable to recover the energy back into the dc supply or into the load, or both. Two classifications of energy recovery circuits exist, either passive or active. A *passive recovery* circuit involves only passive components such as *L* and *C* while *active recovery* techniques involve extra switching devices, as in a switched-mode power supply, smps.

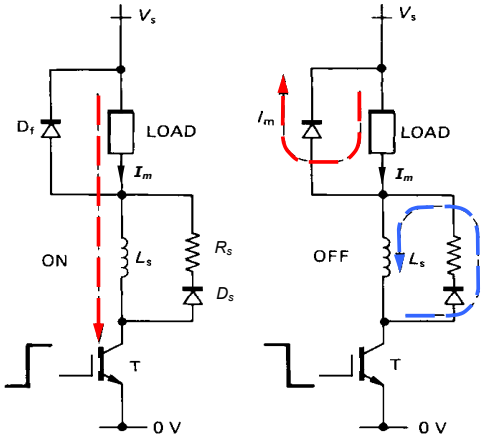


Figure 9.1. *Conventional inductive turn-on snubber principal currents at: (a) turn-on and (b) turn-off.*

9.1 Energy recovery for inductive turn-on snubber circuits – single ended

Figure 9.1 shows the conventional inductive turn-on snubber circuit for a single-ended IGBT transistor switching circuit. Equally the switch may be a GCT or a GTO thyristor, for which an inductive turn-on snubber is mandatory, if switch derating is to be avoided.
At switch turn-on the snubber inductance controls the rate of rise of current as the collector voltage falls to zero. The switch turns on without the stressful condition of simultaneous maximum voltage and

current (V_s, I_m) being experienced. At turn-off the inductor current is diverted through the diode D_s and resistor R_s network and the stored inductor energy $\frac{1}{2}LI_m^2$ is dissipated as heat in the resistance of the L_s-R_s-D_s circuit. The power loss is determined by the switching frequency and is given by $\frac{1}{2}LI_m^2 f_s$. Full design and operational aspects of this turn-on snubber have been considered in chapter 8.3.3.

9.1.1 Passive recovery

i. Recovery into the dc supply

Figure 9.2 shows a magnetic coupled circuit technique for passively recovering the inductive turn-on snubber stored energy back into the dc supply V_s. The inductor is bifilar-wound with a catch winding. The primary winding is designed to give the required (magnetising) inductance based on core dimensions, properties, and number of turns, $L = N^2/R$. At switch turn-off the current in the coupled inductor primary is diverted to the secondary so as to maintain core flux. The windings are arranged to transfer current back into the supply via a diode D_R which prevents reverse current flow.
The operating principles of this turn-on snubber recovery scheme are simple but a number of important circuit characteristics are exhibited. Let the coupled inductor have a primary-to-secondary turns ratio of 1:*N*. At turn-off the catch (secondary) winding conducts and its voltage is thereby clamped to the supply rail V_s. The primary winding therefore has an induced voltage specified by the turns ratio. That is

$$V_{(p} = \frac{1}{N} V_s \quad \text{(V)} \tag{9.1}$$

The switch collector voltage at turn-off is increased by this component, to

$$V_c = \left(1 + \frac{1}{N}\right) V_s \quad \text{(V)} \tag{9.2}$$

The turns ratio *N* should be large so as to minimise the switch voltage rating in excess of V_s.

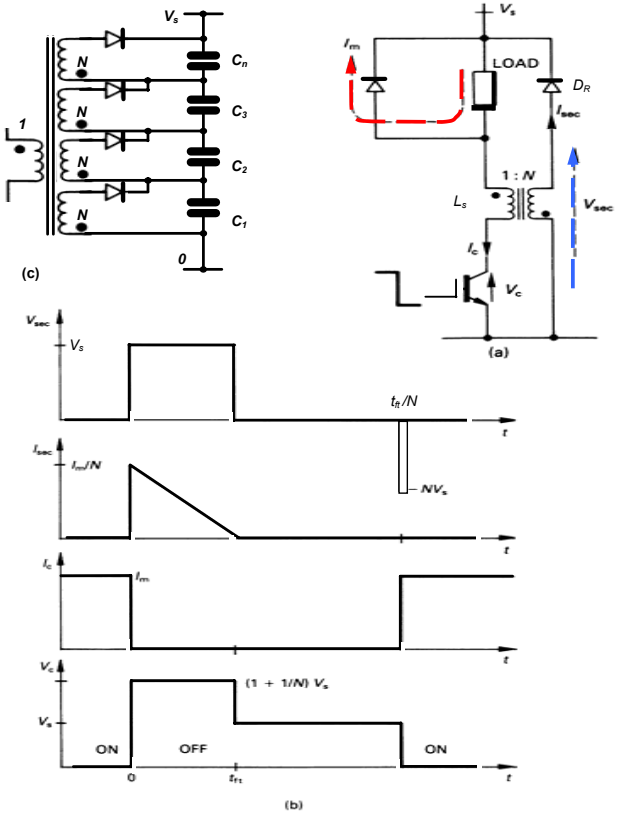


Figure 9.2. *Turn-on snubber with snubber energy recovery via a secondary catch winding: (a) circuit diagram; (b) circuit waveforms; and (c) multilevel recovery.*

At switch turn-on the inductor supports the full rail voltage and, by transformer action, the induced secondary voltage is NV_s. The reverse-blocking voltage seen by the secondary blocking diode D_R is

$$V_c = (1+N)V_s \quad \text{(V)} \tag{9.3}$$

Thus by decreasing the switch voltage requirement with large N, the blocking diode reverse voltage rating is increased, and vice versa when N is decreased.

One further design compromise involving the turns ratio is necessary. The higher the effective pull-down voltage, the quicker the stored energy is returned to the dc supply. The secondary voltage during recovery is fixed at V_s; hence from $v = L\,di/dt$ the current will decrease linearly from I_m/N to zero in time t_{ft}. By equating the magnetically stored primary energy with the secondary energy pumped back into the dc rail source V_s

$$\tfrac{1}{2}L_p I_m^2 = V_s \frac{I_m}{N} \tfrac{1}{2} t_{ft} \quad \text{(J)} \tag{9.4}$$

The core reset time (and the switch minimum off-time), that is the time for the magnetic core energy to be returned to the supply, is given by

$$t_{ft} = L_p \frac{I_m}{V_s} N \quad \text{(s)} \tag{9.5}$$

Thus the lower the turns ratio N, the shorter the core reset time and the higher the upper switching frequency limit. Analysis assumes a short collector current fall time compared with the core reset time.

Primary leakage inductance results in a small portion of the core stored energy remaining in the primary circuit at turn-off. This energy, in the form of primary current, can usually be absorbed and controlled by the capacitive turn-off snubber circuit (*R-C* snubber) across the switch.

Figure 9.2c shows a recovery arrangement with multiple secondary windings, like the link arrangement of a diode clamped multilevel inverter (Chapter 15.3). The reflected voltage, $(1+N/n)V_s$, on to the switch is significantly reduced as the number of secondary windings, n, increases. Auto balancing and regulation of the capacitor voltages is achieved since only the lowest charged (voltage) capacitor has energy transferred to it.

ii. Recovery into the load

Passive inductive energy recovery into the load tends not to significantly affect load voltage regulation since the recovered energy is related to the load current magnitude.

Figure 9.3 shows a passive inductor turn-on snubber with energy recovered into the load and the three recovery stages.

In figure 9.3b, at switch T turn-off, the inductor stored energy $\tfrac{1}{2}L_s I_m^2$ is resonantly transferred to the capacitor C_s in the path L_s - D_s - C_s. The switch is assumed to have a short turn-on time compared to the resonant period. The capacitor C_s voltage and series resonant current are given by

$$\begin{aligned} i(\omega t) &= I_m \cos \omega t \\ V_{Co}(\omega t) &= I_m Z \sin \omega t \end{aligned} \tag{9.6}$$

After time $t = \tfrac{1}{2}\pi\sqrt{L_s C_s}$ the diode D_s blocks preventing continuation of resonance and the final capacitor voltage is

$$V_{Cs} = I_m Z = I_m \sqrt{\frac{L_s}{C_s}} \tag{9.7}$$

When switch T subsequently turns on, the energy stored in C_s is resonantly transferred to the intermediate storage capacitor C_o, through the path C_s - L_r - D - C_o - T shown in figure 9.3c. All the energy in C_s is transferred provided $C_o > C_s$, in which case the diode D_c across C_s conducts, clamping C_s to zero volts. The final voltage on C_o is

$$V_{Co} = V_{Cs}\sqrt{\frac{C_o}{C_s}} = I_m \sqrt{\frac{L_s}{C_o}} \tag{9.8}$$

During the transfer of energy from C_s to C_o the circuit voltage and current waveforms are given by equations (9.11) to (9.14). The voltage on C_o given by (9.8) is retained until subsequent switch turn-off.

The final stage of recovery is shown in figure 9.3d where the capacitor C_o dumps its charge at a constant rate into the load as its voltage falls linearly to zero in a time, independent of the load current

$$t_{Co} = C_o \frac{V_{Co}}{I_m} = \sqrt{L_s C_o} \tag{9.9}$$

during which time the capacitor C_o voltage falls according to

$$V_{Co}(\omega t) = V_{Co}^{t=0} - \frac{I_m}{C_o} t = I_m \sqrt{\frac{L_s}{C_o}} - \frac{I_m}{C_o} t \tag{9.10}$$

The load freewheel diode D_f then conducts the full load current I_m.

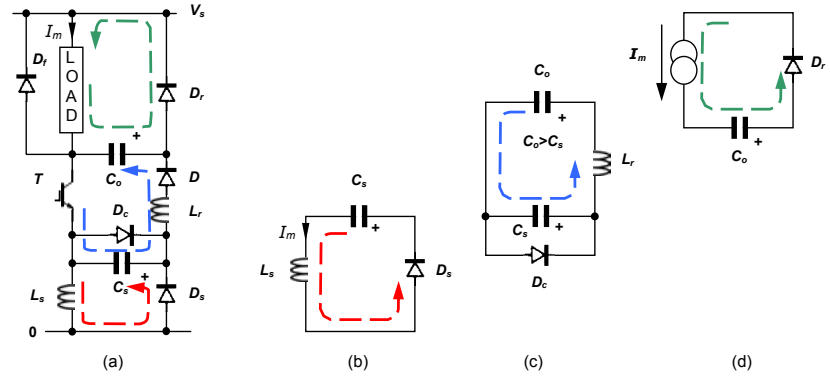


Figure 9.3. *Inductive turn-on snubber with snubber energy recovery intermediate capacitors: (a) circuit diagram; and successive (b) turn-off; (c) turn-on; and (d) turn-off.*

9.1.2 Active recovery

i. Recovery into the dc supply

Figure 9.4 shows an inductive turn-on snubber energy recovery scheme which utilises a switched-mode power supply (smps) based on the boost converter in 15.4, and shown in figure 9.26a.

At switch turn-off the energy stored in the snubber inductor L_s is transferred to the large intermediate storage capacitor C_o via the blocking diode, D_b. The inductor current falls linearly to zero in time $L_s I_m / V_{Co}$. The smps is then used to boost the relatively low capacitor voltage into a higher voltage suitable for feeding energy back into a dc supply. The capacitor charging rate is dependent on load current magnitude. The smps can be controlled so as to maintain the capacitor voltage constant, thereby fixing the maximum switch collector off-state voltage, or varied with current so as to maintain a constant snubber inductor reset time. One smps and storage capacitor can be utilised by a number of switching circuits, each with a blocking/directing diode as indicated in figure 9.4. The diode and switch are rated at $V_s + V_{Co}$. The smps is operated in a discontinuous inductor current mode in order to reduce switch and diode losses and stresses.

If the load and inductive turn-on snubber are re-arranged to be in the cathode circuit, then the complementary smps in figure 9.26b can be used to recover the snubber energy from capacitor C_o.

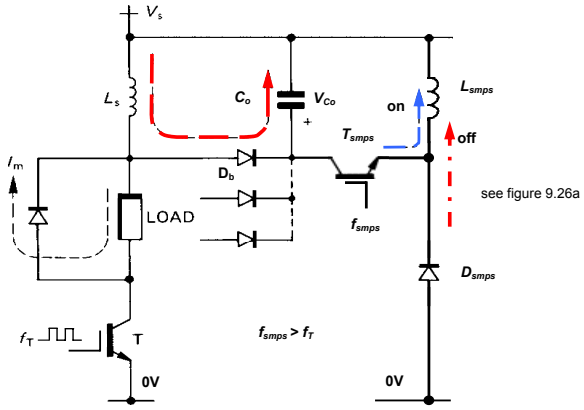


Figure 9.4. *Turn-on snubber with active snubber inductor energy recovery.*

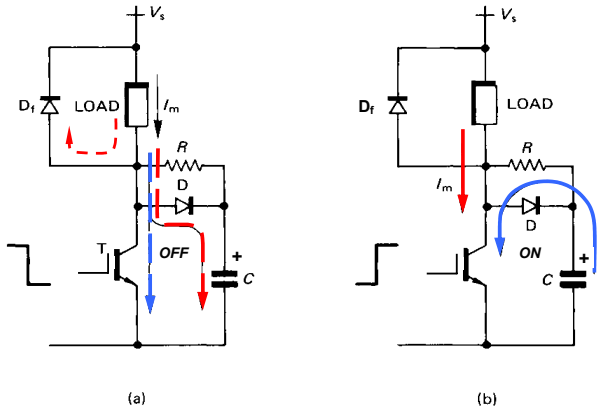


Figure 9.5. *Conventional capacitive turn-off snubber showing currents at IGBT transistor: (a) turn-off and (b) turn-on.*

9.2 Energy recovery for capacitive turn-off snubber circuits – single ended

Figure 9.5 shows the conventional capacitive turn-off snubber circuit used with both the GTO thyristor and the IGBT transistor. At turn-off, collector current is diverted into the snubber capacitor *C* via D. The switch turns off clamped to the capacitor voltage which increases quadratically from zero. At the subsequent switch turn-on the energy stored in *C*, $\frac{1}{2}CV_s^2$ is dissipated as heat, mainly in the resistor *R*. A full functional description and design procedure for the capacitive turn-off snubber circuit is to be found in chapter 8.3.1.

At high voltages and switching frequencies, with slow switching devices, snubber losses ($\frac{1}{2}CV_s^2 f_s$) may be too high to be readily dissipated. An alternative is to recover this energy (either into the load or back into the dc supply), using either passive or active recovery techniques.

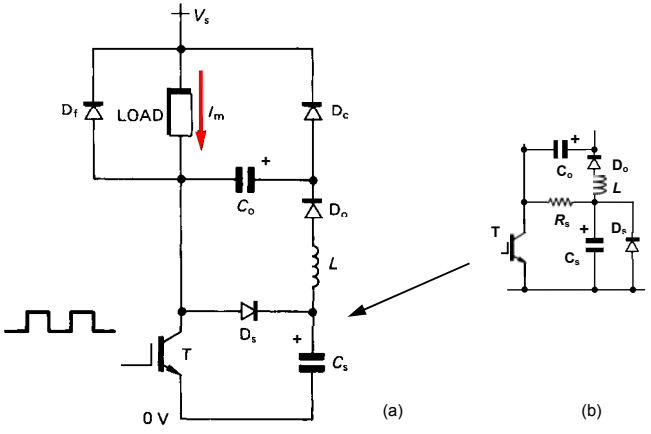


Figure 9.6. *A capacitive turn-off snubber with passive capacitor energy recovery into the load: (a) with a capacitive turn-off snubber and (b) with an RC turn-off snubber.*

9.2.1 Passive recovery

i. Recovery into the load

Figure 9.6 illustrates a passive, lossless, capacitive turn-off snubber energy recovery scheme which dumps the snubber energy, $\frac{1}{2}CV_s^2 f_s$, into the load. The switch turn-off protection is that with a conventional capacitive snubber circuit.

At turn-off the snubber capacitor C_s charges to the voltage rail V_s as shown in figure 9.7a.

At subsequent switch turn-on, the load current diverts from the freewheeling diode D_f to the switch T. Simultaneously the snubber capacitor C_s resonates its charge to capacitor C_o through the path shown in figure 9.7b, T - C_s - L - D_o - C_o.

When the switch next turns off, the snubber capacitor C_s charges and the capacitor C_o discharges into the load. When C_o is discharged, the freewheeling diode conducts. During turn-off C_o and C_s act effectively in parallel across the switching device.

A convenient starting point for the analysis of the recovery scheme is at switch turn-on when snubber energy is transferred from C_s to C_o.

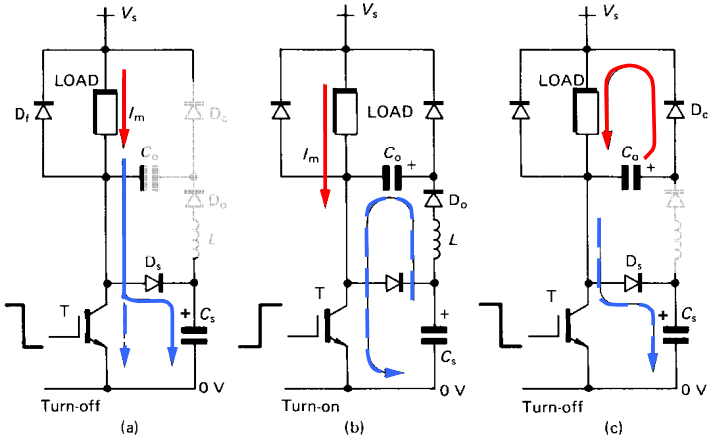


Figure 9.7. *Energy recovery turn-off snubber showing the energy recovery stages: (a) conventional snubber action at turn-off; (b) intermediate energy transfer at subsequent switch turn-on; and (c) transferred energy dumped into the load at subsequent switch turn-off.*

At switch turn-on

The active equivalent circuit portions of figure 9.7b are shown in figure 9.8a.

Analysis of the *L*-*C* resonant circuit with the initial conditions shown yields the following capacitor voltage and current equations. The resonant current is given by

$$i(\omega t) = \frac{V_s}{Z}\sin\omega t \quad \text{(A)} \tag{9.11}$$

where $Z = \omega \ell_s = \frac{1}{\omega C_o} = Z_o\sqrt{\frac{n+1}{n}}$ (ohms) $\quad Z_o = \sqrt{\frac{\ell_s}{C_o}}$ (ohms)

$\omega = \omega_o\sqrt{\frac{n+1}{n}}$ (rad/s) $\quad \omega_o = \frac{1}{\sqrt{\ell_s C_o}}$ (rad/s)

$n = \frac{C_s}{C_o}$

The snubber capacitor voltage decreases from V_s according to

$$V_{Cs} = V_s\left\{1 - \frac{1}{1+n}(1-\cos\omega t)\right\} \quad \text{(V)} \tag{9.12}$$

while the transfer capacitor voltage charges from zero according to

$$V_{Co} = V_s\frac{n}{1+n}(1-\cos\omega t) \quad \text{(V)} \tag{9.13}$$

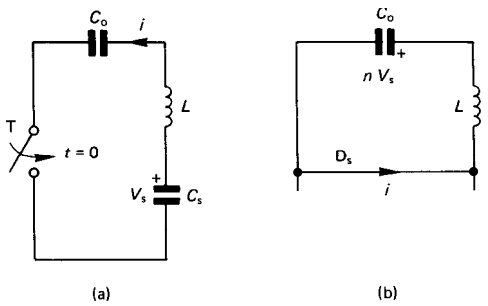


Figure 9.8. *Equivalent circuit for the intermediate energy transfer phase of snubber energy recovery, occurring via: (a) the main switch T and (b) then via the snubber diode D_s.*

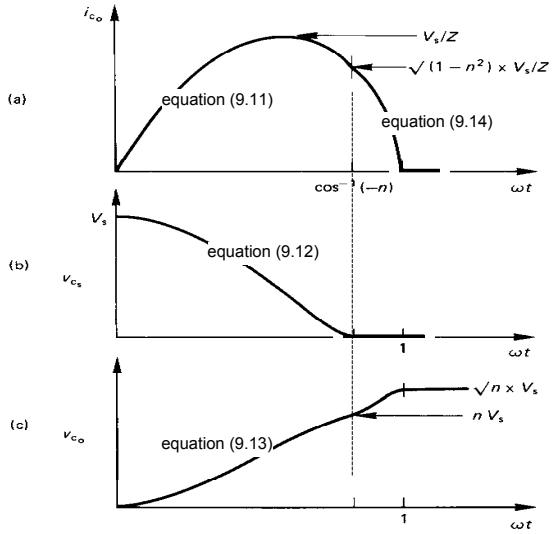


Figure 9.9. *Circuit waveforms during intermediate energy transfer phase of snubber energy recovery: (a) transfer capacitor C_0 current; (b) snubber capacitor voltage; and (c) transfer capacitor voltage.*

Examination of equation (9.12) shows that if $n > 1$, the final snubber capacitor C_s voltage at $\omega t = \pi$ will be positive. It is required that C_s retains no charge, ready for subsequent switch turn-off; thus $n \le 1$, that is $C_o \ge C_s$. If C_o is greater than C_s equation (9.12) predicts C_s will retain a negative voltage. Within the practical circuit of figure 9.6, C_s will be clamped to zero volts by diode D_s conducting and allowing the remaining stored energy in L to be transferred to C_o. The new equivalent circuit for $\omega t = \cos^{-1}(-n)$ is shown in figure 9.8b. The resonant current, hence transfer capacitor voltage are given by

$$i(\omega t) = \frac{V_s}{Z}\sin(\omega_o t + \phi) \quad \text{(A)}$$
$$V_{Co} = \sqrt{n}\,V_s \cos(\omega_o t + \phi) \quad \text{(V)} \qquad (9.14)$$

where $t \ge 0$ and $\phi = -\tan^{-1}\sqrt{\frac{1-n^2}{n}}$.

In maintaining energy balance, from equation (9.14) when the inductor L current $i(\omega t) = 0$, the final voltage on C_o is $\sqrt{n}\,V_s$ and C_s retains no charge, $V_{Cs} = 0$.

The voltage and current waveforms for the resonant energy transfer stage are shown in figure 9.9.

At switch turn-off

Energy dumping from C_o into the load and snubber action occur in parallel and commence when the switch is turned off. As the collector current falls to zero in time t_{fi} a number of serial phases occur. These phases, depicted by capacitor voltage and current waveforms, are shown in figure 9.10.

Phase one

Capacitor C_o is charged to $\sqrt{n}\,V_s$, so until the snubber capacitor C_s charges to $\left(1-\sqrt{n}\right)V_s$, C_o is inactive. Conventional snubber turn-off action occurs as discussed in chapter 8.3.1. The snubber capacitor voltage increases according to

$$V_{Cs} = \tfrac{1}{2}\frac{I_m}{C_s t_{fi}}t^2 \quad \text{(V)} \qquad (9.15)$$

while C_o remains charged with a constant voltage of $\sqrt{n}\,V_s$. This first phase is complete at t_o when

$$V_{Cs} = v_o = \tfrac{1}{2}\frac{I_m t_o^2}{C_s t_{fi}} = \left(1-\sqrt{n}\right)V_s \quad \text{(V)} \qquad (9.16)$$

whence

$$t_o = \sqrt{\frac{2\left(1-\sqrt{n}\,V_s C_s t_{fi}\right)}{I_m}} \quad \text{(s)} \qquad (9.17)$$

and the collector current

$$I_o = I_m\left(1 - {}^{t_o}\!/_{t_{fi}}\right) \quad \text{(A)} \qquad (9.18)$$

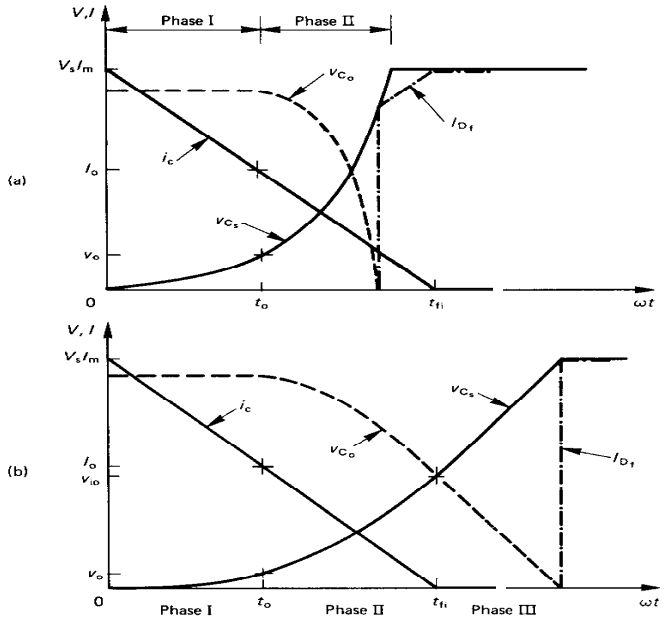


Figure 9.10. *Circuit waveforms at switch turn-off with turn-off snubber energy recovery when: (a) the snubber C_s is fully charged before the switch current at turn-off reaches zero and (b) the switch collector current has fallen to zero before the snubber capacitor has charged to V_s.*

Phase two

When C_s charges to $\left(1-\sqrt{n}\right)V_s$, the capacitor C_o begins to discharge into the load. The equivalent circuit is shown in figure 9.11a, where the load current is assumed constant while the collector current fall is assumed linear. The following Kirchhoff conditions must be satisfied

$$V_s = V_{Cs} + V_{Co} \quad \text{(V)} \tag{9.19}$$

$$I_m = i_{Co} + i_{Cs} + I_o(1 - t/t_{fi}) \quad \text{(A)} \tag{9.20}$$

for $0 \le t \le t_{fi} - t_o$

Under these conditions, the snubber capacitor voltage increases according to

$$V_{Cs} = \frac{n}{1+n}\frac{1}{C_s}\left[\left(I_m - I_o\right)t + \tfrac{1}{2}t^2/t_o\right] + \left(1-\sqrt{n}\right)V_s \quad \text{(V)} \tag{9.21}$$

with a current

$$i_{Cs} = \frac{1}{1+n}\left\{I_m - I_o(1 - t/t_o)\right\} \quad \text{(A)} \tag{9.22}$$

The transfer dump capacitor C_o discharges with a current given by

$$i_{Co} = i_{Cs}/n \tag{9.23}$$

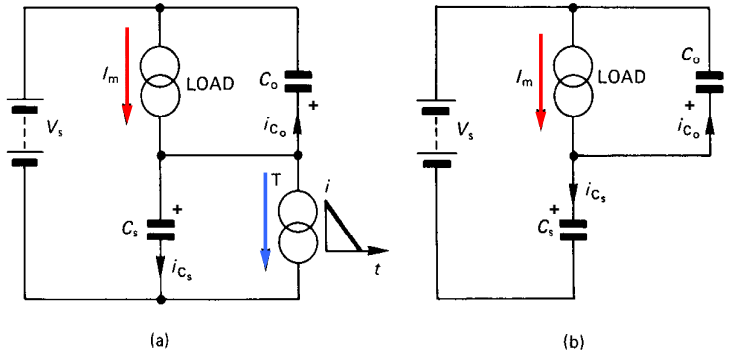


Figure 9.11. *Turn-off snubber equivalent circuit during energy recovery into the load when: (a) C_o begins to conduct and (b) after the switch has turned off.*

Phase three

If the snubber capacitor has not charged to the supply rail voltage before the switch collector current has reached zero, phase three will occur as shown in figure 9.10b. The equivalent circuit to be analysed is shown in figure 9.11b. The Kirchhoff equations describing this phase are similar to equations (9.19) and (9.20) except that in equation (9.20) the component $I_o(1- t/t_0)$ is zero.

The capacitor C_s, charging current is given by

$$i_{Cs} = \frac{n}{1+n} I_m \quad \text{(A)} \tag{9.24}$$

while the dumping capacitor C_o current is

$$i_{Co} = i_{Cs}/n \quad \text{(A)} \tag{9.25}$$

The snubber capacitor charges linearly, according to

$$V_{Cs} = v_{io} + \frac{n}{1+n}\frac{I_m}{C_s}t \quad \text{(V)} \tag{9.26}$$

When C_s is charged to the rail voltage V_s, C_o is discharged and the load freewheeling diode conducts the full load current I_m.

Since the snubber capacitor energy is recovered there is no energy loss penalty for using a large snubber capacitance and the larger the capacitance, the lower the switch turn-off switching loss. The energy to be recovered into the load is fixed, $\frac{1}{2}C_sV_s^2$ and at low load current levels the long discharge time of C_o may inhibit proper snubber circuit action. This is generally not critical since switching losses are small at low load current levels. Output voltage regulation is reduced, since the amount of energy recovered into the load is independent of the load current.

ii. Recovery into the dc supply

Figure 9.12 show two turn-off snubber circuits where the energy is recovered back into the dc supply. The ac circuit operational mechanisms are the same for both circuits.

When the switch T is turned off the snubber capacitor C_s charges to the dc rail voltage V_s.

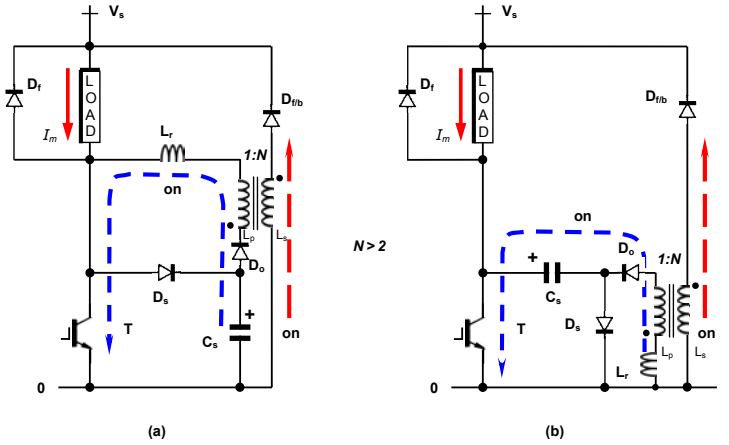


Figure 9.12. *A capacitive turn-off snubber with passive energy recovery into the supply: (a) basic capacitive turn-off snubber and (b) an alternative configuration.*

At switch T turn-on, the snubber capacitor C_s resonates with inductor L_r through the coupled transformer primary L_p, in the loop C_s - D_o – L_p - L_r - T, returning energy to the dc supply through the coupled secondary circuit. The primary voltage is V_s/N, and provided this referred voltage is less than a half V_s, all the energy on C_s is transferred to the dc supply via the transformer. The snubber diode D_s clamps the capacitor C_s voltage to zero, and excess energy in L_r is transferred to the dc supply, in the loop $D_o – L_p$ - $L_r – D_s$, as the inductor L_r current falls linearly to zero when opposed by the referred dc link voltage via the transformer. In figure 9.12a, the winding secondary can be connected to the other terminal of C_s.

Once the energy transfer is complete, the transformer core magnetising current resets to zero in the same Kirchhoff loop, but at a low voltage. Reset must be complete in one complete period of switch T.

iii RC snubber recovery

The IGCThyristor is commonly used and characterised with an RC snubber. The figure 9.6b shows how the snubber diode D_s in figure 9.6a can be replaced by a resistor to form an RC snubber, provided diode D_s is used to clamp the minimum snubber capacitor voltage to zero. The resistor losses are $\frac{1}{2}C_sV^2$. The snubber capacitor stored energy after turn-off, $\frac{1}{2}C_sV^2$, can be recovered at switch turn-on, provided the R_sC_s time constant is at least comparable with the LC resonant period – an unlikely condition.

9.2.2 Active recovery

i. Recovery into the dc supply

Active energy recovery methods for the turn-off snubber are simpler than the technique needed for active recovery of turn-on snubber circuit stored energy. This is because the energy to be recovered from the turn-off snubber is fixed at $\frac{1}{2}C_sV_s^2$ and is independent of load current. In the case of the turn-on snubber, the energy to be recovered is load current magnitude dependent (αI_L^2) which complicates active recovery. Active turn-off snubber energy recovery usually involves the production of an intermediate capacitive energy storage stage involving a positive or negative voltage rail (with respect to the emitter of the principal switch).

a Negative intermediate voltage rail

At switch T turn-on the snubber capacitor stored energy is resonated into a large intermediate storage capacitor C_o as shown in figure 9.13a. Recovery from C_s to C_o at switch T turn-on occurs through the following loops:

at switch T turn-on when $V_{Cs} > 0$: C_s -T- C_o - L- D_a (as shown in figure 9.8a and equations (9.12) - (9.13))

then when $V_{Cs} = 0$: D_s - C_o - L- D_a (as shown in figure 9.8b and equation (9.14))

The switch current is increased by the resonant current, which has a maximum of $V_{Co}/\sqrt{L/C_s}$. It is possible to use the energy in C_o as a negative low-voltage rail supply. This passive recovery technique suffers from the problem that the recovered energy $\frac{1}{2}C_sV_s^2$ may represent more energy than the low-voltage supply requires. An independent buck-boost smps can convert excess energy stored in C_o to a more useful voltage level. Producing the gate drive for the smps switch T_{smps} presents few difficulties since the gate-emitter has a low dc offset and does not experience any *dv/dt* relative to the emitter reference voltage of the main switch T.

The basic recovery circuit, with the buck-boost smps, can form the basis of an active turn-off snubber energy recovery circuit when switches are series connected, as considered in section 9.4.
It may be noticed that the 'Cuk' converter in chapter 17.6 is in fact the snubber energy recovery circuit in figure 9.13a, controlled in a different mode.

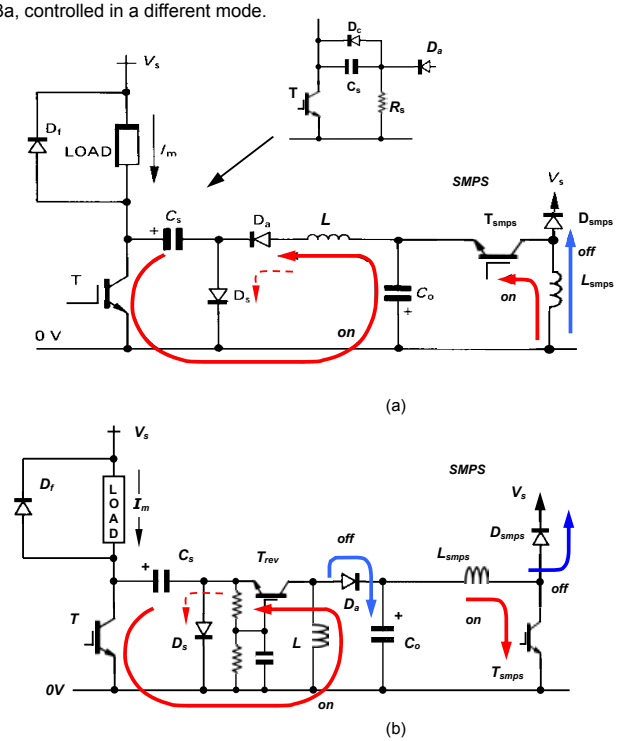


Figure 9.13. *Switching circuit for recovering turn-off snubber capacitor energy, and for providing either (a) a negative voltage rail and/or transferring to V_s, via a buck-boost smps or (b) a positive voltage rail and/or transferring to V_s, via a boost smps.*

b Positive intermediate voltage rail
A positive voltage source, with respect to the main switch emitter, can be produced with the recovery circuit in figure 9.13b. Practically, an extra switch, T_{rev}, is needed in order to minimise the time of current decay in the loop L - D_s, after the switch T is turned on and the voltage on the snubber capacitor C_s has resonated to zero. A passive resistor-capacitor network can be used to synchronise the turn-on (due to the main switch T turning on) and turn-off (due to diode D_s becoming forward biased) of the low-voltage switching device T_{rev}. Recovery from C_s to C_o at switch T turn-on occurs through the following Kirchhoff current loops:
at switch T turn-on when T_{rev} is on and $V_{Cs} > 0$: C_s -T - L- T_{rev} for a period $\frac{1}{2}\pi\sqrt{LC_s}$
then when T_{rev} is off and $V_{Cs} = 0$: C_o - L- D_a for a period $V_s/\omega_o V_{Co}$
A boost smps controls and transfers the energy on C_o to the dc rail through diode D_{smsp}.
The basic recovery circuit, with the boost smps, when cascade connected, can form the basis of an active turn-off snubber energy recovery circuit for series connected switches, as considered in 9.4.

ii. RC snubber recovery
The IGCThyristor is commonly used and characterised with an R-C snubber (as opposed to a parallel connected series capacitor-diode turn-off snubber). The insert in figure 9.13a, for use in figures 9.13a and b, shows how the snubber diode D_s can be replaced by a resistor to form an R-C snubber, provided diode D_c is used to clamp the minimum snubber capacitor voltage to zero. The resistor losses are $\frac{1}{2}C_sV^2$. Most of the snubber capacitor stored energy after turn-off, $\frac{1}{2}C_sV^2$ at switch turn-off, (depending on the R_s-C_s time constant), can be recovered using either of the basic circuits in figure 9.13, or the circuits in figures 9.6 and 9.14, provided the R_sC_s time constant is greater than the LC resonant period.

Whether a positive or negative intermediate voltage is produced on C_o, (typically a few tens of volts, but much higher if part of a turn-on snubber recovery circuit), the energy on C_o is usually smps converted to stable gate voltage levels of the order of ±15V. Since a dual rail polarity gate level supply is needed, the polarity of the voltage on C_o (viz., positive or negative) is inconsequential.

9.3 Unified turn-on and turn-off snubber circuit energy recovery – single ended

9.3.1 Passive recovery

Conventional inductive turn-on and capacitive turn-off snubber circuits can both be incorporated around a switching device as shown in figure 8.20 where the stored energy is dissipated as heat in the reset resistor. Figure 9.14 shows unified turn-on and turn-off snubber circuits which allow energy recovery from both the snubber capacitor C_s and inductor ℓ_s.

i. Recovery into the load
The snubber capacitor energy is recovered by the transfer process outlined in section 9.2.1. Figure 9.14a shows the energy transfer (recovery) paths at switch turn-off. The capacitor C_o and inductor ℓ_s transfer their stored energy to the load in parallel, such that the inductor voltage is clamped to the capacitor voltage V_{Co}.
As C_o discharges, the voltage across ℓ_s decreases to zero, at which time the load freewheel diode D_f conducts. Any remaining inductor energy is dissipated as unwanted heat in circuit resistance. Proper selection of ℓ_s and C_s ($\frac{1}{2}L_sI_m^2 \le \frac{1}{2}C_sV_s^2$) can minimise the energy that is lost although all the snubber capacitor energy is recovered, neglecting diode and stray resistance losses. The energy (controlled by and transferred to the turn-on snubber inductor ℓ_s) associated with freewheel diode reverse recovery current, is also recovered.

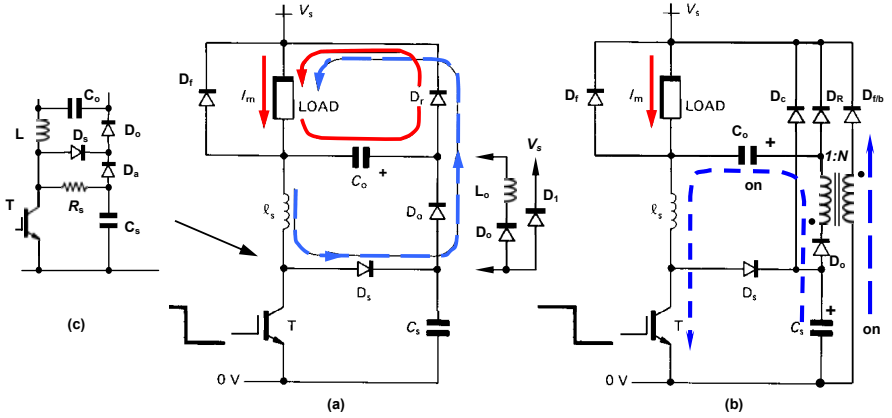


Figure 9.14. *Switching circuits incorporating unified turn-on and turn-off snubber, showing recovery path of energy (a) in C_o and ℓ_s; (b) in C_s and ℓ_s through D_r.; and (c) recovery circuit when an RC snubber is employed.*

At switch turn-on
When the switch is off, the freewheel diode D_f conducts the load current I_m, capacitor voltage $V_{Cs} = V_s$ and $V_{Co} = 0$.

Phase one: t_{P1}^{on}
When the switch is turned on, the series inductor ℓ_s performs the usual turn-on snubber function of controlling the switch *di/dt* according to (assuming the switch voltage fall time is relatively short)

$$i(t) = \frac{V_s}{\ell_s} t \qquad (9.27)$$

The switch current rises linearly to the load current level I_m and then continues to a level I_{RR} higher as the freewheel diode D_f recovers with currents in the paths shown in figure 9.15a. This diode reverse recovery current I_{RR} is included in the analysis since the associated energy transferred to the turn-on inductor is subsequently recovered.

The peak switch current $I_m + I_{RR}$ is reached after the duration t_{P1}^{on}

$$t_{P1}^{on} = (I_m + I_{RR})\frac{\ell_s}{V_s} \qquad (9.28)$$

As long as the freewheel-diode conducts, the load is clamped to near zero volts, thus C_s remains charged to V_s.

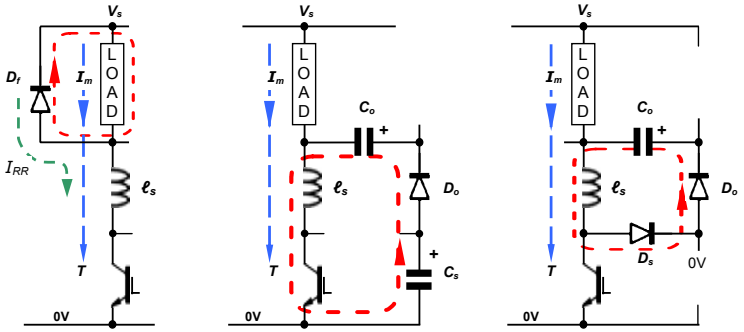


Figure 9.15. *Unified turn-on and turn-off snubber at switch turn-on, showing (a) current build-up in ℓ_s; (b) energy resonant transfer from C_s to C_o; and (c) energy transfer from ℓ_s to C_o through D_s.*

Phase two: t_{P2}^{on}
The turn-off snubber capacitor C_s charge resonates in the path C_s - D_o - C_o - ℓ_s and through the switch T, as shown in figure 9.15b. The capacitor voltages and resonant current are given by ($n = C_s / C_o$)

$$i_{Cs}(\omega t) = i_{Co}(\omega t) = \frac{V_s}{Z}\sin\omega t + I_{RR}\cos\omega t \qquad (9.29)$$

$$V_{Cs}(\omega t) = V_s\left(1 - \frac{1}{1+n}(1-\cos\omega t)\right) + \frac{\omega_o}{\omega}\frac{Z}{n}I_{RR}\sin\omega t \qquad (9.30)$$

$$V_{Co}(\omega t) = V_s\frac{n}{1+n}(1-\cos\omega t) + \frac{\omega_o}{\omega}Z\,I_{RR}\sin\omega t \qquad (9.31)$$

where $Z = \omega\ell_s = \frac{1}{\omega C_o} = Z_o\sqrt{\frac{n+1}{n}}$ (ohms) $\quad Z_o = \sqrt{\frac{\ell_s}{C_o}}$ (ohms) $\quad n = \frac{C_s}{C_o}$

$\omega = \omega_o\sqrt{\frac{n+1}{n}}$ (rad/s) $\quad \omega_o = \frac{1}{\sqrt{\ell_s C_o}}$ (rad/s)

The freewheel diode D_f voltage is

$$V_{Df}(\omega t) = V_s + V_{Co} - V_{Cs} = V_s(1-\cos\omega t) + I_{RR}Z\sin\omega t \qquad (9.32)$$

When the freewheel-diode current reaches its peak recovery level, I_{RR}, it is able to support a voltage which from equation (9.32) sinusoidally increases from zero. Specifically the freewheel-diode reverse bias V_{Df} is controlled such that zero voltage turn-off occurs resulting in low recovery power losses. Stray or inductance deliberately introduced in series with D_o (to decrease the resonant peak current given by equation (9.29) as approximately V_s / Z) produces a freewheel-diode recovery step voltage $V_s\,\ell_s / (\ell_s + L_{stray})$, but the step is always less than V_s.
The resonant period prematurely ends (since $n < 1$) when the snubber capacitor C_s voltage reduces to zero and is clamped to zero by conduction of the snubber diode D_s, as shown in figure 9.15c. Assuming $I_{RR} = 0$ (to obtain a tractable solution), equating equation (9.30) to zero yields the time for period 2, t_{P2}^{on}, that is

$$t_{P2}^{on} = \frac{\cos^{-1}(-n)}{\omega} \qquad (9.33)$$

at which time

$$i_{Co}(t_{P2}^{on}) = \frac{V_s}{Z}(1-n^2) \qquad (9.34)$$

and

$$V_{Co}(t_{P2}^{on}) = nV_s \qquad (9.35)$$

Phase three: t_{P3}^{on}
The remaining energy stored in ℓ_s is resonantly transferred into C_o in the path D_o - C_o - ℓ_s - D_s, with initial conditions given by equations (9.34) and (9.35), according to

$$V_{Co}(\omega_o t) = \sqrt{n}\,V_s\sin(\omega_o t + \phi) \qquad (9.36)$$

and

$$i(\omega_o t) = \sqrt{n}\frac{V_s}{Z_o}\cos(\omega_o t + \phi) \qquad (9.37)$$

The resonant current reaches zero and energy transfer to C_o is complete, after a period

$$t_{P3}^{on} = \frac{1/2\pi - \phi}{\omega_o} \qquad (9.38)$$

If the diode reverse recovery energy is reintroduced, based on energy transfer balance, the final voltage on C_o is

$$V_{Co}(t_{P3}^{on}) = \sqrt{nV_s^2 + (Z_o I_{RR})^2} \qquad (9.39)$$

The turn-on equations (9.29) to (9.37) are essentially the same as equations (9.11) to (9.14) for the turn-off snubber energy recovery circuit considered in section 9.2.1, except free-wheel diode reverse recovery has now been included. The circuit turn-on voltage and current waveforms shown in figure 9.9 are also applicable.

At switch turn-off
When the switch is on, it conducts the load current I_m and the snubber capacitor C_s voltage is zero, while the transfer capacitor voltage $V_{Co}(t_{P3}^{on}) = \sqrt{n}\,V_s = V_o$ (neglecting the I_{RR} component) is a result of the previous switch turn-on. When the switch T is turned off, the collector current decreases linearly from I_m towards zero in time t_{fi}.

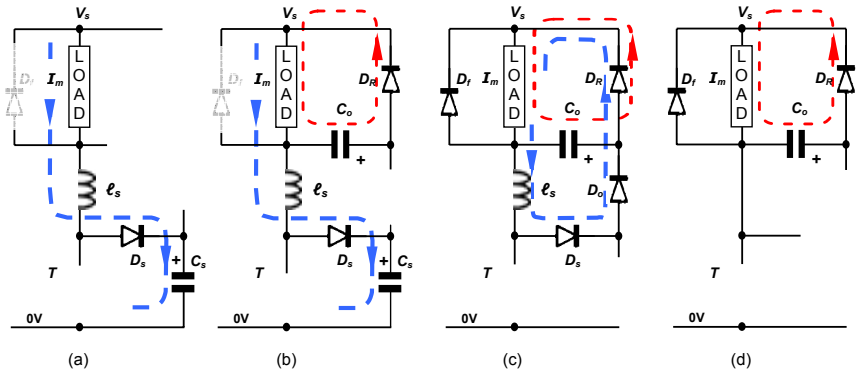


Figure 9.16. *Unified turn-on and turn-off snubber at switch turn-off, showing (a) current diversion to snubber capacitor C_s; (b) transfer capacitor C_o releasing energy (c) energy transfer to the load simultaneously from ℓ_s and C_o through D_R; and (d) energy transfer from C_o into the load through D_R.*

Phase 1: t_{P1}^{off}
The load current is progressively diverted to the snubber capacitor as the collector current decreases, giving a capacitor (and collector) voltage of

$$v_{ce} = V_{Cs}(t) = \frac{1}{C_s}\int_0^t (I_m - i_c)\,dt = \frac{1}{C_s}\int_0^t I_m\frac{t}{t_{fi}}\,dt = \frac{I_m}{C_s}\frac{t^2}{2t_{fi}} \qquad 0 \le t \le t_{fi} \qquad (9.40)$$

If the collector current reaches zero before any other associated recovery processes occurs, then after the collector current has reached zero, the collector and snubber voltages rise linearly (being clamped in parallel), with currents in the paths shown in figure 9,16a, according to

$$v_{ce} = V_{Cs}(t) = \tfrac{1}{2}\frac{I_m t_{fi}}{C_s} + \frac{I_m t}{C_s} \quad \text{provided} \quad \tfrac{1}{2}\frac{I_m t_{fi}}{C_s} \le V_s - V_o \tag{9.41}$$

The collector voltage reaches V_s at a time given from equation (9.41) when $V_{Cs} = V_s - V_{Co}$ as

$$t_{P1}^{off} = \frac{C_s}{I_m}(V_s - V_o) + \tfrac{1}{2}t_{fi} \tag{9.42}$$

where V_o is given by equation (9.39) and the period duration includes the collector linear fall period t_{fi}.

Phase 2: t_{P2}^{off}
When the collector (and snubber) voltage V_{Cs} reaches $V_s - V_o$ capacitor C_o begins to discharge into the load providing the load current I_m. Simultaneously C_s charges to V_s through ℓ_s, as shown in figure 9.16b. The relevant circuit capacitor voltages and current are

$$i_{\ell s}(\omega t) = I_m \frac{n}{1+n}\left(1 + \frac{1}{n}\cos\omega t\right) \tag{9.43}$$

$$V_{Cs}(\omega t) = I_m Z_o \frac{1}{n+1}\left(\frac{1}{\sqrt{n+1}}\sin\omega t + \omega_o t\right) + V_s - V_o \tag{9.44}$$

$$V_{Co}(\omega t) = I_m Z_o \frac{1}{n+1}\left(\frac{1}{\sqrt{n+1}}\sin\omega t - \omega_o t\right) + V_o \tag{9.45}$$

This phase is complete when the snubber capacitor C_s is charged to the supply voltage, V_s, assuming the inductor current is greater than zero at that time. Let the inductor current be I_2 at the end of the off-period t_{P2}^{off} and the capacitor C_o voltage be V_2.

Phase 3: t_{P3}^{off}
The snubber capacitor is clamped to the rail voltage. The transfer capacitor C_o and snubber inductor ℓ_s both release energy in parallel into the load through the paths shown in figure 9.16c. The inductor voltage is clamped to the capacitor C_o voltage. The snubber inductor current is

$$i_{\ell s}(\omega_o t) = I_m + \frac{V_2}{Z_o}\sin\omega_o t + (I_2 - I_m)\cos\omega_o t \tag{9.46}$$

while the transfer capacitor voltage is

$$V_{Co}(\omega_o t) = V_2\cos\omega_o t + Z_o(I_2 - I_m)\sin\omega_o t \tag{9.47}$$

One of two conditions form the completion of this phase

- the transfer capacitor voltage reaches zero before the snubber inductor current reaches zero
- the snubber inductor current reaches zero before the transfer capacitor voltage reaches zero

The **first** condition represents the case where the remaining inductor current associated energy is lost as it freewheels to zero in the low voltage path ℓ_s - D_o - D_R and the load.
In the **second** case, the inductor current given by equation (9.46) reaches zero, while the transfer capacitor C_o continues to discharge into the load as shown in figure 9.16d. The inductor current is prevented from reversing by diode D_s. Once the inductor current has fallen to zero, the transfer capacitor voltage falls linearly to zero as it provides the load current I_m. This second case represents the situation when 100% of all snubber (inductor ℓ_s and capacitor C_s) and diode reverse recovery energy is recovered, that is

$$\tfrac{1}{2}\ell_s(I_m + I_{RR})^2 \le \tfrac{1}{2}C_s V_s^2 \tag{9.48}$$

Snubber reset and recovery is complete when the snubber inductor current and transfer capacitor voltage are both zero, the collector voltage has ramped to V_s, and the free-diode conducts the full load current I_m. From equation (9.47), this stage is complete when $V_{Co}(t_{P3}^{off}) = 0$, that is

$$t_{P3}^{off} = \frac{1}{\omega_o}\tan^{-1}\left(\frac{V_2}{Z_o(I_2 - I_m)}\right) \tag{9.49}$$

Now the switch can be turned on.

ii. *RC-L dual snubber recovery*
The IGCThyristor is commonly used and characterised with an RC snubber and an inductive turn-on snubber. Figure 9.14c shows how the snubber diode D_s in figure 9.14a can be replaced by a resistor to form an RC snubber, provided diode combination D_a - D_s is used to clamp the minimum snubber capacitor voltage to zero. The resistor losses are $\tfrac{1}{2}C_sV^2$. The snubber capacitor stored energy after turn-off, $\tfrac{1}{2}C_sV^2$, can be recovered at switch turn-on, while the inductive turn-on energy $\tfrac{1}{2}L_sI^2$ is recovered at switch turn-off, provided the R_sC_s time constant is greater than the LC resonant period.

iii. Recovery into the load and supply

Figure 9.14b shows a dual snubber energy recovery technique where a portion of the resonance energy is transferred back to the dc supply (as opposed to the load) at switch turn-on, through a magnetically coupled circuit where it is required of the turns ratio that $N > 2$. This reduces the energy transferred from the snubbers to the load, giving better load regulation under light load conditions. Load regulation with light loads is poor since the snubber capacitor energy is fixed, $\tfrac{1}{2}C_sV_s^2$, independent of the load, I_m. In the analysis to follow, the recovery contribution of the freewheel diode reverse recovery energy is neglected.

At switch turn-on
The turn-on phase is essentially the same as the circuit considered in figure 9.14a, except the transformer is seen as an opposing emf voltage source V_s/N.

Phase one: t_{P1}^{on}
The switch current fall period is described by equation (9.27) and the time of the first turn-on period is given by equation (9.28).

Phase two: t_{P2}^{on}
The equations (9.29) to (9.35) are modified to account for the transformer referred voltage V_s/N

$$i_{\ell s}(\omega t) = i_{Cs}(\omega t) = i_{Co}(\omega t) = \frac{N-1}{N} \times \frac{V_s}{Z}\sin\omega t \tag{9.50}$$

$$V_{Cs}(\omega t) = V_s \times \frac{1}{N(1+n)} \times \left[1 + Nn + (N-1)\cos\omega t\right] \tag{9.51}$$

$$V_{Co}(\omega t) = V_s \frac{n(N-1)}{N(n+1)}(1 - \cos\omega t) \tag{9.52}$$

The instantaneous power being returned to the supply through the transformer is given by

$$p(\omega t) = \frac{V_s}{N} \times i_{\ell s}(\omega t) = \frac{V_s}{N} \times \frac{N-1}{N} \times \frac{V_s}{Z}\sin\omega t = \frac{N-1}{N^2} \times \frac{V^2}{Z}\sin\omega t \tag{9.53}$$

The time for this period is given by equation (9.51), when the snubber capacitor voltage is zero

$$t_{P2}^{on} = \frac{1}{\omega} \times cos^{-1}\left(-\frac{nN+1}{N-1}\right) \tag{9.54}$$

The energy returned to the supply is

$$w_{Trans}(t_{P2}^{on}) = \frac{n+1}{N} \times \frac{V_s^2}{\omega Z} = \frac{1}{N} \times C_s V_s^2 < \tfrac{1}{2}C_sV_s^2 \text{ since } N > 2 \qquad \text{(J)} \tag{9.55}$$

Phase three: t_{P3}^{on}
Energy continues to be recovered back into the supply V_s through the transformer when the resonant current transfers to the diode D_s. Capacitor C_s charges to V_s and is clamped to V_s by diode D_c.
The final voltage on the transfer capacitor C_o is

$$V_{Co}(t_{P3}^{on}) = \frac{V_s}{N}\left[\sqrt{1 + nN^2} - 1\right] \tag{9.56}$$

The total energy transferred to the supply through the transformer is the difference between the initial energy in ℓ_s and C_s and the final energy in C_o.

$$w_{Trans}(t_{P2}^{on} + t_{P3}^{on}) = \tfrac{1}{2}C_sV_s^2 + \tfrac{1}{2}\ell_s I_m^2 - \tfrac{1}{2}C_o\frac{V_s^2}{N^2}\left[\sqrt{1 + nN^2} - 1\right]^2 \tag{9.57}$$

If the turn-on inductor current reaches zero before the third phase can commence (due to N being too small), then the turn-off snubber does not fully discharge, and will act as a soft clamp in the subsequent switch turn-off cycle. The capacitors retain the following voltages

$$V_{Cs} = V_s\frac{2 + Nn - N}{N(n-1)} = V_s - \frac{2}{N(n-1)}V_s \tag{9.58}$$

$$V_{Co} = V_s\frac{2n(N-1)}{N(n+1)} \tag{9.59}$$

At switch turn-off
The circuit recovery operation at turn-off is essentially the same as when no transformer is used ($N \to \infty$), except that the voltage on C_o at the begin of turn-off is given by equation (9.59) or equation (9.56), as appropriate.

Operating regions of the dual energy recovery circuits

Both the passive unified recovery circuits analysed can be assessed simultaneously for their operational bounds, since the bounds for the transformerless version in figure 9.14a are obtained by setting N to infinitely in the appropriate equations for the recovery circuit in figure 9.14b. Figure 9.17 shows various operational boundaries for the two unified passive energy recovery circuits analysed. The various boundaries are determined from the operating equations for the circuits.
The boundaries in figure 9.17a show the regions of full snubbering and for soft snubbering where the capacitor C_s is not reset to zero voltage during the resonant cycles at turn-on. The boundaries are summarised as follows

$$n < \frac{N-2}{N} \tag{9.60}$$

$$n < \frac{N}{N-2} \tag{9.61}$$

The boundaries in parts b and d of figure 9.17 satisfy equation (9.57), namely the capacitor energy is less than the inductor energy. The current is normalised with respect to $\sqrt{n}V_s / Z_o$. Part d shows that the relative range for 100% recovery, defined as $(\hat{I} - \breve{I}) / \hat{I}$, is independent of the transformer turns ratio.

Figure 9.17c shows the normalised (with respect to $2\pi\sqrt{n}/\omega_o$) reset time at turn-off. The reset time at turn-on is the sum of periods one and two, but is dominated by the second turn-on period, namely

$$\breve{t}_{on} = \frac{1}{\omega}\cos^{-1}(-n) \tag{9.62}$$

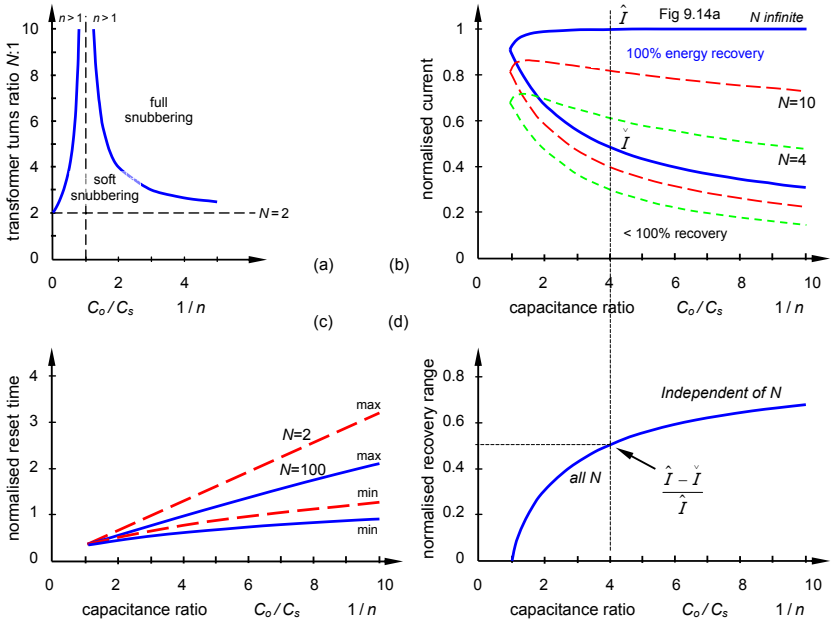


Figure 9.17. *Unified, passive snubbering characteristics: (a) operating regions with recovery transformer; (b) 100% recovery regions with different transformer turns ratios; (c) normalised circuit reset limits; and (d) normalised recovery range independent of transformer turns ratio.*

9.3.2 Active recovery

i. Recovery into the dc supply

Both turn-on and turn-off snubber energy can be recovered into the dc supply using a dedicated buck-boost smps formed by T_{smps}, D_{smps} and L_{smps}, shown in figure 9.18. Both snubbers (capacitor C_s and inductor L_s) transfer their energy to the intermediated storage capacitor, C_o, from which the energy is smps transferred to the dc supply V_s. The buck-boost smps also maintains a fixed voltage on C_o, which facilitates rapid energy transfer of the turn-on snubber inductor L_s energy to C_o at switch T turn-off, in time $L_s I_m / V_{Co}$. The maximum switch off-state voltage is V_s+V_{Co}. At switch T turn-on, the turn-off snubber capacitor C_s energy is resonated to C_o through the loop C_s - T - C_o - ℓ - D_o, as considered in detail in section 9.3.1. The smps is operated in a discontinuous inductor current mode in order to minimise smps switch and diode losses and stresses. The maximum smps switch and diode voltages are V_s+V_{Co}.
Figures 9.18b and c show circuit versions with a reduced component count. With the inductor ℓ removed, the resonant reset current magnitude and period is now only controlled by the turn-on snubber inductor. A further diode can be removed as shown in figure 9.18c, but the number of series components in the turn-on inductor reset path is increase as is the loop inductance associated with the path.

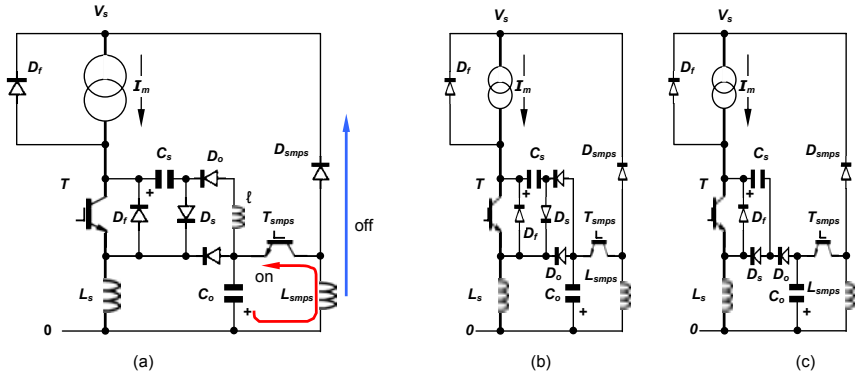


Figure 9.18. *Unified, active turn-on and turn-off snubber energy recovery circuits: (a) basic circuit and (b) and (c) reduced component variations.*

9.4 Inverter bridge legs

Capacitive turn-off snubbers (without any turn-on snubber circuit inductance), both active and passive are not normally viable on bridge legs because of unwanted capacitor discharging and subsequent uncontrolled charging current, as considered in chapter 8.4. At best capacitive soft turn-off voltage clamps (operational at $>V_s$) can be employed to reduce turn-off losses, as shown in figure 8.24.

9.4.1 Turn-on snubbers

i. Active recovery - recovery into the dc supply
Figure 9.19 shows inverter bridge legs where both switches benefit from inductor turn-on snubbers and active energy recovery circuits. The circuits also recover the energy associated with freewheel diode reverse recovery current. The turn-on energy and diode recovery energies are both recovered back into the dc supply, V_s, via a buck-boost smps. At switch turn-off, the energy stored in L_s is transferred to capacitor C_o via diode D_s.
For given turn-on snubber inductance L_s, both circuits give the same *di/dt* in the switches. The capacitor voltages determine the snubber reset time. When both circuits result in the same switch maximum voltages, the reset times are the same. But the capacitor voltages in figure 18.9a are half those for the circuit in figure 9.19b. The main operational difference between the two configurations is the periods when the capacitors are charged. In figure 9.19a, both capacitors are charged at both switch turn-on and turn-off. In figure 9.19b, each capacitor charges once per cycle, one capacitor is charged at turn-on, the other at turn-off.
Coupling of the turn-on inductors results in virtual identical waveforms as to when the inductors are not coupled. No net energy savings or gains result. Close coupling is therefore not necessary.

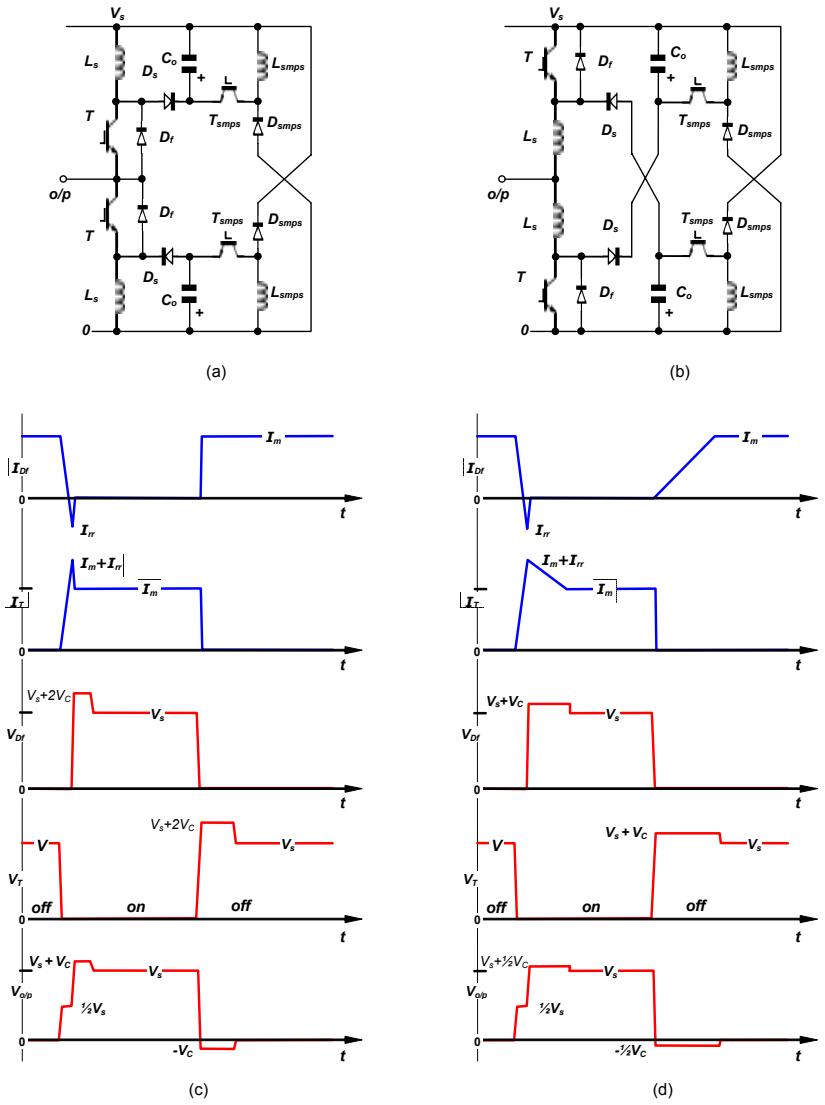


Figure 9.19. *Active inductive turn-on snubber energy recovery circuits: (a) multiple single-ended circuit; (b) cross-coupled high frequency circuit; and (c) and (d) respectively circuit waveforms.*

9.4.2 Turn-on and turn-off snubbers

i. Passive recovery - recovery into the dc supply

Figure 9.20 shows an inverter bridge leg where both switches have inductor turn-on and capacitor turn-off snubbers and passive energy recovery circuits. The circuit also recovers the energy associated with freewheel diode reverse recovery current. Both the turn-on energy and turn-off energy are recovered back into the dc supply, V_s. Although this decreases the energy transfer efficiency, recovery into the load gives poor regulation at low load current levels where the capacitor turn-off energy, which is fixed, may exceed the load requirements. Energy recovery involves a coupled magnetic circuit which can induce high voltage stresses across semiconductor devices. Such conditions can be readily avoided if a split capacitor (multilevel) voltage rail, fed from multiple secondaries, is used, as shown in figure 9.2c.

Dual snubber (inductor and capacitor) energy recovery occurs as follows.

For switch S1, the turn-off snubber is formed by C_{S1} and D_{S1}, and the turn-on snubber comprises L_{S1}.

1. The energy stored in C_{S1} is resonantly transferred to C_{o1} when switch S1 is switched on, in the path C_{S1} - D_{t1} - C_{o1} - L_{S2} - L_{S1} - S1.
2. The energy stored on C_{o1} is resonantly transferred to the dc supply V_s through transformer T1 when switch S1 is turned off and (after an underlap period) S2 is turned on (in the path C_{o1} - L_{r1} - T1 - S2).
3. When S2 is turned on, the turn-on snubber inductor L_{S1} releases its energy in parallel with capacitor C_{o1} (in the path L_{S1} - D_{s1} – D_{t1} - L_{r1} - T1 - S2 - L_{S2}).
4. The diode D_{r1} prevents (by clamping) the transfer capacitor C_{o1} from reverse charging, by providing an alternate path for the remaining energy in the resonant inductor L_{r1} to be returned to V_s via the coupling transformer T1.
5. The transformer T1 magnetising current is also returned to the dc supply V_s, thereby magnetically resetting the coupling transformer T1.

The numerical subscripts '1' and '2' are interchanged when considering the recovery processes associated with switch S2.

The recovery circuit can operate at switching frequencies far in excess of those applicable to the IGCThyristor and the high power IGBT. The limiting operational factor tends to be associated with the various snubber reset periods which specify the switch minimum on and off times. Although adequate for IGCThyristor requirements, minimum on and off times are a restriction with the IGBT.

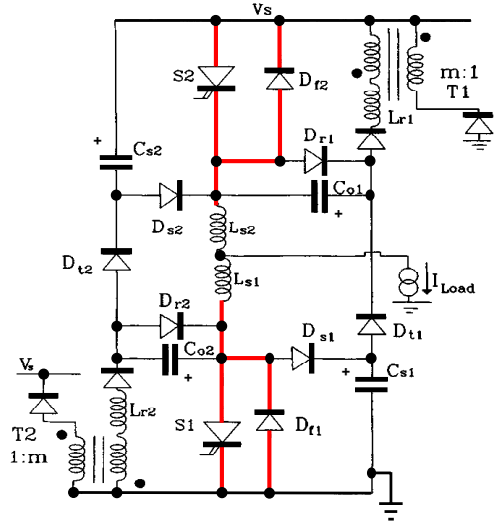


Figure 9.20. *Unified, passive snubber energy recovery circuits for GTO and GCT inverter bridge legs.*

ii. Active recovery - recovery into the dc supply

Figure 9.21 shows two similar turn-on and turn-off snubber, active energy recovery circuits, which are particularly suitable for bridge leg configurations. In figure 9.21a, the turn-on snubber section is similar in operation to that shown in figure 9.4 while the turn-off snubber section is similar in operation to that shown in figure 9.13a. A common buck-boost smps is used for each turn-on and turn-off snubber pair. This arrangement is particularly useful when the two power switches and associated freewheel diodes are available in a single isolated module package.

The active recovery circuit in figure 9.21b shows the inductive turn-on snubbers relocated. The buck-boost smps inputs are cross-coupled, serving the turn-on snubber of one switch and the turn-off snubber of the other switch.

The interaction of turn-off snubbers in both circuits can create high *L-C* resonant currents as discussed in section 8.4. In each case, two buck-boost smps and intermediate storage capacitors C_o can serve numerous bridge legs, as in a three phase inverter bridge.
Theoretically the recovery smps diodes D_r can be series connected, thereby eliminating a diode, as shown in figure 9.21c. But to do so assumes the two inductor recovery currents are both synchronised and equal in magnitude. Extra diodes, D_i are needed to divert any inductor current magnitude imbalance, as shown in figure 9.21c, which negates the diode saving in having series connected the recovery diodes D_r. Alternatively, the single inductor recovery circuit in figure 9.21d may be used provided the smps switches are not conducting simultaneously. Synchronisation of the smps switch to its associated main switch avoids such simultaneous operation. The recovery circuits in figure 9.21 parts c and d are applicable to both the bridge leg circuits in figure 9.21 parts a and b.

The circuit in figure 9.21a is readily reduced for single-ended operation, as shown in figure 9.18.

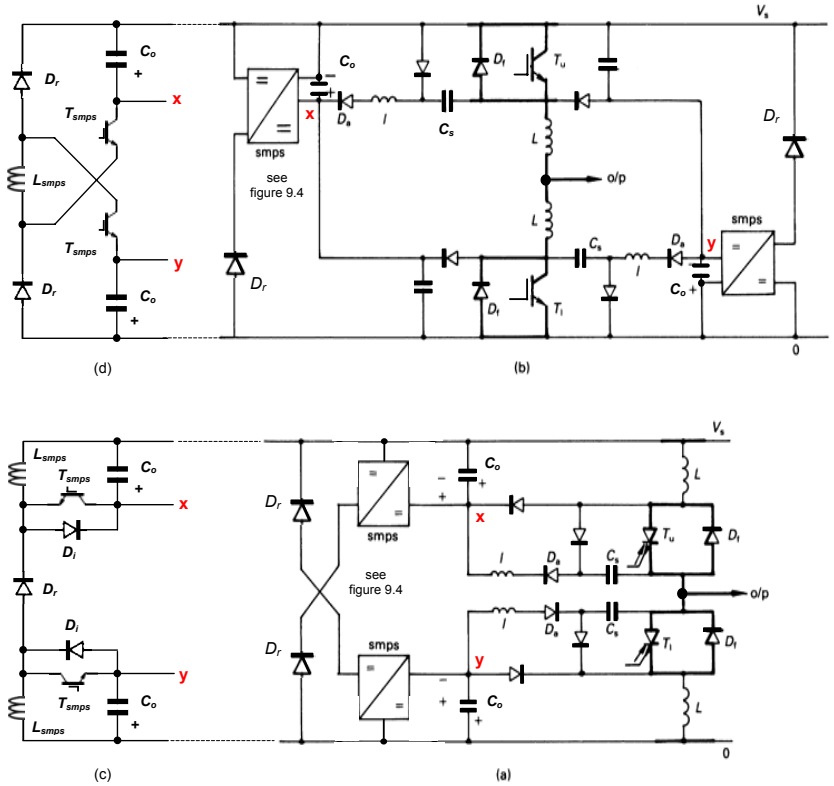


Figure 9.21. *Unified, active snubber energy recovery circuits: (a) multiple single-ended circuit; (b) cross-coupled high frequency circuit; and (c) and (d) coupled smps variations.*

9.5 Snubbers for multi-level inverters

The multi-level inverter introduced in Chapter 15.3 utilises series connected switching elements with each switch operated in a voltage clamped mode. Three multi-level inverter configurations are commonly presented

- the diode clamped multi-level inverter – see figure 15.38
- the flying capacitor clamped multi-level inverter – see figure 15.40 and
- the cascaded H-bridge multi-level inverter – see figure 15.41

9.5.1 Snubbers for the cascaded H-bridge multi-level inverter

Since the cascade multilevel inverter (see figure 15.38) is comprised of identical H-bridge modules, any of the snubbers for bridge legs considered in section 9.4 are applicable. Snubbers can be active or passive, incorporating only an inductive turn-on snubber or a capacitive soft turn-off snubber or both turn-on and turn-off snubbers. When the cascaded H-bridge approach is used for three-phase VAr compensation, real power must be returned to the ac system if the recovered energy is in excess of the inverter losses.

9.5.2 Snubbers for the diode-clamped multi-level inverter

Various snubbers have been proposed for the neutral point clamped inverter which involves a split dc rail composed of two series connected capacitors, as shown in figure 15.38. Generally devices are asymmetrically stressed or indirectly snubbered. Indirect snubbering approaches should be avoided since the main problem with high power multilevel inverters is the decoupling of circuit inductance.
For levels higher than three, only the outer switches have a fixed dc reference, viz., 0V or V_{dc}, hence recovery circuits on these switches can return energy to the outer link capacitors. Energy recovery from snubbers on the inner switches is hampered by the clamping diodes. Thus recovery of snubber energy in a three-level inverter is viable since the two link capacitors are in fact two outer capacitors, referenced to the dc rails. Recovery must be into the associated level capacitor of a given switch, if recovery circuit component voltage ratings are to be limited to that of the main switching elements.

9.5.3 Snubbers for the flying-capacitor clamped multi-level inverter

Turn-off snubbers for the flying capacitor clamped inverter are problematic since the switch clamping principle is based on indirect clamping and the level clamping capacitors support multiple-voltages in excess of the individual device operating voltage ratings. As seen in figure 15.40, the flying capacitors associated with inner switches support lower voltages than the outer capacitors.

As a general rule, if snubbering is being considered, then a series connection approach as in section 9.6 is viable, provided device switching delays are minimised. The turn-off delay of the GCThyristor can be reduced to less than 400ns if high di/dt reverse gate current drive is employed. The key limitation in reverting to series connected device operation is the loss of amplitude modulation offered by multi-level circuits. As a consequence, series connected devices produce higher output *dv/dt* voltages. The diode clamped inverter with numerous series connected devices is a favoured medium voltage compromise.

9.6 Snubbers for series connected devices

Two basic approaches are adopted when power-switching devices are series connected in order to operate circuits at voltages in excess of individual device voltage ratings.

- Use a multilevel structure as considered in Chapter 15.3, where individual switches are effectively soft clamped or
- series connect devices with fast turn-on and turn-off, minimising device switching delays thereby improving transient voltage sharing; possibly using simple R-C snubbers

The use of turn-on and turn-off snubbers greatly increases system complexity and size but does offer a method for reliably operating series connected devices, a modular structure, and the possibility of obtaining gate drive power for individual series connected cells. Fast, noise free, isolated uni/bidirectional signal transmission, without any isolation or *dv/dt* problems, to virtual any voltage potential is possible with fibre optics. The production of isolated gate drive supply power at tens, possibly hundreds of kilovolts is problematic. The usual approach for deriving emitter level supplies involves tapping energy from static voltage sharing resistors, resulting in high resistor losses, or tapping energy from the R-C snubber during switching transitions. Both methods do not provide fail-safe device operation (in the off-state, with static *dv/dt* capability) at the application of the HV dc link voltage.
The use of inductive and capacitive switching snubbers offers two advantages, other than enforcing transient voltage sharing of series connected devices, which may mitigate the associated increased cost and complexity

- Better device *I-V* utilisation and a higher switching frequency
- The derivation of cell level gate power supplies from snubber recovered energy

Many of the previously presented active snubber energy recovery circuits in this chapter are directly transferable to multilevel inverter configurations, thereby extending the current and frequency capabilities of the main switching devices, particularly the GCThyristor, and freewheel diodes. Once

snubbers are employed, traditional series device connection with snubbers is simpler than a multilevel approach, but does not offer the multilevel output voltage features (amplitude modulation and reduced *dv/dt*) of multi-level inverter configurations.
The snubber recovered energy is usually far in excess of that that can be utilised for gate drive power. The topological nature of series connected devices precludes any form of relatively simple snubber energy recovery (active or passive) other than recovery back into the dc link supply.

9.6.1 Turn-off snubber circuit, active energy recovery for series connected devices

i. Recovery into the dc supply
Series connection of switches and diodes requires static voltage sharing (resistors) and transient voltage sharing circuitry, viz., capacitive turn-off snubbers for voltage sharing during turn-off and inductive turn-on snubbers for voltage sharing during turn-on. Figure 9.22 shows series connected devices, each modular cell level incorporating a main switch and inverse parallel connected freewheel diode, plus a turn-off snubber C_s - D_s, a resonant circuit L - D_o, an intermediate energy storage capacitor C_o, and buck-boost smps recovery circuitry T_{smps} - L_{smps} - D_{smps}, as shown in figure 9.13a and considered in 9.2.2. The recovery smps is operated so as to maintain a near constant voltage on the intermediate storage capacitor C_o. The cell energy recovery switches T_{smps} are synchronised, all being turned on for up to the switch minimum on-time (immediately before the switches T are turned off), and turned off when the main switches T are turned off. The timing sequence for the control signal, switch T and recovery switch T_{smps} is shown in figure 9.22b. Note that the transmitted control signal is truncated at the switch T turn-off edge, by the switch minimum on-time, t_{delay}, which is approximately $\frac{1}{2}\pi\sqrt{LC_s}$.
When T_{smps} are turned off, the inductive stored energy in each L_{smps} is returned to the dc link through each corresponding diode D_{smps} as shown in figure 9.22a. Any imbalance in the individual inductor current magnitudes, involves currents in excess of the minimum of all the inductor currents being diverted to the cell snubber capacitor C_s through D_{smps} - C_s - D_s - L_{smps}. The inductor recovery current differentials are minimal compared to the principal current in the switches, hence do not unduly affect capacitive turn-off snubber charging, hence transient turn-off voltage balancing action.

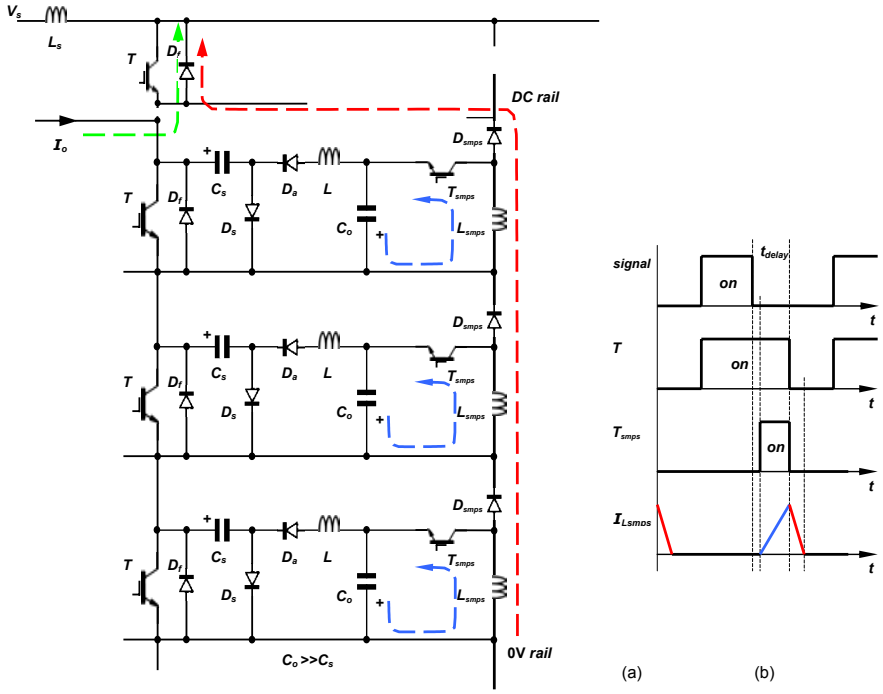


Figure 9.22. *Active turn-off snubber energy recovery for series GCT connected, inverter bridge legs: (a) modular cell circuit and (b) timing diagram.*

The turn-on snubber L_s in figure 9.22 is indirectly clamped, with the stored energy released into the series string of turn-off snubber capacitors. Link inductance is mandatory in order to control recharging of the turn-off snubber capacitors as considered in section 8.4.
Although the smps switch T_{smps} and diode D_{smps} are high voltage devices, rated at the cell voltage level, both are not particularly stressed during energy recovery switching, since the recovery buck-boost smps are operated in a discontinuous inductor current mode. The switch T_{smps} turns on with zero current, without any diode reverse recovery effects, while the diode D_{smps} suffers minimal reverse recovery, since its principal current reduces to zero controlled by L_{smps}, with recovery *di/dt* current (or voltage) controlled (or supported) by the smps inductors L_{smps}. A static voltage-sharing resistor across each cell (not shown in figure 9.22) compensates for various static voltage and current imbalance conditions on both the main switch T and smps diode D_{smps} network, particular during converter start-up and shutdown sequencing.

System start-up
The intermediate transfer stage capacitor C_o can be used to provide a source of gate level power, via a dedicated smps. One of two start-up sequences are used to build-up gate power and cell voltages before normal switching operation can commence. In both cases, an ac to dc single or three phase half-controlled converter is used to ramp charge the intermediate capacitor C_o associated with the lowest potential cell (typically C_o operates at about 50V to 100V). This capacitor C_o in turn provides gate power, via a dedicate 100V dc to ±15Vdc smps, for the lowest level switch T. By using series blocking/directing diodes, rated at the cell voltage rating, one ac to dc converter can supply the lowest potential cell of all bridge legs, as shown in figure 9.23a. Proprietary pre-charging sequences are used to charge C_o on higher cell levels, depending on whether the dc link voltage is established or not. As each C_o is progressively charged, its associated gate supply smps is self-activated, enabling external control of that switching cell. Inverter start-up can involve the application of the dc link voltage before gate level power has been established. This does not present a problem for GCThyristors, but in the case of the IGBT, a low passive impedance gate to emitter circuit is needed to avoid inadvertent device turn-on due to Miller capacitor *dv/dt* effects.

(a) Start-up with an established dc link voltage
In the case of an inverter with an established dc link voltage, each level switch, hence cell, supports half its normal operating voltage, and each snubber capacitor C_s is charged to the cell voltage level. All the intermediate energy storage capacitors C_o are discharged, except for the lowest potential cell capacitor, which has been ramp charged by the ac to dc converter. The recovery smps (and main switch) of the lowest potential cell is operable. T_{smps} of the lowest potential cell is turned on, then off and the current in the associated L_{smps} tends to overcharge C_s of the lowest potential cell. This forces current to increase through the C_o - L - D_s combinations of the higher potential cells as each C_s is forced to decrease its charge, therein charging higher-level capacitors C_o. The voltage on C_s of the lowest potential cell can be doubled before the cell reaches it normal operating voltage level. Thus for n series connected cells, the operating limit of the intermediate capacitor C_o voltage satisfies $(n\text{-}1)V_{Co} < 2V_s/n$. That is, any smps sourcing from C_o used to provide gate supply voltage rails for the main switch T, must be able to function (convert) down to a voltage level satisfied by this inequality equation.
When a cell voltage reaches its operating voltage limit, the associated main switch is turned on briefly to resonantly discharge the snubber capacitor C_s. The supported voltage is redistributed among the other cells, which typically, are only supporting half the normal cell operating voltage.

(b) Start-up with no pre-existing dc link voltage
In the case where the dc link voltage has not been established, a similar charging process is used as for the case of a pre-existing dc link voltage. The dc link capacitance must be on the inverter side of the isolation. The dc link capacitor is initially charged through series diodes D_f to the maximum cell voltage as capacitor C_s of the lowest potential cell is parallel charged from C_o by its associated recovery smps. The lowest potential recovery smps is commutated numerous time in order to charge the dc link capacitance which is usually significantly larger in capacitance than C_s. Once the link capacitor is charged to the maximum allowable cell voltage, the main switch T of the lowest potential cell is turned on to reset its associated snubber C_s voltage to zero. The start-up mechanism used with a pre-existing dc link voltage can then be used. Once C_o in each cell is charged sufficiently to enable its gate voltage smps to become operational, synchronised use of the recovery smps at each level allows charging of the dc link capacitor to the operational voltage level (in fact slightly in excess of the rectified peak level). Then the vacuum circuit breakers before the rectifier, feeding the series connected device circuit, can be closed, which results in zero line current in-rush.
Connection of the load and an interfacing filter may be problematic without dedicated contactors, as is the influence of the output filter on the cell charging mechanism previously outlined.

Other gate power derivation methods
Gate power derived from switching recovered energy cannot be maintained during prolonged standby periods. Using dropper resistors (as for static voltage sharing) to provide all gate level power

requirements results in high dissipation losses, particularly during continuous standby periods (that is, 100% dissipation duty cycle). Although resistors are used for steady-state series voltage sharing, the current associated with this mechanism (≈10mA, depending of the degree of device matching and operating temperature range) is well below that needed for gate power (≈50W for IGCThyristors but much less for IGBTs). But this level of sharing resistor current (≈10mA) may be sufficient to trickle maintain gate level supplies of cells in the off-state during prolong standby periods, using variations on the circuits shown in figure 9.23c.
Depending on the load and output filter, it may be possible during prolong standby periods to sequence the inverter between 000 and 111 states, thereby producing zero average voltage output between phases but activating the snubbers hence resonant recovery circuits that charge each C_o.
Provided sufficient switch voltage redundancy is available, sequential bootstrapping is possible where each level is boot strapped supplied from the immediate next lower level, as shown in figure 9.23b. (See figure 7.4). In the case of a positive voltage as shown in figure 9.23b, each switch, starting from the lowest level is sequentially turned on and off, thereby transferring gate energy from the lowest level to the highest level. (An expanding repetitive simultaneous on-state sequence is used, progressively involving higher potential cells.) This approach is viable in single-ending series connected switch applications. Although each bootstrap diode D_{bs} is rated at the cell level voltage, in the case of inverter legs, only half the inverter leg devices can be supplied, since any bootstrap diode bridging the pole centre take-off node must be rated at the full dc link voltage (actually ½n-1 levels can be charged since the lowest level cell is not bootstrapped).
If the bootstrapping voltage is referenced with respect to the high potential terminal of the cell, then the supply voltage on C_o is bootstrap by transferring energy from the highest potential cell down to the lowest potential cell.
A similar approach can be used with transformer isolated smps's transferring power between adjacent levels, which need only be rated at the cell level voltage. Again, this approach is viable in single-ended applications, but in the case of inverter legs, the pole output take-off node cannot be readily bridged by an smps because of the high dc link voltage blocking and isolation requirement. Also, each smps experiences *dv/dt* stresses when the level switches are commutated.
Possibly the simplest and most reliable method to derive gate power in series connected circuits, up to a few 100kV, is to use ac current transformers with series connected single-turn primaries, where each level short-circuits the secondary when not charging.

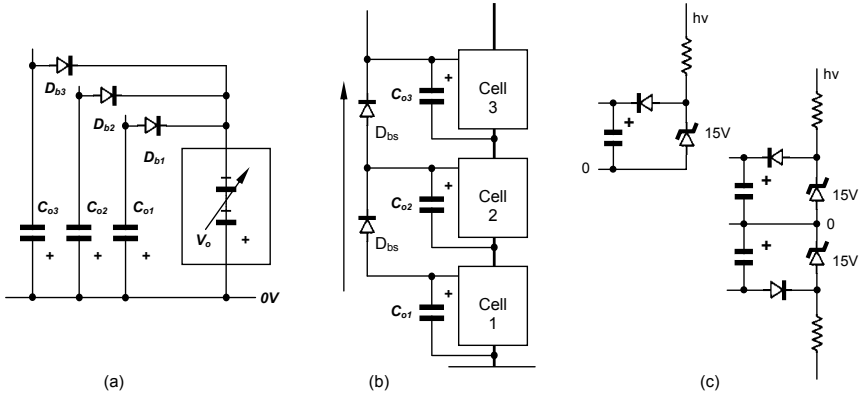


Figure 9.23. *Gate supply derivation methods: (a) ac to dc half-controlled converter for ramp pre-charging of all lowest leg level capacitors C_o; (b) bootstrapping a positive voltage supply; and (c) Zener diode based sources using static voltage sharing resistors or/and R-C snubber resistors.*

9.6.2 Turn-on snubber circuit active energy recovery for series connected devices

i. Recovery into the dc supply
An active energy recovery, inductive turn-on snubber as shown in figure 9.4 (usually with an R-C turn-off snubber), can be adapted and used at each series cell level, therein providing gate level power possibilities from C_o and energy recovery through series connect buck-boost smps recovery circuitry, as shown in figure 9.24a. The capacitor C_o is configured to be connected to the emitter of switch T_{smps}.
Energy stored in the turn-on snubber inductor L_s is transferred to the intermediate storage capacitor C_o via diode D_s at switch T turn-off. The switching sequence is shown in figure 9.24b. Each recovery smps

maintains the voltage near constant on its associated C_o and the higher this voltage the faster the inductor L_s current is linearly reset to zero, in time $t_{reset} = L_s I_m / V_{Co}$. Excess energy on C_o is transfer (recovered) to the dc link by synchronised switching of T_{smps}. Mismatched inductor L_{smps} current magnitudes and durations are diverted to charge C_o of any cell attempting to recover a lower current magnitude, by turning off all T_{smps} just before all the main switches T are turned off, as shown in figure 9.24b. This balancing effect is minimal (but does eliminate any smps diode forward recovery effects) and any current imbalance subsequently tends to overcharge the output capacitance of the main switch of the cells with recovery current in excess of the minimum of all the smps recovery currents. Some form of turn-off snubbering is therefore necessary in order to avoid excessive main switch T voltages at turn-off.
The voltage rating of the various cell circuit semiconductors is increased by the voltage on C_o. A cell static voltage sharing resistor helps maintain steady-state voltage balance of both the main switch T and the smps diode D_{smps}.

a Start-up
One ac to dc converter can be used to pre-charge each lowest level capacitor C_o of each inverter leg, as shown in figure 9.23a, provided the path to each inverter leg incorporates a series blocking/directing diode, rated at the cell voltage level. The start-up sequence, using the lowest level smps to charge higher level C_o and the dc link to the sum of all C_o voltages, is straightforward. Synchronised operation of all the smps can then gradually fully charge the dc rail, if it is not already pre-charged.

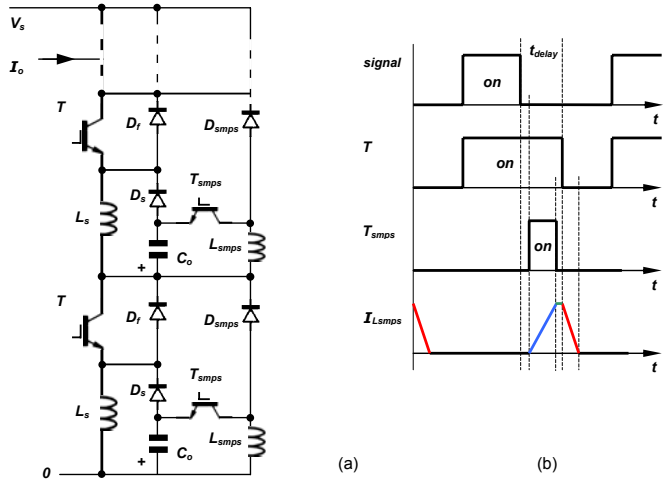


Figure 9.24. *Active turn-on snubber energy recovery for series GCT connected, inverter bridge legs: (a) modular cell circuit and (b) timing diagram.*

9.6.3 Turn-on and turn-off snubber circuit active energy recovery for series connected devices

i. Recovery into the dc supply
If a single inductive turn-on snubber L_s is used in the dc link as in figure 9.22a, its stored inductor energy at switch turn-off is transferred to the capacitive turn-off snubbers of cells supporting off-state voltage. During switching, this causes voltage ringing between the cells and the link inductor. This inductor is rated at the full dc link voltage and cannot be clamped by the usual resistor-diode parallel connected reset circuit as in figure 8.19a. This is because any reset components (*R*-D) need high voltage ratings – in excess of the dc link voltage during diode D_f reverse recovery. For this reason, an inductor snubber (possibly saturable) may be used at each cell level, giving a complete modular cell structure. Active snubber energy recovery of both inductive and capacitive energy is possible, although it may be convenient to resistively dissipated the turn-on inductive snubber energy, which is load current dependant, $½L_sI^2$.

Dual, unified active snubber energy recovery can be achieved by using the recovery circuits shown in figure 9.21b, but with the smps diodes series connected as shown in figure 9.25a. For a modular cell structure, all the cells are configured as for the lower switch in figure 9.21a. This switch configuration in figure 9.21a is preferred since capacitor C_o can be readily pre-charged to initiate the start-up sequence for charging higher level C_o, which can be used to derive gate level power for the associated cell. A relatively low voltage on capacitor C_o (if C_o operates at about 5 to 10% of the cell operating voltage) may necessitate a long switch T minimum off-time in order to ensure reset of the turn-on inductor current to zero. This is not a problem for GTO type devices which have minimum on and off time limitations. Higher operating voltages for C_o necessitates a more complicated smps to derive gate level power for switch T. At higher cell operating voltages, the intermediate storage capacitor C_o can be modified to the circuit in figure 9.25b. The low voltage output *lv* can be used to power cell start-up circuitry.
The resonance inductor ℓ (in series with the turn-on snubber inductance L_s) is used to control the magnitude and duration of the resonant period of C_s transferring its charge to C_o. The minimum value of inductor ℓ can be zero if L_s is large enough to satisfactorily control resonant reset circuit conditions without ℓ. A further simplification can be made by removing a resonant circuit diode as shown in figure 9.25c, which is derived from the circuits in figure 9.18.
The timing sequence in figure 9.22b for turn-off snubbers is used.
One functional design constraint should be observed. At switch turn-on, current builds up in L_{smps} because of the voltage on L_s, during the later part of the cycle when C_s resonates its charge to C_o. This relatively small current magnitude linearly increases to a magnitude dependant on the relative magnitudes of L_s to ℓ and L_{smps}, and the magnitude of the voltage retained on C_o. Once established, a near constant, slowly decreasing current flows in a zero voltage loop, L_{smps} - D_{smps} - T - L_s, and is recovered during recovery smps action at switch turn-off.

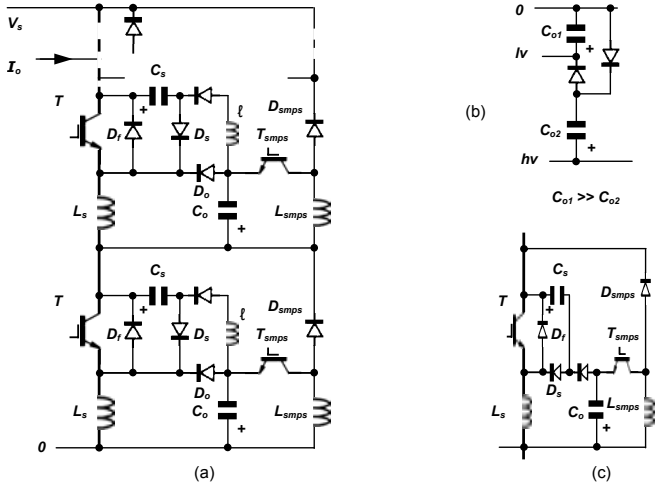


Figure 9.25. *Active turn-on and turn-off snubber energy recovery: (a) circuit for series GCT and IGBT inverter bridge legs; (b) high voltage replacement circuit for C_o; and (c) reduced component variation of part a.*

(a) Start-up
The capacitor C_o of the lowest potential cell (in each bridge leg) is negatively ramp charged by a dedicated ac to dc converter as shown in figure 9.23a. This establishes cell internally generated gate supply power and hence external control of both switches of the lowest potential cell.
The recovery smps of the lowest potential cell is operated in a discontinuous mode, which charges up the turn-off snubber capacitor C_s of that cell. Simultaneously current flows in three other parallel paths, tending to charge up the dc link capacitor, viz.

- the series connected L_{smps} - D_{smps}
- the series connected L_s - D_f
- the series connected C_o - D_o - D_f

Thus provided the smps of the lowest cell delivers a high current, each C_o receives charge before the current is diverted and built up in inductors L_{smps} and L_s. The dc link capacitor simultaneously receives charge. The switch T_{smps} on-time, hence its current, is not restricted during the start-up procedure. Once

gate power, hence external control is established on each cell, judicious operation of each smps and main switch T can facilitate charging of the dc link capacitor and contains all cell voltages to within the rated cell voltage.
The start up mechanism may necessitate a suitable diode connected in series or anti-parallel with T_{smps}.
(b) Shut down
After the dc link has been isolated, under zero inverter output current conditions, using a vacuum circuit breaker on the ac side, the intermediate capacitor of the lowest potential cell (in each bridge leg) is maintained in a partially discharged state by a resistive load which is switch connected to the capacitor C_o of the lowest potential cell. The auxiliary ac to dc converter used to initially charge C_o is disabled during normal operation and shut-down, with all the ac to dc converter thyristors off, therefore blocking current in both directions. Alternatively, if this ac to dc converter has suitable two quadrant operational modes, then the energy continually being transferred to C_o from other cells, can be recovered into the low voltage ac source. The various smps and main switches are operated so as to maintain equal voltage across all cells (by sequentially commutating each main switch on then off), gradually decreasing the dc link voltage as energy is continually, but controlled, being transferred to and removed from the lowest potential cell capacitor C_o.

9.6.4 General active recovery concepts for series connected devices

In each of the three snubber circuits considered for series connected devices, the common key recovery mechanism is performed by a buck-boost smps, with components rated at the cell voltage level.
Figure 9.26 shows two basic underlying recovery techniques for transferring energy from C_o through an inductor, into the dc supply at a higher potential. The key difference between the two techniques is the polarity orientation of the energy source C_o and the dc supply V_s, with respect to their common node.

- Figure 9.26 parts a and b show boost converters, where energy is drawn from C_o when energy is being delivered to the supply V_s, via an inductor.
- Figure 9.26 parts c and b show buck-boost converters, which do not involve C_o during the period when energy is being delivered to the supply V_s, via an inductor.

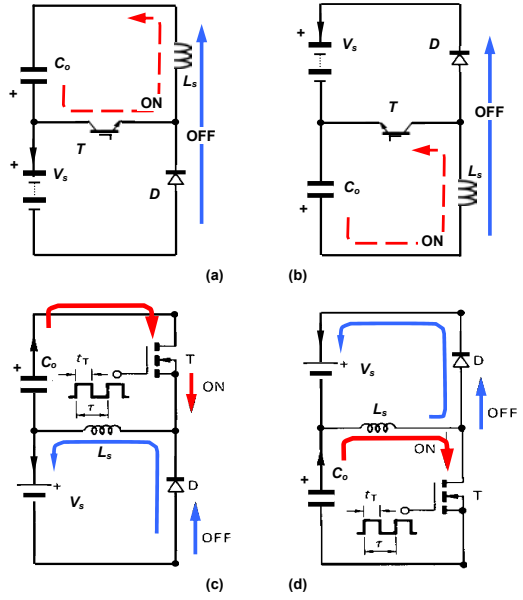


Figure 9.26. *Underlying energy recovery circuits when energy in C_o is stored at different potentials: (a) and (b) boost smps recovery and (c) and (d) buck-boost smps recovery.*

A common requirement is that an smps output (whether inductor-diode for buck-boost and inductor-diode-C_o for boost) span a cell, thereby inherently interconnecting in series any number of cells. Each intermediate storage capacitor C_o must therefore be connected to one cell terminal. To confine further

the possibilities, it is unlikely that C_o referenced with respect to the cell collector will yield a useful active recovery circuit. If the capacitor C_o is referenced with respect to the switch collector/anode, C_o undergoes high *dv/dt* voltages with respect to the switch gate. This complicates any smps using the stored capacitor C_o energy for gate drive purposes. The polarity orientation of C_o and the recovery smps components are therefore restricted to the four possibilities shown in figure 9.27. Series recovery assumes the smps inductors conduct an identical instantaneous maximum magnitude and same duration current.

(a) Start up
The general cell structures and their recovery smps can inherently be used to charge other series connected cells and the dc link, and to provide a dc source (the intermediate storage capacitor C_o) from which to derive cell level power supplies for the gate level circuitry. Specific proprietary switching sequences are required at start-up, depending on the cell circuit arrangement, the output filter and load, the dc link and ac rectifier input arrangement and initial conditions.

(b) Shut down
At shut down, once the inverter is in standby, the dc link supply is isolated (by opening the ac side vacuum circuit breakers) under zero current conditions, then the dc link voltage is cyclically discharged into the load via the series connected cells. Such link discharge using cell switching sequences is problematic when

- each cell voltage reaches a level where C_o falls below a level to maintain operation of the smps used to provide gate level voltage which allows the cell switches to operate; or
- cells in another inverter legs cease to operate sooner.

Such limitations are mitigated by ensuring the smps that operates across C_o has a wide (low minimum bounds) input voltage operating range.
If the load is isolated at shut down, then the dc link energy can be sequentially transferred to C_o of the lowest potential cell in each leg and dissipated in a single ended resistive dumping circuit or recovery from C_o via the ac to dc converter (fully controlled) used during the start-up sequence, as shown in figure 9.23a. The sequence involves progressively, but sequentially, not using higher-level recovery smps.
Fail-safe start-up and shut down sequencing, so as not to over-volt any cell, usually require cell operational coordination. The fibre optic communications link for cell level on/off control of the main switch T, is therefore bidirectional.

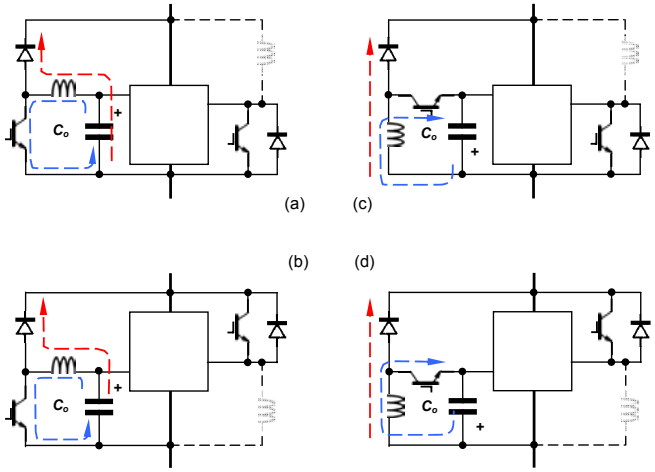


Figure 9.27. *Cell active energy recovery from C_o with:*
(a) and (b) a boost converter and (c) and (d) a buck-boost converter.

9.7 Snubber energy recovery for magnetically-coupled based switching circuits

Coupled circuits can induced circuit and in particular switch voltages that exceed the supply voltage. These increased voltages are associated with two factors:

- leakage or uncoupled inductance energy release
- time-displaced energy-transfer coupled-circuits, as with the buck-boost converter or coupled voltages as with push-pull centre tapped transformer circuits

Both factors come into operation with the two buck-boost isolated output converters shown in figure 9.28. When energy is drawn by the coupled circuit secondary, a voltage is induced into the primary, increasing the voltage experience by the switch in the off-state. Energy associated with leakage inductance further increases the switch T voltage. If a basic R-C-D turn-off snubber is used, the capacitor stored energy is increased from $\frac{1}{2}C_s V_s^2$, if the switch voltage were to be limited to V_s, to in excess of $\frac{1}{2}C_s\left(V_s + V_o/N\right)^2$, where N is the transformer turns ratio as defined in figure 9.28. The leakage energy adds to the voltage component.

9.7.1 Passive recovery
Figure 9.28a shows a passive turn-off snubber energy recovery configuration for an isolated buck-boost converter. It is based on the circuit in figure 9.31j, where the transformer leakage inductance, L_ℓ, is effectively the turn-on snubber inductance.
When the switch T is turned off, the snubber capacitor C_s charges from - V_s to a voltage v_o/N, controlled by the leakage inductance L_ℓ which causes the capacitor C_s to charge to a higher voltage. Turn-off capacitor C_s snubbering of the switch is achieved indirectly, through the dc supply V_s.
At switch T turn-on, the charge on C_s resonates in the loop C_s - T – L_r - D_r, reversing the polarity of the charge on C_s. This reverse voltage is clamped to V_s, as the diode D_s conducts and the remaining energy in L_s is transferred (recovered) to the dc supply V_s. The switch minimum on-time is $\frac{1}{2}\pi\sqrt{L_r C_s}$, whilst the energy recovered from L_r to V_s occurs independent of the state of the switch.
At switch T turn-off, after snubber capacitor C_s is fully charged, an oscillation can occur through L_r - D_r - C_s and the transformer primary back into the supply V_s. Although a lossless oscillation, it can effect the output voltage regulation, increase output rectifier recovery losses, but can be prevented by using a series switch in the L_r - D_r path as shown in figure 9.28d. Then recovery occurs during switch T on-period, back into the supply V_s, without affecting the output regulation. Once a switch has been used, other active recovery possibilities may be more attractive.
The same leakage voltage control and recovery technique can be used on the push-pull converter in figure 9.28c, where two recovery circuits are used.

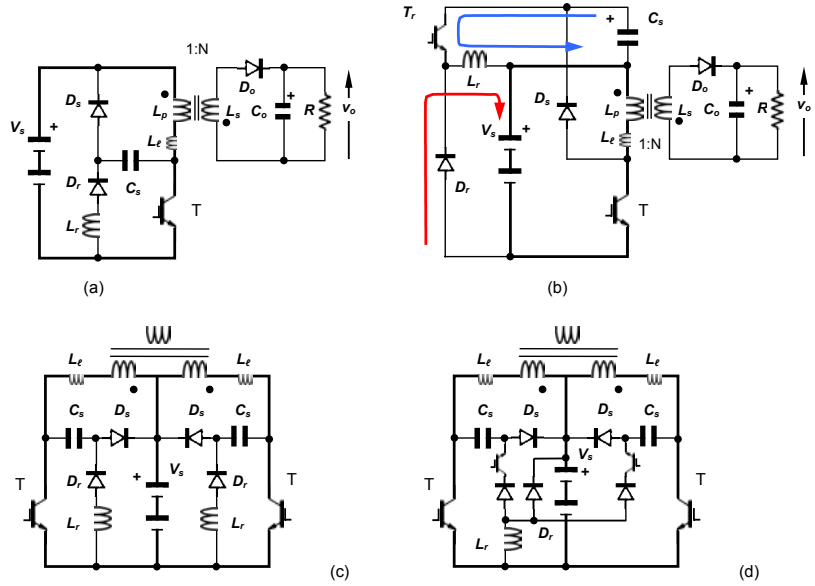


Figure 9.28. *Recovery of leakage inductance energy:*
(a) and (c) passive and (b) and (d) active recovery.

9.7.2 Active recovery

Figure 9.28b show the circuit of an active turn-off soft snubber energy recovery configuration.
Coupled circuit leakage inductance L_t energy is transferred to the intermediate storage capacitor C_s via D_s at switch turn-off. The voltage on C_s is maintained at a voltage related to v_o/N by the buck-boost smps formed by T_r, L_r and D_r, which returns leakage energy to the dc supply V_s. The circuit function is to clamp the switch voltage rather than to perform a turn-off snubber action.
The maximum switch voltage is near constant, where as the voltage experience by the switch at turn-off in figure 9.28a, although variable, is snubbered, but dependant of the output voltage v_o. In both circuits, an R-C snubber may be required across the switch T since the recovery snubber circuits do not decouple stray inductance not associated with the coupled magnetic circuit.

Similar active snubber or clamping circuits can be used with push-pull converters which utilise a centre-tapped transformer, as in figure 9.28d, where, with a full-wave rectified forward converter secondary circuit, the overvoltage is independent of the transformer turns ratio. The recovery circuit switches prevent undesirable lossless oscillations after main switch turn-off, particularly when the switch duty cycle is less than 50%. The diode D_r allows the active recovery switches to be activated with the same control signal timing as the corresponding main switch T, provided the switch minimum on-time is at least $\frac{1}{2}\pi\sqrt{L_r C_s}$. In the active recovery form, only one common reset inductor L_r is necessary.

9.8 General passive snubber energy recovery concepts for single-ended circuits

Snubbers are used for stress reduction at

- switch turn-on - involving series inductance
- switch turn-off – involving shunt capacitance
- freewheel diode recovery - involving series inductance

and the snubber may incorporate more than one of these stress arresting functions.
A single ended switching circuit usually incorporates a switch T, a freewheel diode D_f and an inductive load, where the load may be configured to be in

- the emitter/cathode circuit of T or
- the collector/anode circuit of T.

The input energy source, the switch, diode and load may be configured to perform any of the following functions

- forward converter
- buck converter
- boost converter or
- buck-boost converter

The differentiation between the forward converter and the buck converter is that the inductive element is part of the active load in the case of the forward converter.

Figure 9.29a shows a switch-diode and inductor circuit combination, assuming a collector load circuit, which can be configured as any type of converter viz., forward, buck, boost, etc. Equivalent emitter load circuits, as well as collector loadings, are shown in figures 9.30 and 9.31, which present systematically a more complete range of circuit possibilities, in each case, all with exactly the same functional snubber circuit.
Energy recovery into the load is usually associated with a parallel connected capacitor discharging (instantaneous change in capacitor current to match the load current is possible) while recovery back into the source is usually associated with a parallel connected inductor or magnetically coupled circuit releasing its energy (instantaneous change in inductor terminal voltage to equal the supply voltage is possible).
Ac and dc circuit theory allows all these circuit configuration combinations to be generalised. This is because a snubber is an ac circuit – performing a transient function - while the source and load tend to be dc components (constant voltage and constant current sources respectively). Therefore it is possible to interchange the connections of the snubber (an ac circuit) with the connections to the dc voltage source, since ac-wise, a dc source appears as a short circuit. The snubber function can be achieved directly (across the switch) or indirectly (assuming a well decoupled supply).
An operational mechanism to be appreciated is the topological relative orientation within the principal circuit of the turn-on snubber inductor or turn-off snubber capacitor.

Turn-off snubber - capacitor:
Circuits in figure 9.29c and d show the turn-off snubber D_s - C_s combination parallel to the switch (direct snubbering) or alternatively connected across the freewheel diode to the dc rail (indirect snubbering). AC circuit wise these are the same connection since the dc source can be considered as a short circuit at high frequency. When D_s - C_s are parallel connected to the switch (direct snubbering), the capacitor charges as the switch voltage rises at turn-off, while in the case of the snubber being across the freewheel diode (indirect snubbering), the capacitor discharges, and by Kirchhoff's voltage law, the switch voltage is indirectly controlled to be the difference between the capacitor voltage and the source voltage. Practically it is preferred to place the D_s - C_s snubber directly across the element to be protected, the switch, since the source may not be well decoupled.

Turn-on and diode reverse recovery snubber - inductor:
Circuits in figure 9.29a and b show the inductor L configured such that the snubber turn-on inductor is in series with the switch (direct snubbering) or alternatively in series with the freewheel diode (indirect snubbering). Both arrangements perform the same function at switch turn-on. Assuming a constant current in the inductor L, by Kirchhoff's current law, whether the turn-on inductor controls the rate of rise of current in the collector (direct snubbering) or rate of current fall in the diode (indirect snubbering), the complementary element has its current inversely controlled.

Figure 9.30 shows variations of a snubber for recovering the energy associated with freewheel diode reverse recovery. All twelve circuits have the same functional ac operating mechanism, although a number have been published – even patented - as different. US patent 5633579, 1997, according to the three claims, explicitly covers the boost converter snubber circuit in figure 9.30a. In protecting the specific boost converter circuit, all the other topological variations are inadvertently and unwittingly implicitly precluded. Although a highly skilled expert in the art, Irving, IEEE APEC, 2002, published the next recovery circuit, figure 9.30b, as a new diode recovery snubber for the boost converter.
Passive inductive turn-on snubber energy recovery circuit variations are shown in figure 9.31, for collector and emitter connected buck, boost, forward and buck boost converters. Six versions exist with the circuitry in each of the switch emitter and collector circuits.
Figure 9.32 shows turn-off and turn-off plus turn-on passive recovery circuit variations. The circuitry can be in the emitter or collector (as shown) circuit.

Reading list

Boehringer, A. *et al.,* '*Transistorschatter im Bereich hoher Leistungen und Frequenzen*', *ETZ,* Bd. 100 (1979) pp. 664-670.

Peter, J. M., *The Power Transistor in its Environment,* Thomson-CSF, Sescosem, 1978.

Williams, B. W., *et al*., '*Passive snubber energy recovery for a GTO thyristor inverter bridge leg*', *Trans. IE IEEE,* Vol. 47, No. 1, Feb. (2000) pp. 2-8.

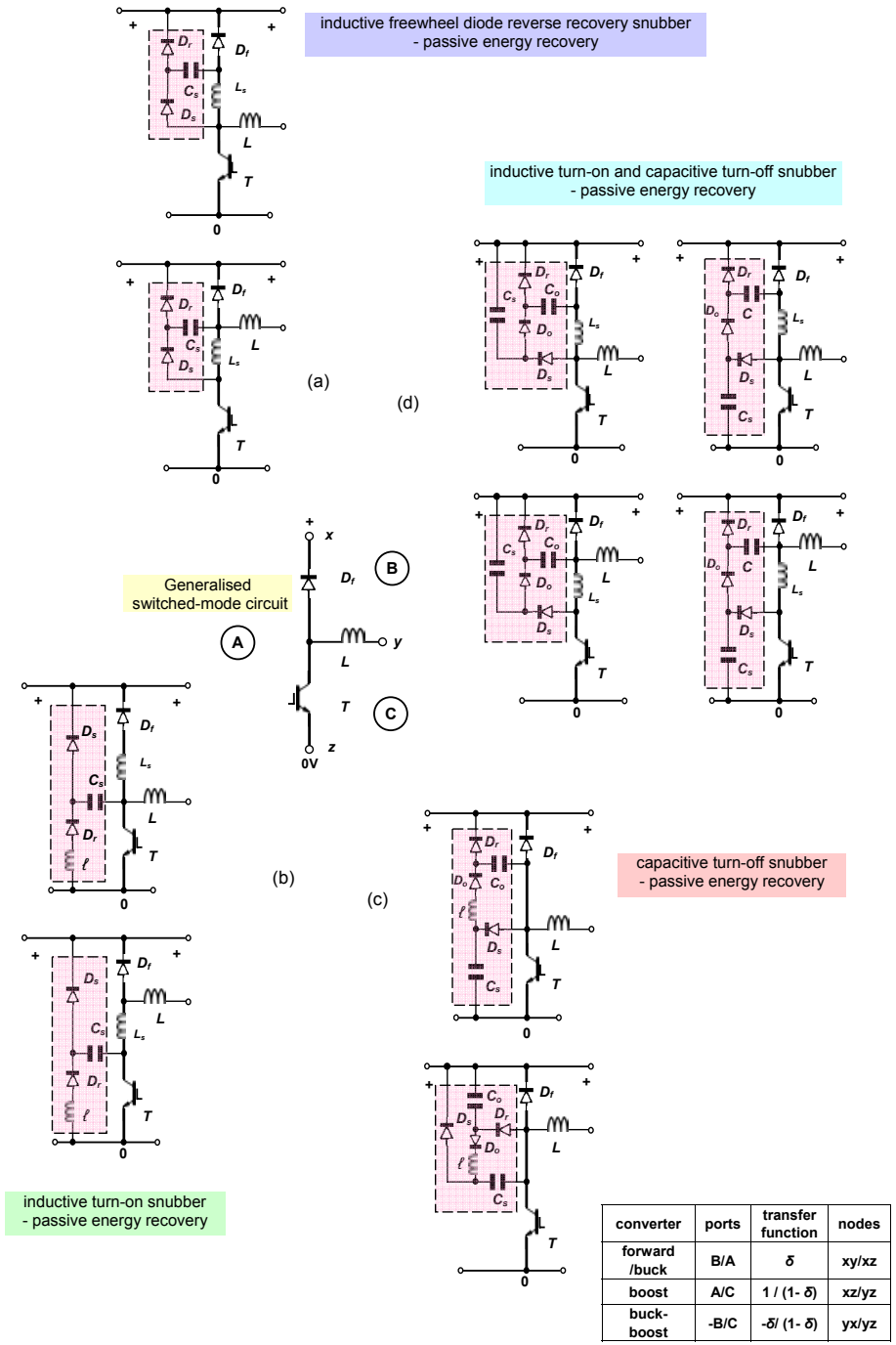
Williams, B. W., '*High-voltage high-frequency power-switching transistor module with switching-aid-circuit energy recovery*', *Proc. IEE, Part B,* Vol. 131, No. 1, (1984) pp. 7-12.

Finney, S. J. *et al*., '*High-power GTO thyristor chopper applications with passive snubber energy recovery*', *Proc. IEE, EPA,* Vol. 144, No. 6, (1997) pp. 381-388.

Problems

9.1. For the circuit in Figure 9.14a show that the upper current limit for total energy recovery is given by $\frac{1}{2}L_s I_m^2 \le \frac{1}{2}C_s V_s^2$.

9.2. Derive capacitor C_s voltage and current equations which describe the operation of the turn-off snubber energy recovery circuit in figure 9.13. Assume the storage capacitor C_o to be an ideal voltage source with polarity as shown.



converter	ports	transfer function	nodes
forward /buck	B/A	δ	xy/xz
boost	A/C	1 / (1- δ)	xz/yz
buck-boost	-B/C	-δ/ (1- δ)	yx/yz

Figure 9.29. *Snubber energy recovery circuits for generalised switch-diode-inductive element circuit.*

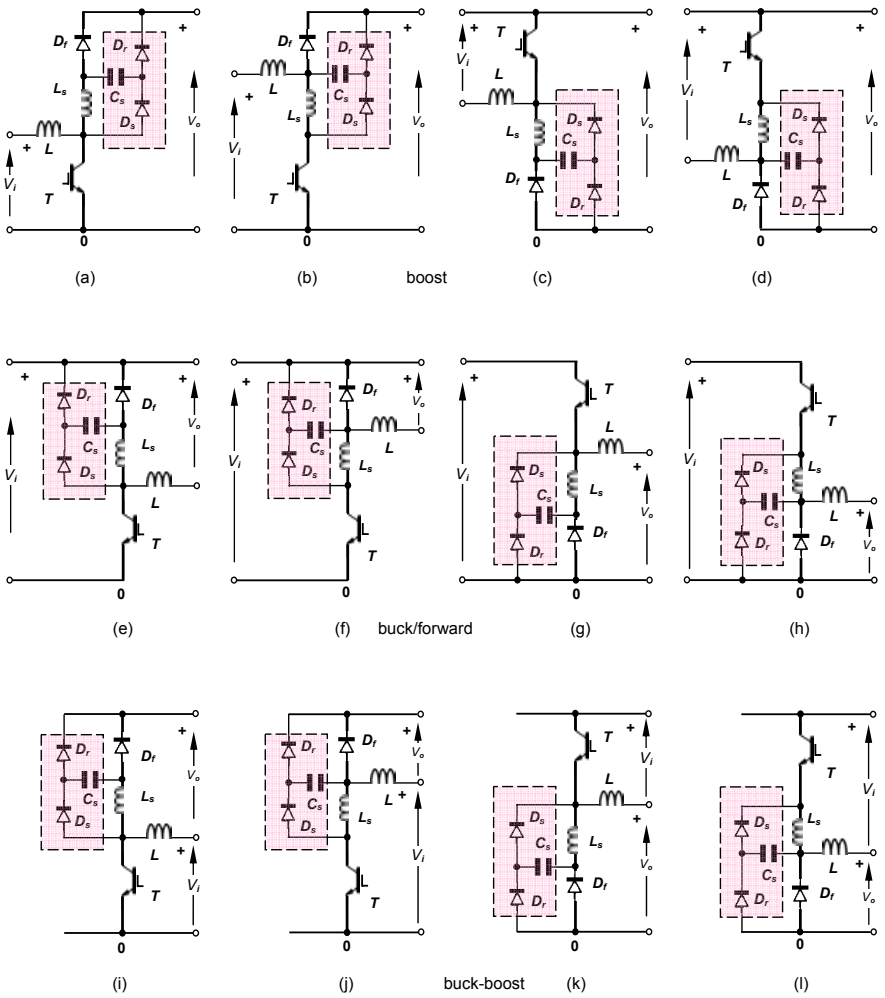


Figure 9.30. *Passive energy recovery of freewheel diode recovery energy: (a)-(d) a boost converter; (e)-(h) a buck/forward converter; and (i)-(l) a buck-boost converter.*

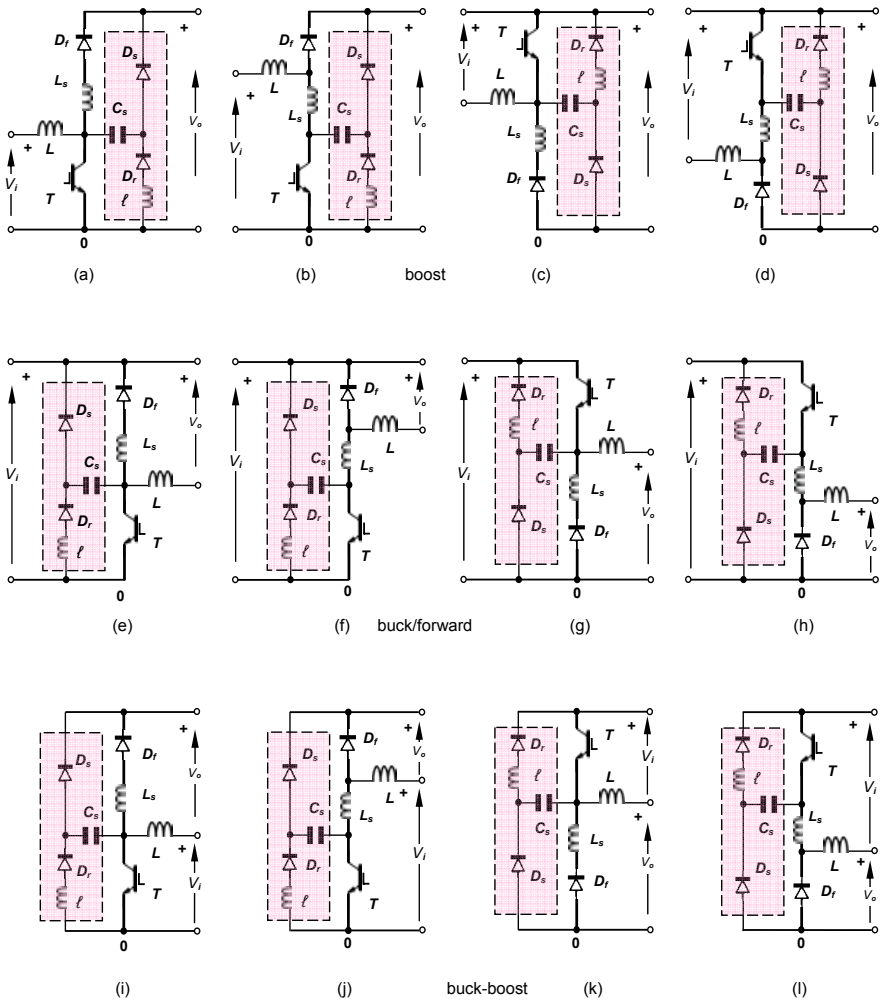


Figure 9.31. *Passive energy recovery for inductive turn-on snubber: (a)-(d) a boost converter; (e)-(h) a buck/forward converter; and (i)-(l) a buck-boost converter.*

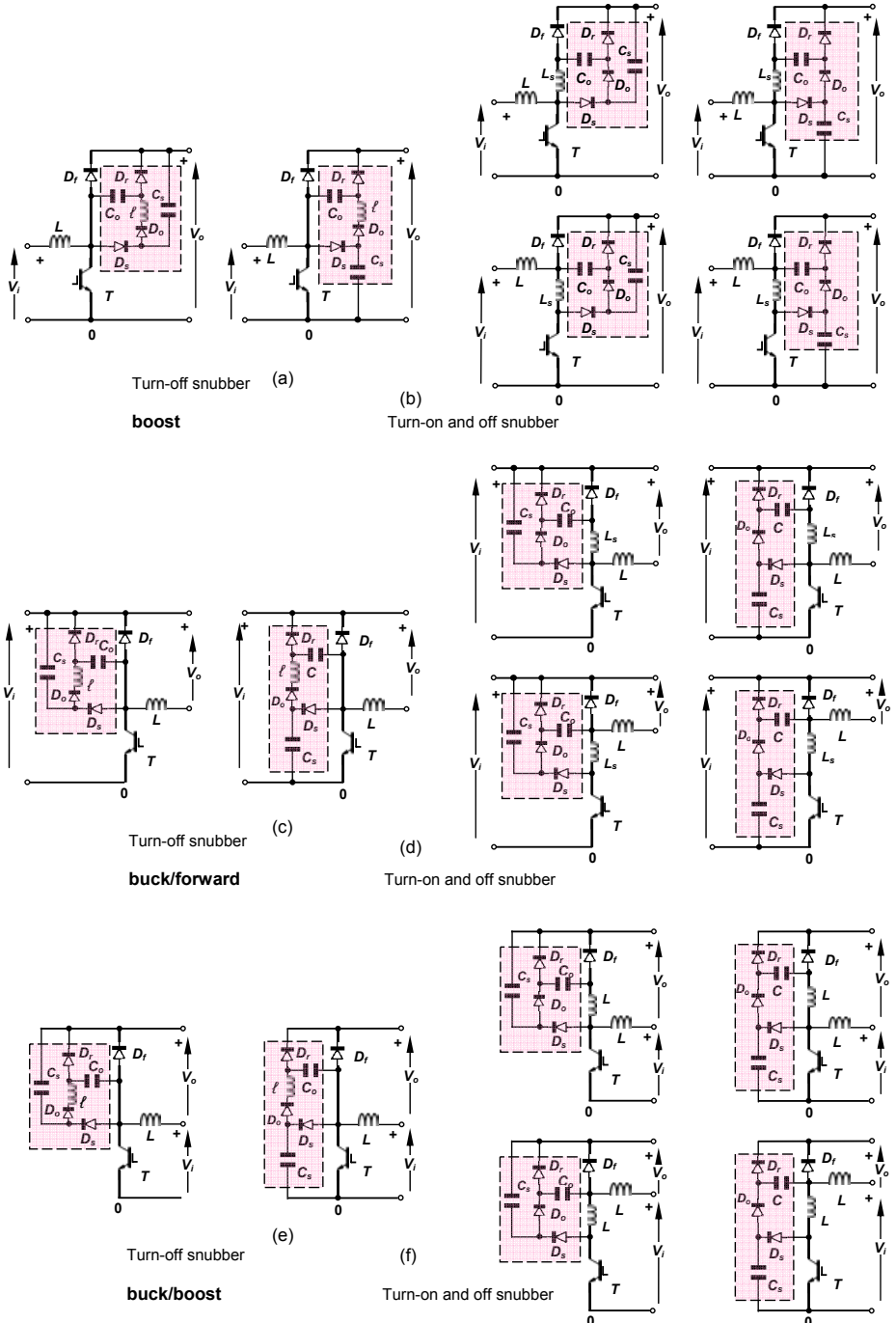


Figure 9.32. *Passive energy recovery circuits for the capacitive turn-off snubber and both turn-on and turn-off snubber circuits, for the different types of switched mode converters.*

10

Device Series and Parallel Operation, Protection, and Interference

This chapter considers various areas of power device application that are often overlooked, or at best, underestimated. Such areas include parallel and series device utilisation, over-current and overvoltage protection, radio frequency interference (rfi) noise, filtering, and interactive noise effects.

10.1 Parallel and series connection and operation of power semiconductor devices

The power-handling capabilities of power semiconductor devices are generally limited by device area utilisation, encapsulation, and cooling efficiency. Many high-power applications exist where a single device is inadequate and, in order to increase power capability, devices are paralleled to increase current capability or series-connected to increase voltage ratings. Extensive series connection of devices is utilised in HVDC transmission thyristor and IGBT modules while extensive paralleling of IGBTs is common in inverter applications. Devices are also series connected in multilevel converters.
When devices are connected in series for high-voltage operation, both steady-state and transient voltages must be shared equally by each individual series device. If power devices are connected in parallel to obtain higher current capability, the current sharing during both switching and conduction is achieved either by matching appropriate device electrical and thermal characteristics or by using external forced sharing techniques.

10.1.1 Series semiconductor device operation

Owing to variations in blocking currents, junction capacitances, delay times, on-state voltage drops, and reverse recovery for individual power devices, external voltage equalisation networks and special gate circuits are required if devices are to be reliably connected and operated in series (or parallel).

10.1.1i - Steady-state voltage sharing

Figure 10.1 shows the forward off-state voltage-current characteristics of two power switching devices, such as SCRs or IGBTs. Both series devices conduct the same off-state leakage current but, as shown, each supports a different voltage. The total voltage blocked is $V_1 + V_2$ which can be significantly less than the sum of the individual voltage capabilities. Forced voltage sharing can be achieved by connecting a resistor of suitable value in parallel with each series device as shown in figure 10.2.

These equal value sharing resistors will consume power and it is therefore desirable to use as large resistance as possible. For worst case analysis consider n cells in series, where all the cells pass the maximum leakage current except cell D_1 which has the lowest leakage. Cell D_1 will support a larger blocking voltage than the remaining n - 1 which share voltage equally.
Let V_D be the maximum blocking voltage for any cell which in the worst case analysis is supported by D_1. If the range of maximum rated leakage or blocking currents is from $\hat{I}_b$ to $\check{I}_b$ then the maximum imbalance occurs when member D_1 has a leakage current of $\check{I}_b$ whilst all the remainder conduct $\hat{I}_b$.

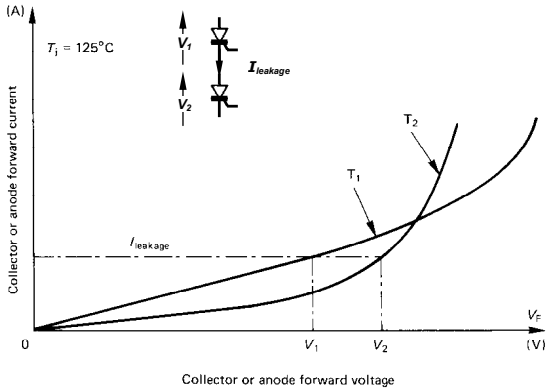


Figure 10.1. *Collector (transistor) or anode (thyristor) forward blocking I-V characteristics showing voltage sharing imbalance for two devices in series.*

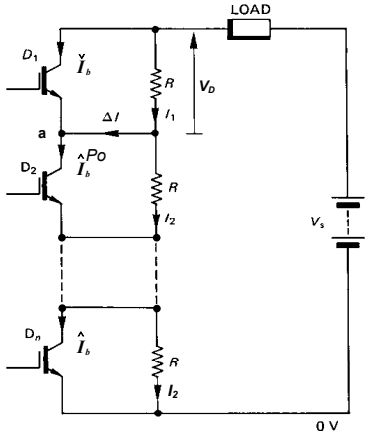


Figure 10.2. *Series IGBT string with resistive shunting for sustaining voltage equalisation in the off-state.*

From figure 10.2, Kirchhoff's current law at node 'a', gives

$$\Delta I = \hat{I}_b - \check{I}_b \quad \text{(A)} \tag{10.1}$$

$$= I_1 - I_2 \quad \text{(A)} \tag{10.2}$$

where $I_1 > I_2$. The voltage across cell D_1 is

$$V_D = I_1 R \quad \text{(V)} \tag{10.3}$$

By symmetry and Kirchhoff's voltage law, the total string voltage to be supported, V_s, is given by

$$V_s = (n - 1) I_2 R + V_D \quad \text{(V)} \tag{10.4}$$

Eliminating ΔI, I_1, and I_2 from equations (10.1) to (10.4) yields

$$\hat{R} \le \frac{nV_D - V_s}{(n-1)\left(\hat{I}_b - \check{I}_b\right)} \quad \text{(ohms)} \tag{10.5}$$

for $n \ge 2$.
Generally only the maximum leakage current at rated voltage and maximum junction temperature is specified. By assuming $\check{I}_b$ = 0, a conservative value of the maximum allowable resistance is obtained, namely

$$\hat{R} \le \frac{nV_D - V_s}{(n-1)\hat{I}_b} = \frac{n(1-k_s)V_D}{(n-1)\hat{I}_b} \quad \text{(ohms)} \tag{10.6}$$

The extent to which nV_D is greater than V_s, is termed the *voltage sharing factor*, namely

$$k_s = V_s / nV_D \le 1 \qquad (10.7)$$

As the number of devices is minimized the sharing factor approaches one, but equation (10.5) shows that undesirably the resistance for sharing decreases, hence losses increase.
The power dissipation of the resistor experiencing the highest voltage is given by

$$\hat{P}_d = V_D^2 / \hat{R} \quad \text{(W)} \qquad (10.8)$$

If resistors of ± l00a per cent resistance tolerance are used, the worst case occurs when cell D_1 has a parallel resistance at the upper tolerance while all the other devices have parallel resistance at the lower limit. After using $V_D = (1+a)I_1R$ and $V_s = (n-1)\times(1-a)I_2R + V_D$ for equations (10.3) and (10.4), the maximum resistance is given by

$$\hat{R} \le \frac{n(1-a)V_D - (1+a)V_s}{(n-1)(1-a^2)\hat{I}_b} \quad \text{(ohms)} \qquad (10.9)$$

for $n \ge 2$.

The maximum loss in a resistor is

$$\hat{P}_D = V_D^2 / \hat{R}(1-a) \qquad (10.10)$$

If the dc supply toleration is incorporated, then V_s in equations (10.6) and (10.9) is replaced by $(1+b)\times V_s$ where +100*b* is the supply percentage upper tolerance. This leads to a decreased resistance requirement, hence increased resistor power losses.

$$\hat{R} \le \frac{n(1-a)V_D - (1+a)(1+b)V_s}{(n-1)(1-a^2)\hat{I}_b} \quad \text{(ohms)} \qquad (10.11)$$

The effects and importance of just a few per cent resistance or supply voltage tolerance on the maximum value for the sharing resistors and their power losses, are illustrated by example 10.1.

Example 10.1: ***Series device connection – static voltage balancing***

Ten, 200 V reverse-blocking, ultra fast 35 ns reverse recovery diodes are to be employed in series in a 1500 V dc peak, string voltage application. If the maximum device reverse leakage current is 10 mA (at maximum junction temperature) calculate the voltage sharing factor, and for worst case conditions, the maximum value of sharing resistance and power dissipation.

i. If 10 per cent tolerance resistors are employed, what is the maximum sharing resistance and its associated power rating?
ii. If a further allowance for supply voltage tolerance of ±5% is incorporated, what is the maximum sharing resistance and its associated power rating?

Solution

When n = 10, V_D = 200 V dc, V_s = 1500 V dc, and $\hat{I}_b$ = 10mA, the voltage sharing factor is $k_s = 1500\text{V}/10\times200\text{V} = 0.75$. Equation (10.6) yields the maximum allowable sharing resistance

$$\hat{R} \le \frac{nV_D - V_s}{(n-1)\hat{I}_b} = \frac{10\times200\text{V} - 1500\text{V}}{(10-1)\times10\text{mA}} = 5.55\text{k}\Omega$$

The nearest (lower) preferred value, 4.7kΩ, would be used.

Maximum resistor power losses occur when the diodes are continuously blocking. The maximum individual supporting voltage appears across the diode which conducts the least leakage current. Under worst case conditions this diode therefore supports voltage V_D, hence maximum power loss $\hat{P}_D$ is

$$\hat{P}_D = V_D^2 / \hat{R}$$
$$= 200\text{V}^2/4700\Omega = 8.5\text{ W}$$

Since the worse device, (in terms of sharing has lowest leakage current), is randomly located in the string, each 4.7kΩ resistor must be capable of dissipating 8.5W.

The maximum 1500V dc supply leakage current is 42.5mA (10mA+1500V/10×4.7kΩ) giving 63.8W total losses (1500V×42.5mA), of which 15W (10mA×1500V) is lost in the diodes.

i. If 10% resistance tolerance is incorporated, equation (10.9) is employed with ***a* = +**0.1, that is

$$\hat{R} \le \frac{n(1-a)V_D - (1+a)V_s}{(n-1)(1-a^2)\hat{I}_b}$$
$$\hat{R} \le \frac{10\times(10-0.1)\times200\text{V} - (1+0.1)\times1500\text{V}}{(10-1)\times(1-0.1^2)\times10\text{mA}}$$
$$= 2.13\text{ k}\Omega$$

The nearest (lower) preferred value is 1.8kΩ, which is much lower resistance (higher losses) than if closely matched resistors were to be used.
Worst case resistor power dissipation is

$$\hat{P}_D = V_D^2 / \hat{R}(1-a)$$
$$= 200\text{V}^2/1800\Omega\times(1-0.1)$$
$$= 27.7\text{ W}$$

The maximum total module losses are 165W (1500V×103mA) arising from 103mA (10mA + 1500V/1.8kΩ×(1- 0.1)) of leakage current.

ii. If the device with the lowest leakage is associated with the worse case resistance (upper tolerance band limit), and simultaneously the supply is at its upper tolerance limit, then worse case resistance is given by equation (10.11), that is

$$\hat{R} \le \frac{n(1-a)V_D - (1+a)(1+b)V_s}{(n-1)(1-a^2)\hat{I}_b}$$
$$= \frac{10\times(1-0.1)\times200\text{V} - (1+0.1)\times(1+0.05)\times1500\text{V}}{(10-1)\times(1-0.1^2)\times10\text{mA}} = 758\Omega$$

Each resistor (preferred value 680Ω) needs to be rated in excess of

$$200\text{V}^2/680\Omega\times(1-0.1) = 68.6\text{ W}$$

♣

When resistance tolerances are considered, sharing resistors of lower value must be used and the wider the tolerance, the lower will be the resistance and the higher the power losses. A number of solutions exist for reducing power losses and economic considerations dictate the acceptable trade-off level. Matched semiconductor devices would allow a minimum number of string devices (voltage sharing factor $k_s \to 1$) or, for a given string device number, a maximum value of sharing resistance (lowest losses). But matching is complicated by the fact that semiconductor leakage current varies significantly with temperature. Alternatively, by increasing the string device number (decreasing the sharing factor k_s) the sharing resistance is increased, thereby decreasing losses. By increasing the string device number from 10 (k_s = ¾) to 11 (k_s = 0.68) in example 10.1, the sharing resistance requirement increases from 4.7kΩ to 6.8kΩ and resistor losses are reduced from a total of 50.8 W to 31 W. Another method of minimising sharing resistance losses is to minimise resistance tolerances. A tolerance reduction from 10 per cent to 5 per cent in example 10.1 increases the sharing resistance requirements from 1.8kΩ to 3.9kΩ, while total power losses are reduced from 140 W to 64 W. These worse case losses assume a near 100% off-state duty cycle.

10.1.1ii - Transient voltage sharing

During steady-state or at very low frequencies, sharing resistors as shown in figure 10.2 are sufficient to prevent individual device overvoltage. Mismatching of turn-on delay times of thyristors and transistors can be minimised by supplying high enough turn-on drive with very fast rise times. A higher initial *di/dt* is then allowable.
Before a conducting string of diodes or thyristors can reverse-block, reverse recovery charge must flow. Those elements with least recovery charge requirements recover first and support the reverse bias. The un-recovered devices recover slowly, since recovery now occurs as a result of the low leakage current though the recovered devices, and natural recombination.
The transient reverse-blocking voltage can be shared more equally by placing capacitance across each string element as shown in figure 10.3. The capacitor action is to provide a transient current path bypassing a recovered device to allow a slower device to recover and to support volts. In the case of thyristors, low value resistance is connected in series with each capacitor to avoid high capacitor discharge through the thyristors at turn-on. Figure 10.4 shows the *I-V* characteristics of two unmatched thyristors or diodes during reverse recovery.

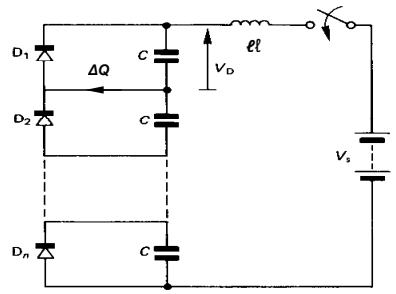


Figure 10.3. *A series diode string with shunting capacitance for transient reverse blocking voltage sharing.*

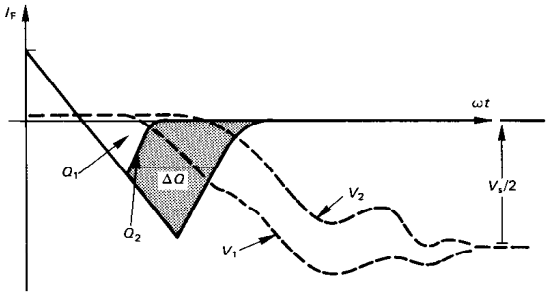


Figure 10.4. *Reverse recovery current and voltage for two mismatched series connected diodes.*

The worst case assumptions for the analysis of figure 10.3 are that element D_1 has minimum stored charged $\check{Q}$ while all other devices have the maximum requirement, $\hat{Q}$. The charge difference is

$$\Delta Q = \hat{Q} - \check{Q} \qquad \text{(C)} \qquad (10.12)$$

The total string dc voltage V_s, comprises the voltage across the fast-recovery device V_D plus the sum of each of the voltages across the slow n - 1 devices, V_{slow}. That is

$$V_s = V_D + (n - 1)V_{slow} \qquad \text{(V)} \qquad (10.13)$$

The voltage across each slow device is given by

$$V_{slow} = \frac{1}{n}\left(V_s - \Delta\hat{V}\right) \qquad \text{(V)} \qquad (10.14)$$

where $\Delta\hat{V} = \Delta\hat{Q}/C$.

Eliminating V_{slow} from equations (10.13) and (10.14) yields

$$\check{C} \geq \frac{(n-1)\Delta Q}{nV_D - V_s} = \frac{(n-1)\Delta Q}{n(1-k_s)V_D} \qquad \text{(F)} \qquad (10.15)$$

This equation shows that as the number of devices is minimized, the sharing factor, k_s, which is in the denominator of equation (10.15), tends to one and the capacitance requirement undesirably increases. Manufacturers do not specify the minimum reverse recovery charge but specify the maximum reverse recovery charge for a given initial forward current, reverse recovery *di/dt,* and temperature. For worst case design, assume $\hat{Q}$ = 0, thus

$$\check{C} \geq \frac{(n-1)\hat{Q}}{nV_D - V_s} = \frac{(n-1)\hat{Q}}{n(1-k_s)V_D} \qquad \text{(F)} \qquad (10.16)$$

Voltage sharing circuit design is complicated if the effects of reverse steady-state leakage current in ac thyristor blocking are taken into account.

Supply and sharing capacitance tolerances significantly affect the minimum capacitance requirement. Worst case assumptions for capacitance tolerances involve the case when the fastest recovering diode is in parallel with capacitance at its lower tolerance limit while all the other sharing capacitances are at their upper tolerance limit. Assuming the minimum reverse recovery charge is zero, then the minimum sharing capacitance requirement is

$$\check{C} \geq \frac{(n-1)\hat{Q}}{(1-a)(nV_D - V_s)} = \frac{(n-1)\hat{Q}}{n(1-a)(1-k_s)V_D} \qquad \text{(F)} \qquad (10.17)$$

where -100*a* is the capacitor negative percentage tolerance and $n \geq 2$. Voltage sharing resistors help minimise capacitor static voltage variation due to capacitance variations.

If the supply tolerance is incorporated, then V_s in equations (10.16) and (10.17) are replaced by $(1+b)\times V_s$ where +100*b* is the supply percentage upper tolerance. This leads to an increased capacitance requirement, hence increased energy losses, $\frac{1}{2}CV_D^2$.

$$\check{C} \geq \frac{(n-1)\hat{Q}}{(1-a)(nV_D - (1+b)V_s)} \qquad \text{(F)} \qquad (10.18)$$

Example 10.2: ***Series device connection – dynamic voltage balancing***

The string of ten, 200 V diodes in worked example 10.1 is to incorporate capacitive reverse recovery transient sharing. Using the data in chapter 5, figure 5.9, specify a suitable sharing capacitance based on zero capacitance and supply tolerances (*a* = *b* = 0), ± 10 per cent capacitance tolerances (*a* = 0.1, *b* = 0), ± 5 per cent supply tolerance (*a* = 0, *b* = 0.05), then both tolerances (*a* = 0.1, *b* = 0.05). Estimate in each case the capacitor energy loss at capacitor discharge.

Solution

Figure 5.9 shows that worst case reverse recovery conditions occur at maximum junction temperature, *di/dt,* and I_F, and a value of $\hat{Q}$ = 6μC is appropriate.
The minimum possible sharing capacitance occurs when the capacitance and dc rail voltage are tightly specified. From equation (10.16)

$$\check{C} \geq \frac{(n-1)\hat{Q}}{nV_D - V_s} = \frac{(10\text{-}1)\times 6\mu\text{C}}{10\times 200\text{V} - 1500\text{V}} = 108\text{nF @ 200Vdc}$$

The sharing capacitance requirement with 10% tolerance capacitors, is given by equation (10.17)

$$\check{C} \geq \frac{(n-1)\hat{Q}}{(1-a)(nV_D - V_s)} = \frac{(10 - 1)\times 6\mu\text{C}}{(1 - 0.1)\times(10\times 200\text{V-}1500\text{V})} = 0.12\mu\text{F @ 200Vdc}$$

A further increase in capacitance requirements results if the upper tolerance dc rail voltage is used. From equation (10.18)

$$\check{C} \geq \frac{(n-1)\hat{Q}}{(1-a)(nV_D - (1+b)V_s)}$$

$$= \frac{(10 - 1)\times 6\mu\text{C}}{(1 - 0.1)\times(10\times 200\text{V} - (1+0.05)\times 1500\text{V})} = 0.14\mu\text{F @ 200Vdc}$$

In each tolerance case the next larger preferred capacitance value should be used, namely, 120nF, 120nF, and 150nF respectively, all rated at 200V dc.
The total series capacitance, using the upper tolerance limit is

$$C_T = \frac{(1+a)\check{C}}{n}$$

The stored energy with a 1500V dc rail in the 10 series connect 120nF capacitors, and subsequently loss when the string voltages reduces to zero at diode forward bias, is therefore

$$W_T = \tfrac{1}{2}C_T\hat{V}_s^2 = \tfrac{1}{2}\frac{(1+a)\check{C}}{n}V_s^2(1+b)^2$$

$$= \tfrac{1}{2}\frac{(1 + 0.1)\times 120\text{nF}}{10}\times 1500\text{V}^2\times(1+0.05)^2 = 16.4\text{mJ}$$

The energy stored in the 10 series connect 150nF capacitors, and subsequently loss when the string voltage reduces to zero at diode forward bias, is

$$W_T = \tfrac{1}{2}\times\frac{(1 + 0.1)\times 150\text{nF}}{10}\times 1500\text{V}^2\times(1 + 0.05)^2 = 20.5\text{mJ}$$

♣

When capacitive sharing is used with switching devices, at turn-on the transient sharing capacitor discharges into the switching device. The discharge current magnitude is controlled by the turn-on voltage fall characteristics. If a linear voltage fall at turn-on is assumed, then the transient sharing capacitor maximum discharge current i_{dis} is a constant current pulse for the fall duration, of magnitude

$$i_{dis} = C\frac{\Delta V_D}{\Delta t} = C\frac{V_D}{t_{fv}} \quad \text{(A)} \qquad (10.19)$$

The discharge current can be of the order of hundreds of amperes, incurring initial *di/dt* values beyond the capabilities of the switching device. In example 10.2 the discharge current for a switch rather than a diode is approximately 150nF×200V/1µs =30A, assuming a 1µs voltage fall time. This 30A may not be insignificant compared to the switches current rating. But, advantageously, the sharing capacitors do act as turn-off snubbers, reducing switch turn-off stressing.
In the case of the thyristor, the addition of low-valued, low inductance, resistance in series with the transient capacitor can control the capacitor discharge current, yet not significantly affect the transient sharing properties. The resultant *R-C* discharge current can provide thyristor latching current while still offering transient recovery sharing, *dv/dt,* and voltage spike suppression. Thyristor snubber operation and design are considered in chapter 8.1.2.
Figure 10.5 shows the complete steady-state and transient-sharing networks used for diodes, thyristors, and transistors. Transient voltage sharing for transistors involves the use of the conventional *R-D-C* snubber shown in figure 10.5c and considered in chapter 8. The series inductor used with thyristor and transistor strings provides transient turn-on voltage protection. The inductor supports the main voltage while each individual element switches on. Such an inductive turn-on snubber is mandatory for the GCT and the GTO thyristor. No one device is voltage-stressed as a consequence of having a longer turn-on delay time, although gate overdrive at turn-on minimises delay variations.

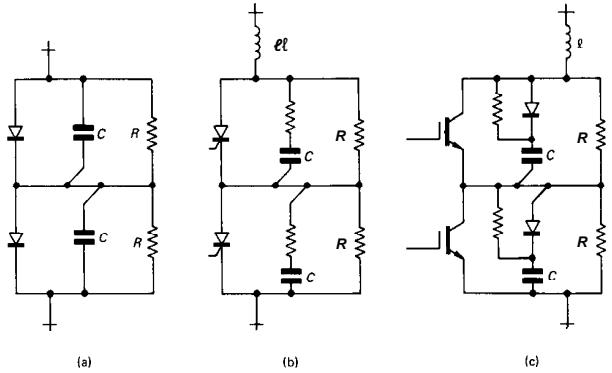


Figure 10.5. *Transient and steady-state voltage sharing circuits for series connected: (a) diodes; (b) thyristors; and (c) igbt transistors.*

10.1.2 Parallel semiconductor device operation

It is common practice to parallel power devices in order to achieve higher current ratings or lower conducting voltages than are attainable with a single device. Although devices in parallel complicate layout and interconnections, better cooling distribution is obtained. Also, built-in redundancy can give improved equipment reliability. A cost saving may arise with extensive parallel connection of smaller, cheaper, high production volume devices.
The main design consideration for parallel device operation is that all devices share both the steady-state and transient currents. Any bipolar device carrying a disproportionately high current will heat up and conduct more current, eventually leading to thermal runaway as considered in section 4.1.
The problem of current sharing is less severe with diodes because diode characteristics are more uniform (because of their simpler structure and manufacturing) than those of thyristors and transistors. Two basic sharing solutions exist

- matched devices
- external forced current sharing.

10.1.2i - Matched devices

Figure 10.6 shows the static *I-V* on-state characteristics of two SCR's. If these two devices are connected in parallel, for the same on-state voltage, the resultant current flow is $I_1 + I_2$ where I_1 and I_2 can be very different in value. The total current rating of the pair is not the sum of the maximum current rating for each but rather a value which can be just larger than the rating of one device alone. The percentage parallel derating *pd* for *n* parallel connected devices is defined as

$$pd = \left(1 - \frac{I_T}{nI_m}\right)\times 100 \;\; = \left(1 - k_p\right)\times 100 \quad \text{per cent} \qquad (10.20)$$

where I_T = total current through the parallel arrangement
I_m = maximum allowable single device current rating
n = number of parallel devices
k_p = current parallel sharing factor = $I_T/nI_m \le 1$

Parallel connection of IGBT die within a module is made possible by using die from the same wafer/batch. On-state voltage matching for single large area wafers is expensive and complicated by the high temperature dependence of both static and dynamic electrical device characteristics.
Derating does not account for effects such as layout and electrical and thermal impedance imbalance. The amount of derating is traded off against the extra cost involved in selecting devices with closer (matched) static characteristics.

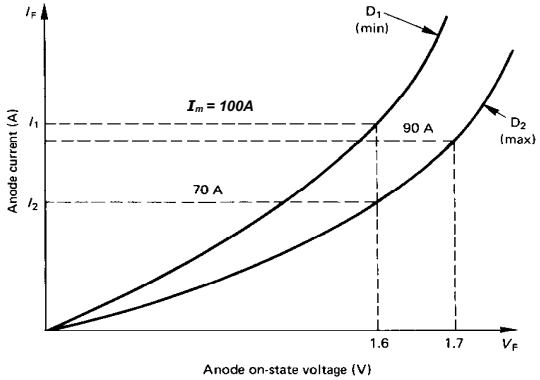


Figure 10.6. *Forward conduction characteristics of two unmatched devices.*

10.1.2ii - External forced current sharing

Forced current sharing is applicable to both steady-state and transient conditions. For a current derating of less than 5 per cent it is usually cheaper to use forced sharing techniques rather than matched devices.
Figure 10.6 shows the maximum variation of *I-V* characteristics in devices of the same type. When parallel connected the maximum current is restricted to I_m+I_2, (= 100A+70A = 170A at 1.6V). The maximum current rating for each device is I_m, (100A); hence with suitable forced sharing a combination in excess of $I_m + I_2$ (170A) should be possible. The resistive network in figure 10.7 is used for forced current sharing and in example 10.3 it is required that I_m, 100A, flows through D_1 and (1-2×*pd*)×$I_m > I_2$, (90A) flows through D_2, for a *pd* (5%) overall derating.
From Kirchhoff's voltage law in figure 10.7

$$\begin{aligned} V_1 + V_3 &= V_2 + V_4 \\ V_{D_1} + I_m R &= V_{D_2} + \left(I_T - I_m\right)R \end{aligned} \qquad (10.21)$$

From equation (10.20), rearranged for two devices, *n* = 2

$$I_T = 2\times(1 - pd)I_m = 2k_p I_m$$

Substituting for I_T in equation (10.21) gives

$$R = \frac{V_{D_2} - V_{D_1}}{2\ pd\ I_m} \quad \text{(ohms)} \qquad (10.22)$$

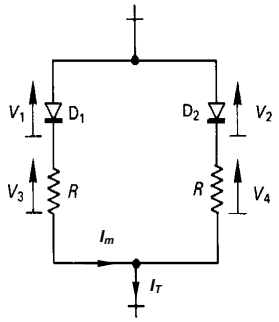


Figure 10.7. *Forced current sharing network for parallel connected devices.*

For *n* devices connected in parallel, equation (10.21) becomes

$$V_{D_1} + I_m R = V_{D_2} + \frac{(I_T - I_m)}{n-1} R \qquad (10.23)$$

which after substituting for I_T from equation (10.20), for maximum device voltage variation, gives

$$R = \frac{\hat{V}_D - \check{V}_D}{I_m} \frac{(n-1)}{n \times pd} \qquad \text{(ohms)} \qquad (10.24)$$

Although steady-state sharing is effective, sharing resistor losses can be high. The total resistor losses in general terms for *n* parallel connected devices and a conduction duty cycle *δ*, are given by

$$P_t = \delta \left\{ 1 + \left(1 - \frac{n}{n-1} \times pd \right)^2 \right\} I_m^2 R \qquad \text{(W)} \qquad (10.25)$$

Since the devices are random in characteristics, each resistor must have a power rating of $I_m^2 R$.

Example 10.3: ***Resistive parallel current sharing – static current balancing***

For the two diodes shown in figure 10.6, with $\hat{I} = 100\text{A}$, what derating results when they are parallel connected, without any external sharing circuits?
The maximum current rating for each device is I_m, 100A; hence with suitable forced sharing a 190A combination should be possible. Using the network in figure 10.7 for current sharing, it is required that 100A flows through D_1 and 90A through D_2. Specify the per cent overall derating, the necessary sharing resistors, their worse case losses and diode average, rms, and ac currents at a 50% duty cycle and worse case.

Solution

The derating for the parallel situation depicted in figure 10.6, without external sharing, is

$$pd = \left(1 - \frac{170\text{A}}{2 \times 100\text{A}}\right) \times 100 = 15 \text{ per cent} \qquad \left(k_p = \frac{100\text{A}+70\text{A}}{2 \times 100\text{A}} = 0.85\right)$$

With forced resistive sharing, the objective derating is reduced from 15% to

$$pd = \left(1 - \frac{190\text{A}}{2 \times 100\text{A}}\right) \times 100 = 5 \text{ per cent} \qquad \left(k_p = \frac{100\text{A}+90\text{A}}{2 \times 100\text{A}} = 0.95\right)$$

From figure 10.6

$$1.6\text{V} + 100\text{A} \times \text{R} = 1.7\text{V} + 90\text{A} \times \text{R}$$

that is

$$R = 10 \text{ milliohm}$$

Equation (10.22), being based on the same procedure, gives the same result. The cell voltage drop is increased to $1.6\text{V}+100\text{A}\times0.01\Omega = 1.7\text{V}+90\text{A}\times0.01\Omega = 2.6\text{V}$.
Thus, for an on-state duty cycle *δ*, the total losses are *δ*×2.6V×190A = *δ*×494W.

For $\delta = ½$

$$\bar{I}_{D1} = \delta \times I_{D1} = ½ \times 100\text{A} = 50\text{A} \qquad \bar{I}_{D2} = \delta \times I_{D2} = ½ \times 90\text{A} = 45\text{A}$$
$$I_{D1rms} = \sqrt{\delta} \times I_{D1} = \sqrt{½} \times 100\text{A} = 70.7\text{A} \qquad \bar{I}_{D2rms} = \sqrt{\delta} \times I_{D2} = \sqrt{½} \times 90\text{A} = 63.6\text{A}$$
$$I_{D1ac} = \sqrt{I_{D1rms}^2 - \bar{I}_{D1}^2} = \sqrt{70.7^2 - 50^2} = 50\text{A} \qquad I_{D2ac} = \sqrt{I_{D2rms}^2 - \bar{I}_{D2}^2} = \sqrt{63.6^2 - 45^2} = 45\text{A}$$
$$P_{R1} = I_{D1rms}^2 R_1 = 70.7^2 \times 0.01\text{m}\Omega = 50\text{W} \qquad P_{R2} = I_{D2rms}^2 R_2 = 63.6^2 \times 0.01\text{m}\Omega = 40.5\text{W}$$
$$P_{D1} = \bar{I}_{D1} V_{D1} = 50\text{A} \times 1.6\text{V} = 80\text{W} \qquad P_{R2} = \bar{I}_{D2} V_{D2} = 45\text{A} \times 1.7\text{V} = 76.5\text{W}$$
$$P_{total} = P_R + P_D = (50\text{W} + 40.5\text{W}) + (80\text{W} + 76.5\text{W}) = 90.5\text{W} + 156.5\text{W} = 247\text{W}$$

For worse case losses, $\delta \to 1$

$$\bar{I}_{D1} = \delta \times I_{D1} = 1 \times 100\text{A} = 100\text{A} \qquad \bar{I}_{D2} = \delta \times I_{D2} = 1 \times 90\text{A} = 90\text{A}$$
$$I_{D1rms} = \sqrt{\delta} \times I_{D1} = \sqrt{1} \times 100\text{A} = 100\text{A} \qquad \bar{I}_{D2rms} = \sqrt{\delta} \times I_{D2} = \sqrt{1} \times 90\text{A} = 90\text{A}$$
$$I_{D1ac} = \sqrt{I_{D1rms}^2 - \bar{I}_{D1}^2} = \sqrt{100^2 - 100^2} = 0\text{A} \qquad I_{D2ac} = \sqrt{I_{D2rms}^2 - \bar{I}_{D2}^2} = \sqrt{90^2 - 90^2} = 0\text{A}$$
$$P_{R1} = I_{D1rms}^2 R_1 = 100^2 \times 0.01\text{m}\Omega = 100\text{W} \qquad P_{R2} = I_{D2rms}^2 R_2 = 90^2 \times 0.01\text{m}\Omega = 81\text{W}$$
$$P_{D1} = \bar{I}_{D1} V_{D1} = 100\text{A} \times 1.6\text{V} = 160\text{W} \qquad P_{R2} = \bar{I}_{D2} V_{D2} = 90\text{A} \times 1.7\text{V} = 153\text{W}$$
$$P_{total} = P_R + P_D = (100\text{W} + 81\text{W}) + (160\text{W} + 153\text{W}) = 181\text{W} + 313\text{W} = 494\text{W}$$

The general form in equation (10.25) gives the same total resistor losses for each conduction duty cycle case, namely for $\delta = ½$: 50W+40.5W = 90.5W and for $\delta \to 1$: 100W+81W = 181W.

♣

A more efficient method of current sharing is to use coupled reactors as shown in figure 10.8. In these feedback arrangements, in figure 10.8a, if the current in D_1 tends to increase above that through D_2, the voltage across L_1 increases to oppose current flow through D_1. Simultaneously a negative voltage is induced across L_2 thereby increasing the voltage across D_2 thus increasing its current. This technique is most effective in ac circuits where the core is more readily designed to reset, not saturate.

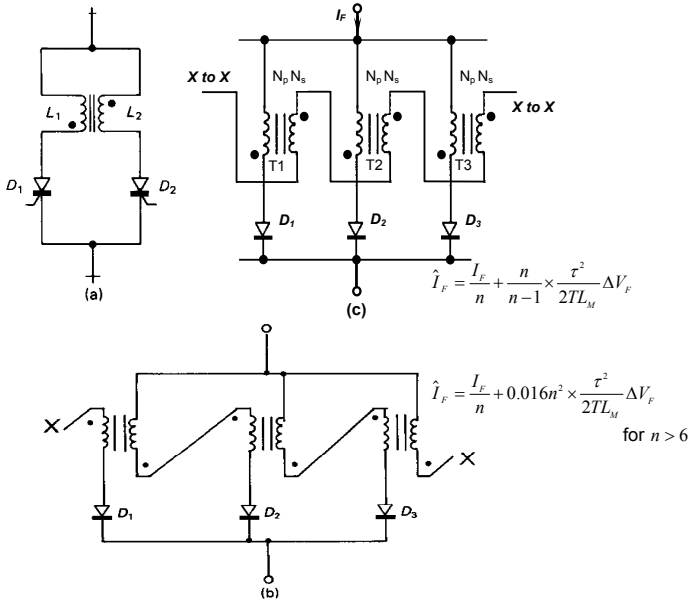


Figure 10.8. *External forced current sharing networks using cross-coupled reactors: (a) for two devices; and (b) and (c) for many devices.*

Equalising reactor arrangements are possible for any number of devices in parallel, as shown in figures 10.8b and c, but size and cost become limiting constraints. The technique is applicable to steady-state and transient sharing. At high current densities, the forward *I-V* characteristic of diodes and thyristors (and some IGBTs) has a positive temperature dependence which provides feedback aiding sharing.
The mean current in the device with the highest current, therefore lowest voltage, of *n* parallel connected devices in figure 10.8c (with one coupled circuit in series with each device), is given by

$$\bar{I}_F = \frac{I_F}{n} + \Delta I_F = \frac{I_F}{n} + \frac{n-1}{n} \times \frac{\tau^2}{2TL_M} \Delta V_F = \frac{I_F}{n} + \frac{n-1}{n} \times \frac{\delta^2}{2 f_s L_M} \Delta V_F \qquad (10.26)$$

where ΔV_F is the maximum on-state voltage drop difference
L_M is the self-inductance (magnetising inductance) of the coupled inductor
T is the cycle period, $1/f_s$ and
τ is the conduction period ($\tau < T$)

(a) current sharing analysis for two devices:– $r_o = 0$
Consider two thyristors (n = 2) connected in parallel as show in figure 10.9. The coupled circuit magnetising current is modelled with the magnetising inductor L_M. The transformer turns ratio is 1:1, hence the winding voltages and currents are equal, taking into account the relative winding flux orientation shown by the dots. Commutation inductance overlap is ignored.
From Kirchhoff's voltage law

$$v_{T1} + v_1 = v_{T2} - v_1 \qquad (10.27)$$

That is

$$v_1 = \tfrac{1}{2} \times \left(v_{T1} - v_{T2} \right) = \tfrac{1}{2} \times \Delta v \qquad (10.28)$$

From Kirchhoff's current law

$$I_M = i_1 - i_2 \qquad (10.29)$$

From Faraday's equation

$$v_1 = L_M \frac{dI_M}{dt} \qquad (10.30)$$

which after integrating both sides gives

$$I_M = \frac{1}{L_M} \int_0^{\tau} v_1 dt = \tfrac{1}{2} \frac{1}{L_M} \Delta V_F \tau \qquad (10.31)$$

As a condition it is assumed that the voltage difference Δv does not decrease as the operating point moves along the *I-V* characteristics. That is, both devices are modelled by $v = v_o + i \times r_o$, where the linear resistance r_o, is zero, each have different zero current voltages that is different v_o, $\Delta v_o = \Delta V_F$. Actually D_1 moves further up the *I-V* characteristic with time as it conducts more current while D_2 moves towards the origin, as shown in figure 10.9b.

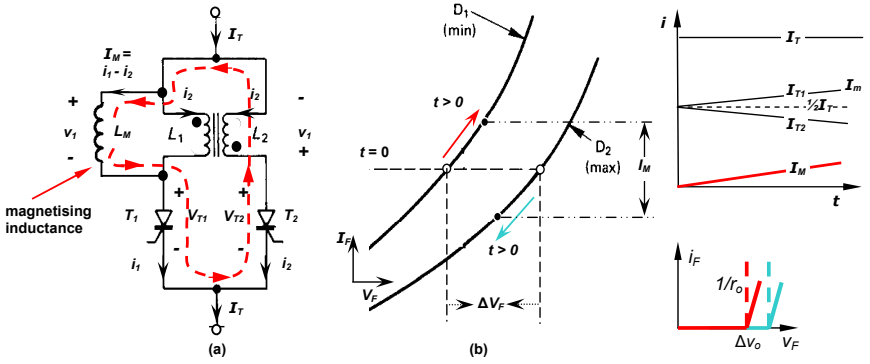


Figure 10.9. *External forced current sharing network using cross-coupled reactors: (a) circuit (including magnetising inductance L_M) for two devices and (b) I-V operating points.*

(b) current sharing analysis for two devices:– $r_o \neq 0$
If static resistance is included into the device model for current sharing analysis, then equation (10.30), assuming both devices have the equal resistance, becomes

$$\Delta v_o = L_M \frac{dI_M}{dt} + 2 I_M r_o \qquad (10.32)$$

The solution to this differential equation gives the magnetizing current as

$$I_M = \frac{\Delta v_o}{2r_o} \left[1 - e^{-t \frac{2r_o}{L_M}} \right] \qquad (10.33)$$

The maximum magnetizing current increases from zero and reaches a maximum at the end of the current conduction period τ. Re-arranging equation (10.33) gives the magnetizing inductance as

$$L_M = \frac{2 r_o \tau}{\ell \mathrm{n} \left(\frac{\Delta v_o}{\Delta v_o - I_M 2 r_o} \right)} \qquad (10.34)$$

(c) current sharing analysis for *n* devices:– $r_o = 0$
When more than two devices are parallel connected, sharing can be enforced with the multiple transformer technique shown in figure 10.8c, where the ***n*** transformer secondary windings are series connected. Each transformer has a turns ratio of $\eta = N_p{:}N_s$, and the magnetising inductance is assumed to be on the primary side of each transformer.
The semiconductor devices are assumed to have a constant on-state voltage v_o. The total current is I_T, and zero commutation inductance is assumed.

Using Kirchhoff's voltage law on the primary side:
Since the secondary voltages sum to zero

$$v_{s1} + v_{s2} + v_{s3} + \ldots + v_{sn} = 0 \qquad (10.35)$$

then the transformer primary voltages also sum to zero

$$v_{p1} + v_{p2} + v_{p3} + \ldots + v_{pn} = \frac{N_s}{N_p} \left(v_{s1} + v_{s2} + v_{s3} + \ldots + v_{sn} \right) = 0 \qquad (10.36)$$

Since the legs are parallel connected

$$V_{T1} + v_{p1} = V_{T2} + v_{p2} = \ldots\ldots = V_{Tn} + v_{pn} \qquad (10.37)$$

For worst case analysis, let one device (n = 1) operate at minimum on-state voltage, $\check{V}_T$, while the other n - 1 devices have a maximum on-state voltage $\hat{V}_T$, therefore potentially conduct less current than the device operating at minimum voltage.

$$\check{V}_T + v_{p1} = \hat{V}_T + v_p = \ldots\ldots = \hat{V}_T + v_p \qquad (10.38)$$

These equations yield the following primary voltages

$$v_{p1} = \frac{n-1}{n} \left(\hat{V}_T - \check{V}_T \right) \quad \text{and} \quad v_{p2} = v_{p3} \ldots = v_{pn} = -\frac{1}{n} \left(\hat{V}_T - \check{V}_T \right) \qquad (10.39)$$

Using Kirchhoff's current law on the primary side:

$$I_T = I_{T1} + I_{T2} + \ldots + I_{TN} \qquad (10.40)$$

But a thyristor current, which is the transformer primary current, can be expressed in terms of the transformer secondary current plus the parallel magnetising current on the primary side. That is

$$I_{Ti} = i_{pi} + i_{Mi} = \frac{N_s}{N_p} i_s + i_{Mi} = \frac{1}{\eta} i_s + i_{Mi} \qquad (10.41)$$

where, because the secondary windings are series connected, the secondary current is the same for each transformer. The transformer magnetising current i_{Mi} is the same for transformers i = 2 to n, i_M. Thus the total current

$$I_T = \sum_{i=1}^{n} I_{Ti} = \sum_{i=1}^{n} \left(\frac{1}{\eta} i_s + i_{mi} \right)$$
$$I_T = n \frac{1}{\eta} i_s + i_{m1} + (n-1) i_m \qquad (10.42)$$

Using Kirchhoff's voltage law on the secondary side:
Since the transformers are identical, each has the same value of magnetising inductance (self-inductance) L_M. Because the secondary windings are series connected the sum of the secondary voltages, hence sum of primary voltages, are zero.

$$\begin{aligned} v_{p1} + v_{p2} + v_{p3} + \ldots + v_{pn} &= 0 \\ = L_M \frac{d\,i_{M1}}{dt} + L_M \frac{d\,i_{M2}}{dt} + \ldots\ldots + L_M \frac{d\,i_{Mn}}{dt} &= 0 \\ = L_M \left(\frac{d\,i_{M1}}{dt} + (n-1)\frac{d\,i_M}{dt} \right) &= 0 \\ = L_M \frac{d}{dt}\left[i_{M1} + (n-1)\,i_M \right] &= 0 \end{aligned} \qquad (10.43)$$

This component inside the square bracket must be a constant.

$$i_{M1} + (n-1)\,i_M = \mathrm{c} \qquad (10.44)$$

Substituting the constant c into equation (10.42) gives the secondary current as

$$i_s = \eta \frac{1}{n}(I_T - \mathrm{c}) \qquad (10.45)$$

In conjunction with Faraday's Equation, the magnetising current is a linear function of time, starting from zero. Applying these conditions to the worst case device, T1, then as the magnetising current in transformer Tr1 increases and the associated thyristor current I_{T1} increases, from equation (10.44), the opposing magnetising current in the other transformers reduces the associated device principal current. At the maximum on-time, the current in device T1 should not exceed its permitted rated limit, I_m.

$$\Delta i_{M1}(t) = \frac{1}{L_M} \times v_{p1} t \qquad (10.46)$$

From equation (10.39), when $t = \tau$, the maximum magnetising current, in terms of the device voltage extremes, is

$$\Delta i_{M1}(t = \tau) = \Delta \hat{i}_{M1} = \frac{1}{L_M} \times \frac{n-1}{n} \left(\hat{V}_T - \check{V}_T \right) \times \tau \qquad (10.47)$$

Re-arranging gives the necessary minimum transformer self-inductance with respect to the primary side.

$$L_M = \frac{1}{\Delta \hat{i}_{M1}} \times \frac{n-1}{n} \left(\hat{V}_T - \check{V}_T \right) \times \tau = \frac{1}{\Delta \hat{i}_{M1}} \times \frac{n-1}{n} \times \Delta V_F \times \tau \qquad (10.48)$$

The maximum magnetising current $\Delta \hat{i}_{M1}$ can be expressed in terms of devices current rating I_m and device percentage derating, *pd*, or device utilisation, $k_p = 1 - pd$.
If the device current rating is I_m, then n devices in parallel can theoretically conduct $n \times I_m$. When derated by *pd* to k_p, the total current is $k_p \times n I_m$ where each device initially conducts $k_p \times I_m$. The current in the worst case device increases from $k_p I_m$ to I_m ($\Delta \hat{i}_{M1} = (1 - k_p) I_m = pd \times I_m$) in the maximum period the device conducts, τ.

$$\hat{I}_T = k_p I_m + \Delta \hat{i}_M = k_p I_m + (1 - k_p) I_m = I_m \qquad (10.49)$$

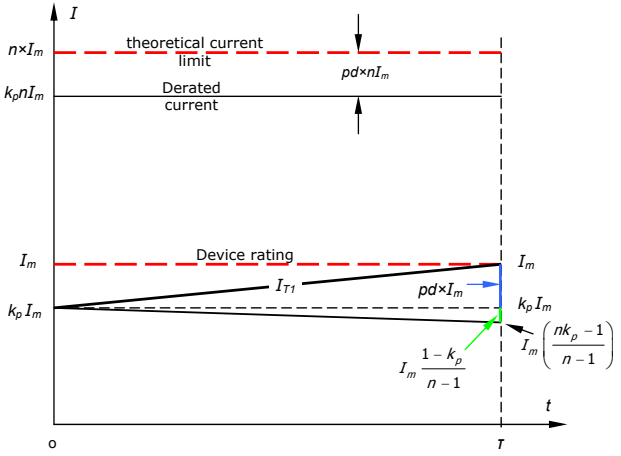


Figure 10.10. *External forced current sharing network using series connected secondary windings.*

The current in each of the remaining n-1 devices decreases from $k_p I_m$ by (1-k_p) I_m/n-1 to

$$\check{I}_T = k_p I_m - I_m \frac{1 - k_p}{n-1} = I_m \left(\frac{n k_p - 1}{n-1} \right) \qquad (10.50)$$

such that the necessary total current is maintained:

$$\hat{I}_T + (n-1)\check{I}_T = I_m + (n-1) I_m \left(\frac{n k_p - 1}{n-1} \right) = n k_p I_m$$

These various current components are shown in figure 10.10.
By assuming a current quadratic dependence on time, equations similar to equations (10.26) can be obtained.

Example 10.4: ***Transformer current sharing – static and dynamic current balancing***

Two thyristors with the same forward conduction characteristics as the diodes in figure 10.6 are parallel connected using the coupled circuit arrangement in figure 10.8a.
The maximum current rating for each device is I_m, 100A; hence with suitable forced sharing a 190A combination should be possible. Using the network in figure 10.9a for current sharing, it is required that no more than rated current flow through the lower conducting voltage device, D_1. Specify the per cent overall derating and the necessary sharing transformer properties assuming a half-wave, 180º conduction, phase-controlled, 50Hz, highly inductive load application.
What are the transformer core reset requirements?
Estimate inductance requirements if the thyristors have a static on-state resistance of 1mΩ.

Solution

As in example 10.3, the derating for the parallel situation depicted in figure 10.6, without external sharing, is

$$pd = \left(1 - \frac{170\mathrm{A}}{2 \times 100\mathrm{A}} \right) \times 100 = 15 \text{ per cent} \quad (k_p = 0.85)$$

With forced transformer sharing, the objective derating is reduced from 15% to

$$pd = \left(1 - \frac{190\mathrm{A}}{2 \times 100\mathrm{A}} \right) \times 100 = 5 \text{ per cent} \quad (k_p = 0.95)$$

When the two thyristors are turned on, the magnetizing current is assumed zero and transformer action will force each device to conduct 95A, giving 190A in total. From figure 10.6, the voltage difference between the thyristors, ΔV_F is about 0.1V, thus the transformer winding voltages will be 0.05V each, with polarities as shown in figure 10.9a. In time the magnetizing current increases and the current in T1 increases above 95A due to the increasing magnetizing current, while the current in T2 decreases below 95A, such that the total load current is maintained at 190A.

The worse case conduction period in this ac application, giving maximum magnetising current, is for 180º conduction, that is, 10ms. Thus it is required that T1 current rises to 100A and T2 current falls to 90A after τ =10ms, that is, the magnetising current is 100A - 90A = 10A.
Substitution into equation (10.31) gives

$$L_M = ½ \frac{1}{I_M} \int_0^{10\mathrm{ms}} \Delta v dt = ½ \times \frac{1}{10\mathrm{A}} \times 0.1\mathrm{V} \times 10\mathrm{ms} = 50\mu\mathrm{H}$$

where it is assuming that the voltage differential ΔV_F between the two devices is constant during the conduction period. In fact figure 10.9b shows that the voltage difference decreases, so assuming a constant value gives an under-estimate of requirements.

The core volt-µs during conduction is 0.05V×10ms = 500 V-µs. That is, during core reset the reverse voltage time integral must be at least 500 V-µs to ensure the core flux is reset, (magnetising current reduced to zero).
Using equation (10.34), with r_o = 1mΩ, gives

$$L_M = \frac{2 r_o \tau}{\ell \mathrm{n}\left(\frac{\Delta V_o}{\Delta V_o - I_M 2 r_o} \right)} = \frac{2 \times 1\mathrm{m\Omega} \times 10\mathrm{ms}}{\ell \mathrm{n}\left(\frac{0.1\mathrm{V}}{0.1\mathrm{V} - 10\mathrm{A} \times 2 \times 1\mathrm{m\Omega}} \right)} = 90\mu\mathrm{H}$$

The inductance, 50µH, given by equation (10.31) when neglecting model resistance, under-estimates requirements.

♣

10.2 Protection overview - over-voltage and over-current

All electrical systems are vulnerable to inference and damage from lightning or other short duration electrical surges or long duration supply system swells. As systems become more electronically complex, they also become more vulnerable to external and internally generated interference. A fault can be caused by a device failure or noise which causes undesired device turn-on. This will cause semiconductor device and equipment failure unless protective measures are utilised.
Protection against fault current effects usually involves fuses which clear in time to protect endangered devices, or voltage transient absorption devices which absorb spike energy and clamp the equipment voltage to a safe level. The crowbar fault protection technique can be employed to divert the fault from sensitive components to the crowbar which is a robust circuit. The crowbar clamps the sensitive circuit to zero volts and initiates an isolation breaker or fuse action.
An electrical surge is a temporary increase in voltage, current or both. The size, waveform, and form of the transient surge which can occur within a system are many and varied.

i. Lightning - Although direct strike lightning current can potentially generate transients in the millions of volts and tens of thousands of amps, electronic equipment is rarely exposed to surges of this magnitude. The greatest exposure in power electronics systems is through interconnection and transmission lines. Domestic ac lines can only carry voltages up to 5kV and currents of the order of 1kA. Therefore, for the vast majority of instances where the chance of a lightning strike directly to the equipment is low, 5kV and 1kA is the limit of the direct strike or inductively generated surges. Exposed equipment, such as wind turbines, although suitable earthed, can experience significantly higher electrical surge stresses.

ii. Power Induction - Although power induction voltages can be high in voltage and current, they are often limited in duration. These voltages are caused by faults on the power system which couple into the system (usually inductively as a consequence of the surge causing a large fault current). In virtually all modern power transmission systems, these faults are quickly terminated by circuit breaker and re-closer equipment. This can occur in as short as a couple of cycles of power frequency voltage and rarely takes longer than a second. These transients are typically modelled as a 600Vrms waveform lasting up to a second.

iii. Power Cross - Alternatively, power cross voltages are low voltage events but the exposure can occur for long durations. They are often caused by maintenance error or cabling faults and can result in moderate currents flowing for a long period of time, for example, in domestic applications, 25A for 15 minutes. They are predominately at mains power supply voltage levels (100 to 220Vrms).

iv. Earth Potential Rise (EPR) - EPR can be categorized in two forms:

1. as a result of power system faults and
2. lightning discharges.

In normal industry, where fault currents from the power system are limited in magnitude by fuses and circuit breakers, power system EPR is not usually a considerable risk. EPR only becomes a significant risk when power earthing systems are significantly below standard or where high power transmission systems are used such as at power generation and distribution facilities, within the high power industry, and in the vicinity of electrical traction systems (electric rail). Lightning EPR can only result from a direct strike to the building housing the equipment or in its immediate vicinity. Such events are uncommon, unless the installation is particularly vulnerable due to location or extreme height (for example, wind turbine and cellular phone base-station antennae). The equipment exposure as a result of EPR can be high, and at high earth resistance locations, may become a significant portion of the lightning current.

Surge protection is the process of protecting electronic systems or equipment from voltages and currents which are outside their safe operating limits. These surge voltages and currents can be generated by short circuits, lightning or faults from a power system and usually enter the electronic system along inter-equipment wiring. The surges may be galvanically coupled into the system as in the case of a direct lightning strike, through an inadvertent connection of the power system to the wiring, or as a result of an earth potential rise. They may be capacitively coupled into the system which may occur when a data system is used in the vicinity of a high voltage power line. They may be inductively coupled into the system as may occur if the wiring is run in parallel with large currents running in a power circuit feeding a high power motor. Such events can result in a wide variety of potential consequences.
Electrical surge protection performs several key functions:

- it must prevent or minimize damage caused by a surge;
- it must ensure that the system returns to a working condition with minimal disruption to service.
- under normal conditions the protection must not interfere with any signals or control circuitry, creating challenges for power electronics technologies.
- the protection must operate and fail in a safe manner during overstress.

10.2.1 Ideal secondary level protection

Power electronic equipment is generally within a system that has primary protection, associated with the ac grid protection, for example. The installed equipment therefore may only require secondary protection. Secondary protection prevents the let-through energy of the primary protector (the energy of the surge which gets past the primary protector) from damaging the load.
The peak open circuit voltage of the let-through energy past the primary protector is smaller than the initial external surge. Therefore, a secondary protector can effectively block (series) and or divert (shunt) the reduced surge energy.
The requirements for this ideal blocking device are:

i. As the device needs to block the let-through energy of the primary protector, it can be a series component (in series with the transmission line), located just after the primary protector. As a series component, the device will react to the current through the device rather than, as with a shunt protector, voltage across the interface.

ii. A series device should have a predictable, stable and low trigger current (current at which the device changes between its conductive and non-conductive state) to provide effective protection for sensitive downstream equipment. A shunt device should operate (clamp or fold-back) at a level just above the maximum working voltage.

iii. It should be fast acting (less than 10ns) to protect equipment from surges which rise at 5kV/μs - as with direct lightning strikes or lightning EPR;

iv. As a series device it should have low impedance (resistive, capacitive and inductive) so that it does not effect normal circuit operation, while for the same reasons, a shunt device should have a high standby impedance;

v. In the blocking mode, a series protector should have high impedance so that it does not dissipate significant energy during long duration surges, while for the same reasons, a shunt device should have a low clamping impedance;

vi. It should reset after the surge to reinstate the system and continue to allow normal system operation;

vii. Reset to normal after an incident, returning the equipment to pre-event opration;

viii. Debatably, after excessive stress, a shunt device should fail-safe open circuit, while a series device should failure short-circuit, so as to enable continued unprotected operation, but system protection is afforded.

In addition, for practical and economic reasons it should be small in size and low in cost.

Electrical protection devices fall into two key category, overvoltage (usually shunt connected) and over-current (usually series connected). Over-voltage devices divert or shunt surge current produced by an over-voltage (such as lightning), as shown in Figure 10.11, while most over-current devices increase in resistance (possibly becoming open-circuit) to limit the surge current flowing from longer duration surge currents (50/60 Hz power fault), as shown in the parts of figures 10.12
. There are two types of voltage limiting protectors: switching devices (gas discharge tube GDT and thyristor) that crowbar (voltage fold-back) the line, and voltage clamping devices (metal oxide varistor MOV and transient voltage suppressor TVS). The waveforms of figure 10.11 highlight that switching devices result in lower stress levels than clamping devices (shaded area) for protected equipment during their operation. Functionally, all voltage protectors reset after the surge, while current protectors may or may not reset, depending on their operating mechanisms. For example, PTC thermistors are resettable; fuses are non-resettable.

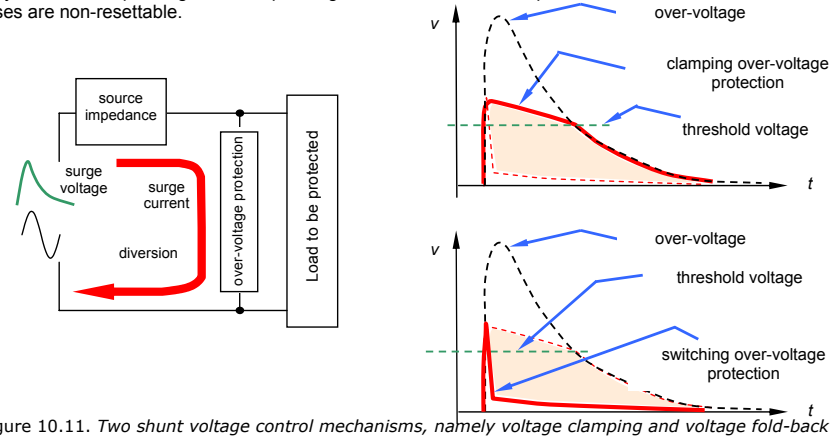


Figure 10.11. *Two shunt voltage control mechanisms, namely voltage clamping and voltage fold-back by switching action, with source and load voltages shown.*

10.2.2 Overvoltage protection devices

Gas Discharge Tubes (GDT) create a quasi short circuit across the line when the internal gas is ionized by an overvoltage, returning to a high impedance state after the surge has ceased. These robust devices have the highest impulse current capability of any technology, and combined with negligible capacitance, make them attractive for the protection of high-speed digital and ac switching converter applications.

Thyristor-based devices initially clamp the line voltage, and then switch to a low voltage on-state. After the surge, when the current drops below the holding current, the protector recovers and returns to its original high impedance blocking state.

Transient Voltage Suppressor (TVS) or Zener diodes operate by rapidly moving from a high impedance to a non-linear resistance characteristic that clamps surge voltages. TVS diodes provide a fast-acting and controlled clamping voltage, however they have high capacitance and low energy capability which restricts the maximum surge current.

Electrostatic discharge (ESD) devices clamping protectors consists of multilayer varistors (MLV) designed to protect equipment against ESD conditions. They have low leakage currents that make the devices transparent under normal operation. ESD transients cause the device to clamp the voltage by reducing its effective resistance and it will reset to a high impedance state after the disturbance. Diode arrays for ESD protection combine thin film on silicon wafer fabrication technology and chip scale packaging. Such devices are used in portable electronics applications where a particular electrical response characteristic is specified for a minimum volume.

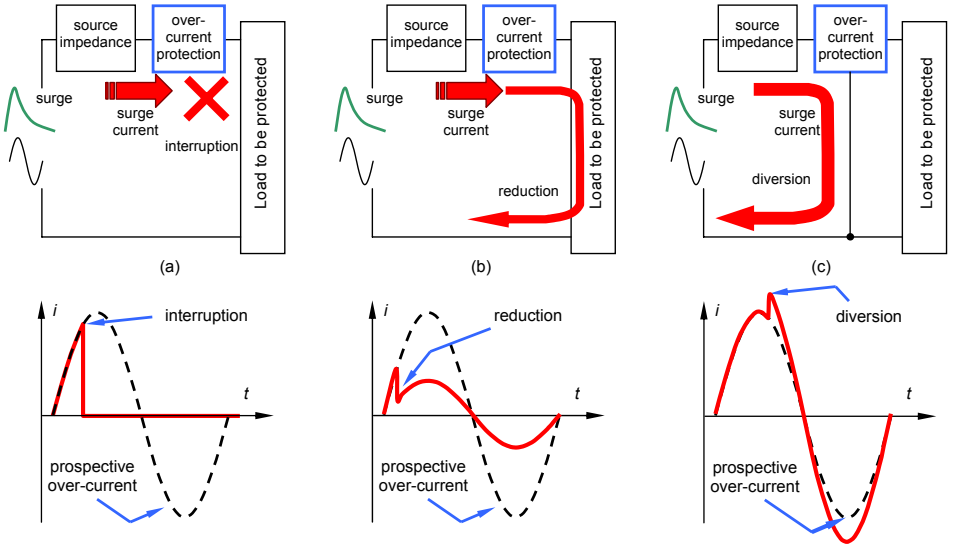


Figure 10.12. *Three current limiting mechanisms:*
(a) current flow interruption, (b) current reduction, and (c) current diversion.

10.2.3 Over-current protection devices

Polymer Positive Temperature Coefficient (PPTC) Thermistor resettable fuses are used in circuit current protection applications. Under high-current fault conditions, its resistance increases by many orders of magnitude and remains in a tripped state, continuing to provide continuous circuit protection until the fault is removed. Then after the power is cycled, the device returns to its normal low-resistance, low-loss state.

Traditional fuses are constructed from a metal element encapsulated in a ceramic housing. The fuse element heats up at the rate related to I^2R. When the metal element temperature exceeds its melting point, it vaporizes and opens the circuit. The low resistance and losses of fuses are attractive for ac applications.

Line Protection Modules (LPM) are based on a basic form of current protection; the Line Feed Resistor (LFR), normally fabricated as a thick-film resistor on a ceramic substrate. LPMs can withstand high-voltage impulses without breaking down. AC current interruption results when the high temperature developed by the resistor produces mechanical expansion stresses that cause the ceramic to break

open. Low current power induction may not break-open the LFR, creating long-term surface temperatures of more than 300°C. To avoid heat damage to adjacent components, the maximum surface temperature can be limited to about 250°C by mounting a series thermal fuse link on the LFR. This capability is exploited in modules incorporating both over-current and over-voltage devices on one ceramic substrate. The incorporation of silicon die and discrete components gives modules with high performance and specific functionality.

A concise overview of generally available over-voltage and over-current protection devices is presented in Table 10.1. The following sections will consider each technology in detail.

Table 10.1: Overview of over-current and over-voltage protection devices and technologies

Device	action	connection	speed	accuracy	current rating
Over-voltage					
GDT	voltage switching	shunt	fair	fair	very high
Thyristor	voltage switching	shunt	fair	good	high
MOV	voltage clamping	shunt	fair	poor	high
TVS	voltage clamping	shunt	fast	good	low
Over-current					
Polymer PTC thermistor	resettable	series	fair	good	low
Ceramic PTC thermistor	resettable	series	slow	good	low
Fuse	non-resettable	series	very slow	Fair	medium/high
Heat coil	non-resettable	shunt or series	very slow	Poor	low
Thermal switch, LFR	non-resettable	series	very slow	poor	high

10.3 Over-current Protection

Current limiting devices provide a slow response, and are primarily aimed at protection from surges lasting hundreds of milliseconds or more, including power induction or contact with AC power. By combining a fixed resistor in series with a resettable protector, an optimum balance of nominal resistance and operating time is obtained. The inherent resistance of certain over-current protectors can also be useful in coordination between primary and secondary overvoltage protection.

Positive Temperature Coefficient (PTC) Thermistors

Heat generated by current flowing in a PTC thermistor causes a step function increase in resistance towards an open circuit, gradually returning close to its original value once the current drops below a threshold value. The resistance stability after surges over time is a key aspect for preserving line balance. PTCs are commonly referred to as resettable fuses, and since low-level current faults are common, automatically resettable protection can be particularly important.

Fuses

A fuse heats up during surges, and once the temperature of the metallic element exceeds its melting point, the normal low resistance creates an open circuit. The low resistance of fuses is attractive for power applications, but their operation is relatively imprecise and time-dependant. Once operated, they do not reset. Fuses also require additional resistance for primary coordination.

Since overvoltage protection usually consists of establishing a low impedance path across the equipment input, overvoltage protection itself will cause high currents to flow. Although relatively slow acting, fuses play a safety role in removing longer-term faults that would damage protection circuitry, thus reducing the size and cost of other protection elements. It is important to consider the I-t performance of the selected fuse, since even multiples of the rated current may not cause a fuse to rupture except after a significant delay. Coordination of this fuse behaviour with the I-t performance of other protection is critical to ensuring that there is no combination of current-level and duration for which the protection is ineffective. By including structures intended to rupture under excess current conditions or separate components, it is also possible to produce hybrid fusible resistors.

Heat Coils
Heat coils are thermally activated mechanical devices connected in series with the line being protected, which divert current to ground. A series coil operates a parallel shunt contact, typically by melting a solder joint that is restraining a spring-loaded contact. When a current generates enough heat to melt the joint, the spring mechanically forces two contacts together, short-circuiting the line. Heat coils are ideal to protect against 'sneak currents' that are too small to activate other methods. Their high inductance makes them unsuitable for digital lines. It is also possible to construct current interrupting heat coils which open the circuit as a result of over-current.

Line Feed Resistors
A Line Feed Resistor (LFR) is the most fundamental form of current protection, normally fabricated as a thick-film device on a ceramic substrate. With the ability to withstand high voltage impulses without breaking down, AC current interruption occurs when the high temperature developed by the resistor causes mechanical expansion stresses that result in the ceramic breaking open.
Low current power induction may not break open the LFR, creating long-term surface temperatures of more than 300°C. To avoid heat damage to the adjacent components, the maximum surface temperature can be limited to about 250°C by incorporating a series thermal fuse link on the LFR. The link consists of a solder alloy that melts when high temperatures occur for periods of 10 seconds or more.
Along with the high precision needed for balanced lines, LFRs have significant flexibility to integrate additional resistors, multiple devices, or even different protection technology within a single component. One possible limitation is the need to dimension the LFR to handle the resistive dissipation under surge conditions. Along with combining multiple non-inductive thick-film resistors on a single substrate to achieve matching to <1%, a resistor can be combined with other devices to optimize their interaction with the overall protection design. For example, a simple resistor is not ideal for protecting a wire, but combining a low value resistor with another over-current protector provides closer protection and less dissipation than either device can offer alone. Both functions can be integrated onto a single thick-film component using fusible elements, PTC thermistors, or thermal fuses. Similarly, more complex hybrids are available, adding surface mount components such as thyristor protectors, to produce coordinated sub-systems.

Thermal Switches
These switches are thermally activated, non-resetting mechanical devices mounted on a voltage-limiting device (normally a GDT). There are three common activation technologies: melting plastic insulator, melting solder pellet or a disconnect device.
Melting occurs as a result of the temperature rise of the voltage-limiting device's thermal overload condition when exposed to a continuous current flow. When the switch operates, it shorts out the voltage-limiting device, typically to ground, conducting the surge current previously flowing through the voltage limiting device.
A plastic-melting based switch consists of a spring with a plastic insulator that separates the spring contact from the metallic conductors of the voltage limiting device. When the plastic melts, the spring contacts both conductors and shorts out the voltage limiting device.
A solder-pellet-melting based switch consists of a spring mechanism that separates the line conductor(s) from the ground conductor by a solder pellet. In the event of a thermal overload condition, the solder pellet melts and allows the spring contacts to short the line and ground terminals of the voltage-limiting device.
A 'snap action' switch typically uses a spring assembly that is held in the open position by a soldered standoff and will short out the voltage-limiting device when its switching temperature is reached. When the soldered connection melts, the switch is released and shorts out the line and ground terminals of the voltage limited.

10.3.1 Protection with fuses
It is not economical to design a circuit where fault overloads are catered for by using devices and components which will withstand worst case faults. A fuse link is normally used for circuit fault current protection.
A fuse link is a current sensitive device designed to serve as the intentional weak link in the electrical circuit. Its function is to provide protection of discrete components, or of complete circuits, by reliably melting under current overload conditions.
A fuse link protecting a semi-conductor is required to carry normal and overload currents but to open the circuit under fault conditions before the semiconductor is damaged. The resultant circuit induced fuse arcing voltage must not cause damage to the circuit. Other fuse links or circuit breakers should be unaffected when the defective cell is disconnected. This non-interaction property is termed *discrimination.*
The fuse element is one or more parallel conductors of pure silver rolled into thin bands, 0.04 to 0.25 mm thick. Each silver band has a number of traverse rows of punched holes (or notches) as shown in

figure 10.13. The area between the holes determines the pre-arcing I^2t integral of the fuse and, along with thermal aspects, is related to the fuse current rating. The number of rows of holes determines the fuse voltage rating. When fusing occurs the current is shared between the holes (the necks), while the arcing voltage is supported between the series of rows of holes. The arcing characteristics are enhanced by packing the silver element in a filler, such as sand or glassed sand. The sand and silver element are contained in a ceramic body and the element is welded or soldered to the fuse contacts (blades or ferrules). These end connector plates are copper flashed and tinned. The element is a calibrated conductor where its configuration, mass, and the materials employed are selected to achieve the desired electrical and thermal characteristics. During normal operation, the I^2R heat generated by the element is absorbed by the sand and transferred through the fuse body to the surrounding air. When an overload current occurs, the element generates heat at a faster rate than the heat can be transferred to the sand. If the overload persists, the element reaches its melting point, melts and then open circuits with an arc. The filler aids fuse performance by absorbing arc energy when the fuse clears an overload or short circuit.
Increasing the applied current will heat the element faster and cause the fuse to open sooner. Thus fuses have an inverse time current characteristic, that is, the greater the over-current the less time required for the fuse to open the circuit. This characteristic is desirable because it parallels the characteristics of conductors, motors, transformers and other electrical apparatus.

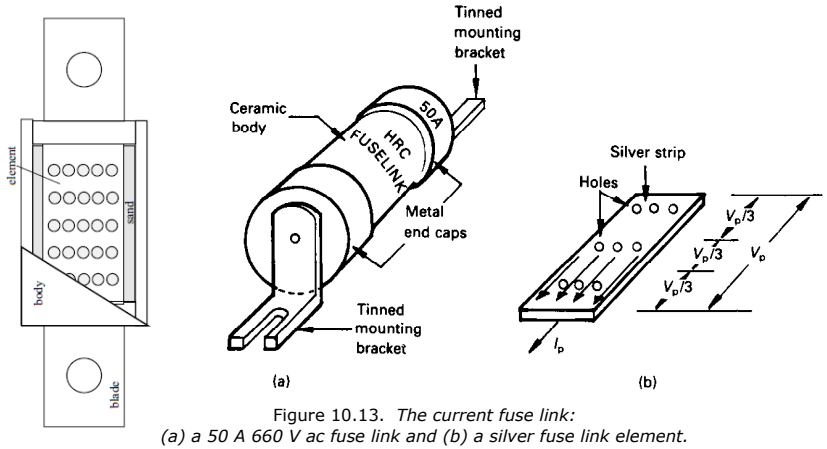


Figure 10.13. *The current fuse link:*
(a) a 50 A 660 V ac fuse link and (b) a silver fuse link element.

The action of a typical fuse link is shown in figure 10.14. Owing to the rms *prospective fault current* I_a the fuse melts at point A, time t_m. Depending on the fuse design and the circuit, the current may continue to rise further to point B, termed the *peak let-through current* I_p. Beyond this point the impedance of the arcing fuse forces the fault current down to zero at the point C. Thus *fuse-clearing* or *total interrupting time* t_c consists of a *melting time* t_m and an *arcing time* t_a.
A series *L-R* circuit can be used to model the prospective fault. The current characteristic is given by

$$i_{sc}(\omega t) = \sqrt{2}\, I_a \left\{ \sin(\omega t - \psi - \phi) - \sin(\psi - \phi)\, e^{-\omega t / \tan\phi} \right\} \qquad (10.51)$$

where ψ is the angle of the short circuit, after the zero voltage cross-over. $\tan\phi = \omega L / R$. The maximum peak fault current therefore occurs when the short appears at zero voltage cross-over, $\psi = 0$.

Differentiation of equation (10.51) gives the current *di/dt*, and the maximum initial *di/dt* is

$$\left. \frac{di}{dt} \right|_{t=o} = \sqrt{2}\, I_a \times \frac{\sin\psi}{\sin\phi} \qquad (10.52)$$

This equation shows that the maximum initial *di/dt* occurs for a short circuit occurring at the peak of the ac supply, $\psi = \frac{1}{2}\pi$, and is independent of the circuit *R-L*, that is independent of ϕ.
The load fault energy, for a fuse link resistance *R*, is

$$W_{tot} = \int_0^{t_c} i_{sc}^2 R \, dt \qquad \text{(J)}$$

If the load current, shown in figure 10.13a, during fuse action is assumed to be triangular, then the clearing integral of the fuse is

$$W_c = \tfrac{1}{3} I_p^2 t_c R \qquad \text{(J)} \qquad (10.53)$$

If the resistance R is assumed constant (because of its low resistivity temperature co-efficient), the value of I^2t ($\frac{1}{3}I_p^2t_c$) is proportional to the energy fed to the protected circuit. The I^2t term is called the *total let-through energy* or the *virtual clearing integral* of the fuse. The energy which melts the fuse is proportional to $\frac{1}{2}I_p^2t_m$ and is termed the *pre-arcing* or *melting* I^2t.

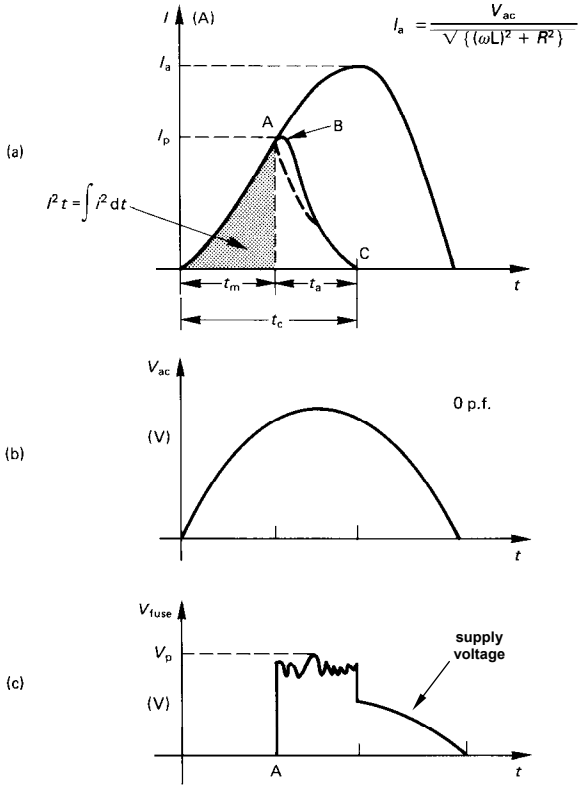


Figure 10.14. *Parameters of a fuse link operating:*
(a) current waveforms; (b) supply voltage; and (c) fuse arcing voltage.

10.3.1i - Pre-arcing I^2t

Before a fuse melts, the fuse is affected only by the current flowing. The pre-arcing or melting I^2t characteristics of fuse links are therefore only a function of prospective fault current and are independent of voltage. For melting times longer than 5-10 ms, the time-current characteristics are usually used for design. Typical time-current characteristics for four different current rated fuses are shown in figure 10.15. For times less than a millisecond, the melting I^2t reduces to a minimum and the pre-arcing I^2t characteristics shown in figure 10.16 are most useful.
The peak let-through current I_p, is a function of prospective fault current I_a for a given supply voltage. Typical current cut-off characteristics are shown in figure 10.17.

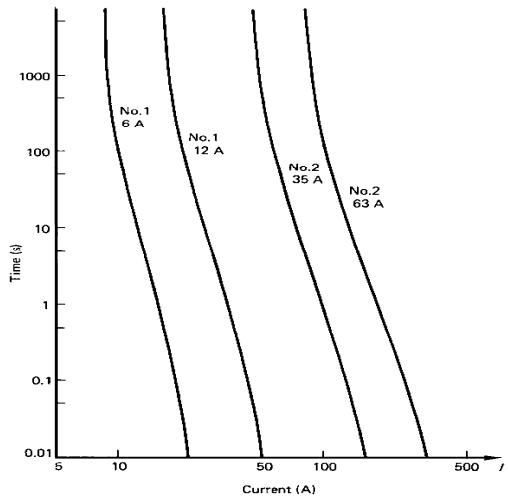


Figure 10.15. *Fuse-link time-current characteristics for 4 fuses and symmetrical sinusoidal 50Hz currents.*

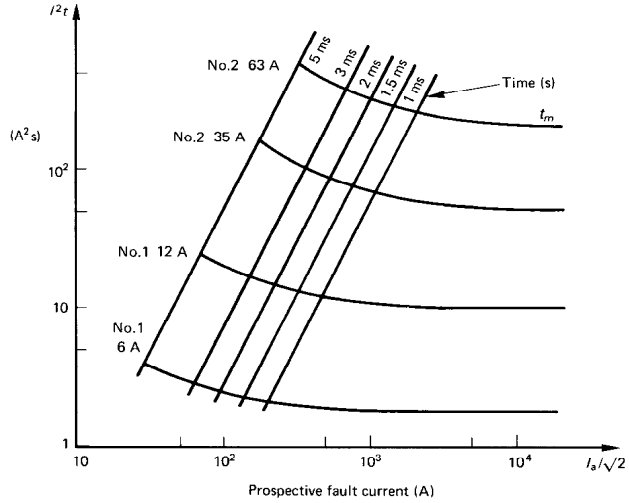


Figure 10.16. *Pre-arcing I^2t characteristics of four fuse links.*

10.3.1ii - Total I^2t let-through

For fuse operating times of less than about 10 milliseconds the arcing I^2t can be considerably larger than the pre-arcing I^2t and it varies considerably with system voltage, fault level, power factor, and the point on the wave when the fault is initiated. The higher the voltage the more onerous is the duty of the fuse link because of the increase in energy absorbed by the fuse link during the arcing process. Under short-circuit conditions this leads to an increase in I^2t let-through with voltage. The I^2t let-through will decrease with increased supply frequency whereas the cut-off current will increase.

The peak arc voltage after melting is usually specified for a given fuse link type and is a function of supply voltage, as indicated by the typical arcing voltage characteristics in figure 10.18. The faster the fault is cleared, the higher the arc voltage V_p. Typical total I^2t let-through values for total operating times of less than 10 ms, at a given voltage, are shown in figure 10.19. Derating factors for temperature, frequency, and power factor are shown in figure 10.20.

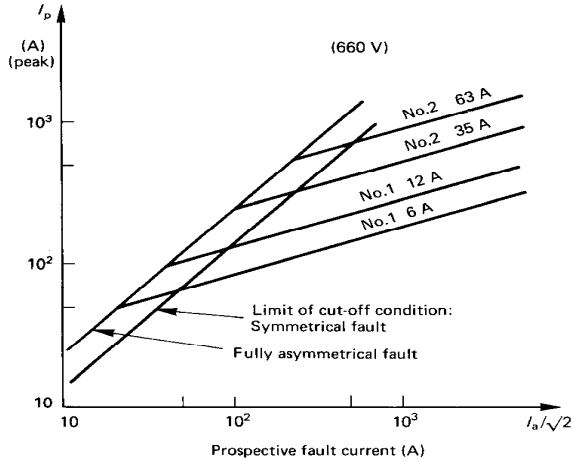


Figure 10.17. *Fuse-link cut-off characteristics at 660 V rms.*

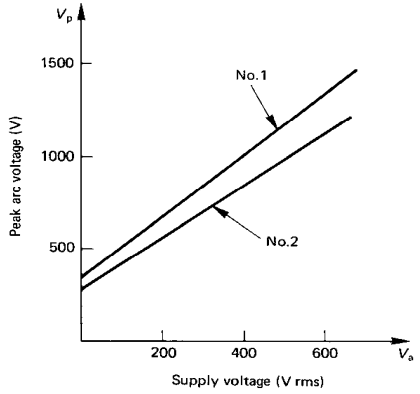


Figure 10.18. *Typical peak arc voltage for two different fuse-link types.*

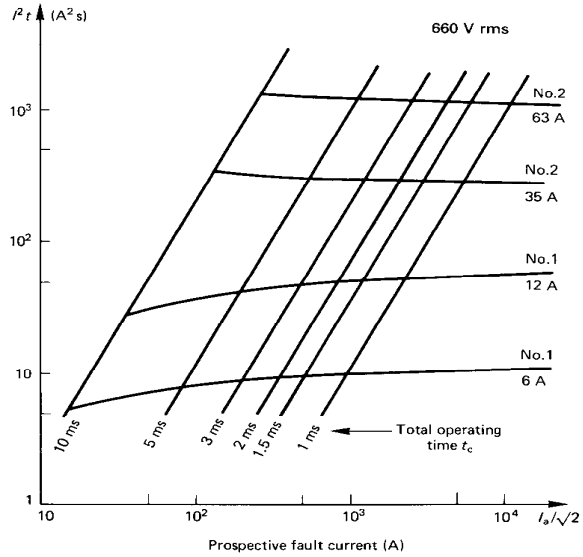


Figure 10.19. *Total let-through current for total fuse-link operating times of less than 10 ms and at 660 V rms.*

10.3.1iii - Fuse link and semiconductor I^2t co-ordination

Difficulties arise in matching fuses with semiconductors because each has very different thermal and electrical properties.

Semiconductor manufacturers publish (mainly for diodes and thyristors) I^2t withstand values for their devices for times less than 10 ms. To ensure fuse link protection the total I^2t let-through by the fuse link under appropriate circuit conditions should be less than the I^2t withstand ability of the semiconductor.

Fuse link manufacturers usually give the data shown in figures 10.15 to 10.20. In ac applications the parameters on which the semiconductor withstand capability is normally compared to the fuse link are

- Peak let-through current versus clearing time or clearing I^2t
- Applied voltage
- Power factor

The voltage rating indicates that the fuse can be relied upon to safely interrupt its rated short circuit current in a circuit where the voltage is equal to, or less than, its rated voltage. The standard voltage ratings used by fuse manufacturers for most small dimension and midget fuses are 32, 63, 125, 250 and 600.

In electronic equipment with relatively low output power supplies, with circuit impedance limiting short circuit currents to values of less than ten times the current rating of the fuse, it is common practice to specify fuses with 125 or 250 volt ratings for secondary circuit protection of 500 volts or higher.

As mentioned previously, fuses are sensitive to changes in current, not voltage, maintaining their “status quo” at any voltage up to the maximum rating of the fuse. It is not until the fuse element melts and arcing occurs that the circuit voltage and available power become an issue.

To summarize, a fuse may be used at any voltage that is less than its voltage rating without detriment to its fusing characteristics.

10.3.1iv – Fuse link derating and losses

For 25ºC ambient temperatures, it is recommended that fuses be operated at no more than 75% of the nominal current rating established using the controlled test conditions. Fuses are essentially temperature-sensitive devices whose ratings have been established in a 25ºC ambient. Even small variations from the controlled test conditions can greatly affect the predicted life of a fuse when it is loaded to its nominal value, usually expressed as 100% of rating.

To compensate for variable operating factors, for trouble-free, long-life fuse protection of equipment, the fuse should not be used at more than 75% of the nominal rating, whilst ensuring overload and short circuit protection must be adequately provided for. The fuse temperature generated by the current passing through the fuse increases with increased ambient temperature. Increased ambient temperature decreases the nominal current rating of a fuse. Most traditional fuse designs use lower melting temperature materials and are, therefore, more sensitive to ambient temperature changes.
The resistance of a fuse is usually an insignificant part of the total circuit resistance. Since the resistance of fractional amperage fuses can be several ohms, this fact should be considered when using them in low-voltage circuits. Most fuses are manufactured from materials which have positive temperature coefficients, and, therefore, it is common to refer to cold resistance and hot resistance (voltage drop at rated current), with actual operation being somewhere in between. Cold resistance is the resistance obtained using a measuring current of no more than 10% of the fuse's nominal rated current. Hot resistance is the resistance calculated from the stabilized voltage drop across the fuse, with current equal to the nominal rated current flowing through it.

The maximum permissible continuous fuse current $\hat{I}$ is dependant on the ambient temperature T_{amb} and the air flow velocity, according to

$$\hat{I} \le I_n \times \left(1 - 0.005 \times \left(T_{amb} - 20°\text{C}\right)\right) \times \left(1 + 0.05v\right) \times K_b \qquad (10.54)$$

where I_n is the fuse rated current and the air velocity, v, is limited to 5m/s. The fuse load constant K_b is assumed worse case, that is 100% conduction, K_b = 1.

In the absence of manufacturer's curves as in figure 10.20a, being a resistive element, fuse losses are related to the square of the current, that is

$$P_{n\%} = \left(\frac{n\% \text{ of } I_{rated}}{100\% \text{ of } I_{rated}}\right)^2 \times P_{100\%} = \left(\frac{I_{load}}{I_n}\right)^2 \times P_{100\%} \qquad (10.55)$$

where $P_{100\%}$ is the fuse losses at rated current I_n in a 20°C ambient.

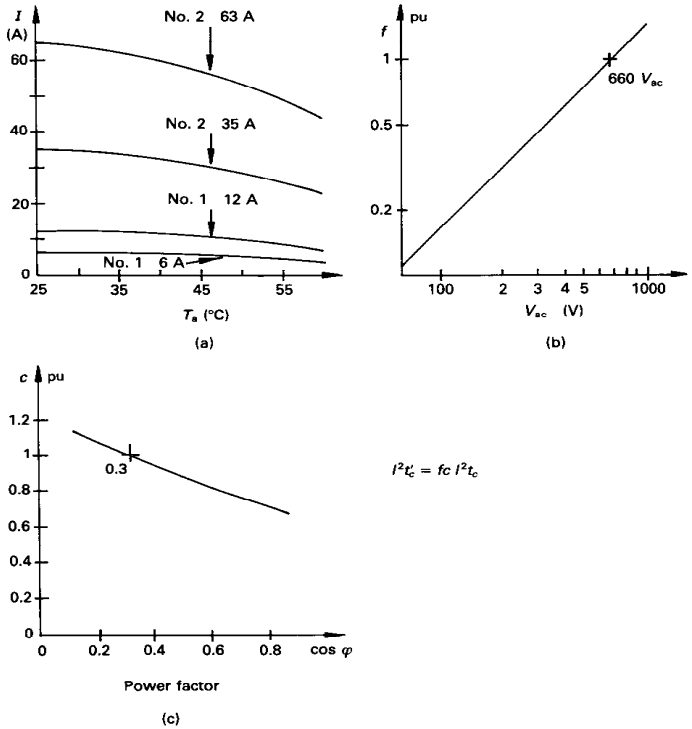


Figure 10.20. *Fuse derating with: (a) ambient temperature; (b) ac supply voltage; and (c) power factor.*

Example 10.5: ***AC circuit fuse link design***

A fast acting fuse is connected in series with a thyristor in a 415 V ac, 50 Hz ac application. The average current in the thyristor is 30 A at a maximum ambient temperature of 45°C. The ratings of the thyristor are

$I_{T(AV)}$ = 45 A @ T_c = 85°C
I_{TRMS} = 80 A
I^2t = 5 k A^{2}s for 10 ms @ 125°C
I^2t = 20 k A$^2\sqrt{}$s
I_{TSM} = 1000 A for 10 ms @ 125°C and V_{RRM}= 0

The fault circuit inductance is 1.32 mH and the resistance is negligible. Using the figures 10.15 to 10.20, select a suitable fuse.

Solution

From figure 10.20a, the 35 A rms No. 2 fuse is rated at 30 A rms in a 45°C ambient.
From figure 10.18 the peak arc voltage for a type No. 2 fuse will be less than 1200 V, hence the thyristor voltage rating must be greater than 1200 V and possibly 1200V+√2×415V ac, depending on the point-on-wave of the fault and the particular circuit configuration.
The short circuit or prospective rms symmetrical fault current is

$$I_{sc} = \frac{I_{sc}}{\sqrt{2}} = \frac{V_s}{X_L} = \frac{415\text{V}}{2\pi \times 50\text{Hz} \times 1.32\text{mH}} = 100\text{A}$$

Figure 10.16 gives a fuse peak let through current of 500 A, which is less than the thyristor peak current rating, I_{TRMS}, of 1 kA.
Figure 10.19 gives the fuse total I^2t of 300 A^{2}s and the total clearing time of t_c =3.5 ms. Since the fuse clears in less than 10 ms (½ ac cycle), the thyristor re-applied V_{RRM} will be zero and an I_{TSM} = 1000 A rating is applicable. The total I^2t is corrected for voltage (415V ac) and power factor (0 pu) with f = 0.6 and c = 1.2 from figures 10.20b and c.

$$I^2t' = f \times c \times I^2t = 0.6 \times 1.2 \times 300\ \text{A}^2\text{s} = 216\ \text{A}^2\text{s}$$

which is significantly less than the thyristor I^2t rating of 5 kA^{2}s.
Since t_c is less than 10 ms, the $I^2\sqrt{t}$ rating of the thyristor is used.

$$I^2t'' = (I^2\sqrt{t})\sqrt{t_c}$$
$$= 20\ \text{kA}^2\sqrt{\text{s}} \times \sqrt{3.5\ \text{ms}} = 1.18\ \text{kA}^2\text{s}$$

which is significantly greater than the I^2t (216 A^{2}s) of the fuse.

Since the fuse peak let through current (500 A) is less than the thyristor peak surge current rating (1000 A), and the fuse I^2t rating (216 A^{2}s) is significantly less than that for the thyristor (1180 A^{2}s), the proposed 35 A fast acting fuse should afford adequate protection for the thyristor.
Generally, if the rms current rating of the fuse is less than the average current rating of the thyristor or diode, the fuse will provide adequate protection under fault conditions.

♣

10.3.1v – Pulse derating

Here the general term 'pulses' is used to describe the broad category of wave shapes referred to as 'surge currents', 'start-up currents', 'inrush currents', and 'transients'. Electrical pulse conditions vary considerably from one application to another and different fuse constructions may not react the same to a given pulse condition. Electrical pulses produce thermal cycling and possible mechanical fatigue that could affect the life of the fuse. Initial or start-up pulses are normal for some applications and require the characteristic of a 'slow blow' fuse, incorporating a thermal delay design to enable it to survive normal start-up pulses and still provide protection against prolonged overloads. The start-up pulse should be defined and then compared to the time-current curve and I^2t rating for the fuse.
Nominal melting I^2t is a measure of the energy required to melt the fusing element and is expressed as 'Ampere squared seconds' (A^{2}s). This nominal melting I^2t, and the energy it represents (within a time duration of 8ms [0.008 s] or less and 1ms [0.001 s] or less for thin film fuses), is a value that is constant for each different fusing element. Because every fuse type and rating, has a different fusing element, it is necessary to determine the I^2t for each. This I^2t value is a parameter of the fuse itself and is controlled by the element material and the configuration of the fuse element. In addition to selecting fuses on the basis of normal operating currents, derating, and ambient temperature, it is also necessary to apply the I^2t design approach. This nominal melting I^2t is not only a constant value for each fuse element design, but it is also independent of temperature and voltage. Most often, the nominal melting I^2t method of fuse selection is applied to those applications in which the fuse must sustain large current

pulses of a short duration. These high-energy currents are common in many applications and are critical to the design analysis.

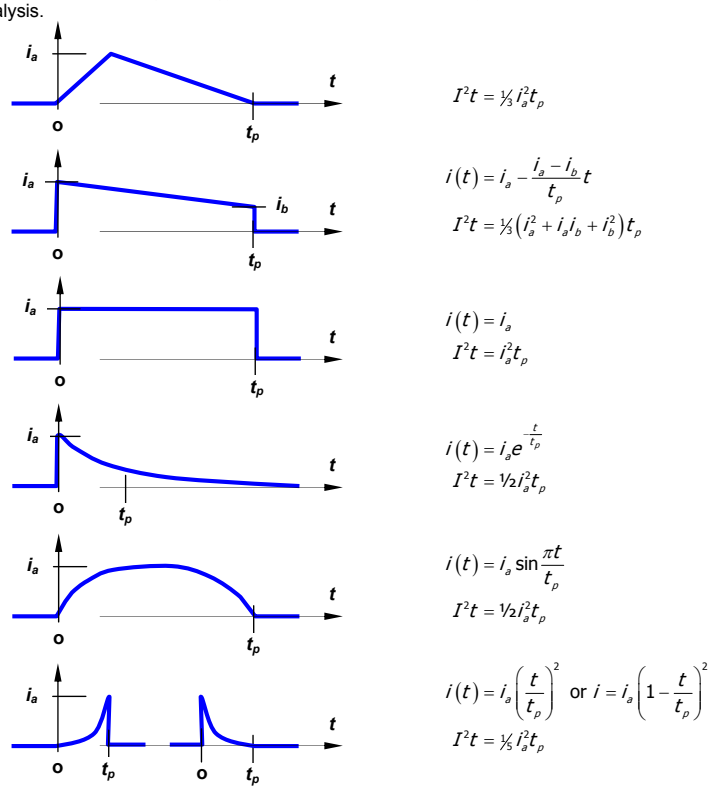


Figure 10.21. *Fuse I^2T formula for various current waveforms.*

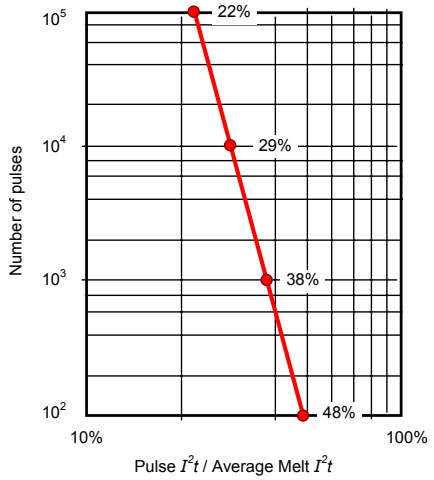


Figure 10.22. *Fuse pulse number derating curves, assuming adequate cooling time between pulses.*

The following example illustrates the application of I^2t for a fusing undergoing repetitive surges.

Example 10.6: ***AC circuit fuse link design for I^2t surges***

Based on figures 10.16 and 10.21, select a 230V, very fast-acting fuse that is capable of withstanding 100,000 pulses of current having a triangular pulse waveform of 20A magnitude and of 3ms duration. The normal operating current is 4.5A at an ambient temperature of 25ºC.

Solution

At 25ºC, no fuse thermal derating is necessary.
Waveform D and the associated effective pulse I^2t formula in figure 10.21 are applicable, where i_p=50A and the effective duration is 3ms.
The applicable value for peak pulse current i_p and time t into the corresponding formula for wave-shape D in figure 10.21 gives:

$$I^2t = \frac{1}{3} i_p^2 t$$
$$= \frac{1}{3} \times 20^2 \times 3\text{ms} = 0.40\text{A}^2\text{s}$$

This value is referred to as the *pulse I^2t.*

The required nominal melting I^2t value for 100,000 occurrences of the calculated pulse I^2t from figure 10.22, involves a derating figure of 22%. The calculated pulse I^2t is converted to the necessary nominal melting I^2t values as follows:

$$\textit{Nom. Melt } I^2t = \frac{\textit{Pulse } I^2t}{.22}$$
$$= \frac{0.4}{0.22} = 1.82\ \text{A}^2\text{s}$$

Examine the I^2t rating data for a 230V, very fast-acting fuse. From figure 10.16, 6A design is rated at 2A^{2}s., which is the minimum fuse rating that will accommodate the 1.82A^{2}s value calculated. This 6A fuse will also accommodate the specified 4.5A normal operating current, when a 25% derating factor is applied to the 6A nominal rating.

♣

10.3.1vi - Other fuse link derating factors

Ambient temperature correction coefficient, A_1
Fuse current ratings are usually established at a reference ambient air temperature T_{ref} of 25°C or 30°C. Typical ambient operating temperatures are greater, T_{op}, so the fuse must be derated. The temperature rise depends on the internal power dissipation, which is a function of the current squared. The derating coefficient, for a maximum allowable fuse temperature of T_{max} (typically 130°C to 150°C, is

$$A_1 = \sqrt{\frac{T_{\max} - T_{op}}{T_{\max} - T_{ref}}}$$

Forced cooling correction coefficient, B_v
Forced air cooling, up to a limit about v=5m/s, increases the fuse continuous rating by up to 25%, according to:

$$B_v = 1 + 0.05v \le 1.25$$

Terminal conductor size coefficient C_1
Cable and busbar size conduct heat away from the fuse, thereby affecting the fuse temperature. A factor 0.8 to 1 can account for busbar conduction and the effects of nearby heat sources. Liquid cooling of the terminals can result in a correction factor of greater than one.

High frequency derating coefficient C_f
The fuse link element is a metal strip, in which at fundamental frequencies *f* above 1kHz, its resistance is increased by skin effects. The I^2R losses are increased, where the current includes all harmonics, which should not exceed 15% more than the fundamental. The derating is applicable from 100Hz up to 20kHz is shown in the following table.

f Hz	C_f
$0 < f \le 100$	1.0
$100 < f \le 500$	0.95
$500 < f \le 1{,}500$	0.9
$1{,}500 < f \le 5{,}000$	0.8
$5{,}000 < f \le 10{,}000$	0.7
$10{,}000 < f \le 20{,}000$	0.6

Current-variation coefficient A_i
Large rms current variation cause thermal fatigue of small notch zones on the fuse link used for semiconductor fuses. The thermal derating is classified as either continuous or cyclic.
On/off operation a few time per day is considered continuous operation with minimal overloads, has an associated 20% derating, A_i =0.8. Equipment turned on and off once per day, or less often, is fuse derated by 10%, A_i =0.9.
Cyclic loading is when the fuse heats and cools to steady-state at a cycle rate of less than a few tens of minutes, $A_i \le 0.6$.

The fuse current rating for a nominal current I_i is derated to a maximum rms continuous current of

$$I_n' = I_n \times A_1 \times B_v \times C_1 \times C_f \times A_i$$

Example 10.7: ***AC circuit fuse link derating***

A 1000A fuse has following operational data:
Maximum operating temperature T_{max} =150°C
reference temperature T_{ref} =30°C
Modest busbars giving C_1=0.85

The operational environment is:
Ambient temperature T_{op} =50°C
Fundamental frequency f=1kHz
Forced air cooling velocity 2m/s
Operation: cyclic every hour, A_i =0.6

What is the maximum allowable continuous current for the fuse?

Solution

From the frequency derating table C_f=0.9 at 1kHz.

$$A_1 = \sqrt{\frac{T_{max} - T_{op}}{T_{max} - T_{ref}}} = \sqrt{\frac{150°C - 50°C}{150°C - 30°C}} = 0.91$$

$$B_v = 1 + 0.025v = 1 + 0.05 \times 2\text{m/s} = 1.1$$

$$C_f = 1 - 0.075 \times \log_{10} f$$
$$= 1 - 0.075 \times \log_{10} 1000\text{Hz} = 0.775$$

The fuse adjusted rating is

$$I_n' = I_n \times A_1 \times B_v \times C_1 \times C_f \times A_i$$
$$= 1000\text{A} \times 0.91 \times 1.1 \times 0.85 \times 0.9 \times 0.6 = 460\text{A}$$

The adjusted rating of 460A is significantly lower than the 1000A applicable to rated fuse conditions.

♣

10.3.1vii – Fuse link dc operation
Fuse link protection in dc circuits presents greater difficulty than for ac circuits. No natural ac period current zeros exist and faults can result in continuous arcing. The breaking capacity of a fuse link in a dc application depends on

- the maximum applied dc voltage, $\hat{E}$
- the feed *L /R* time constant, τ
- the prospective short circuit current of the circuit, I_a

High-speed semiconductor ac fuses can be used in dc applications, after suitable derated. The longer the fault current *L /R* time constant, the lower the allowable operating voltage, since the fuse takes longer to melt due to the slower energy delivery rate. Conversely, the higher the prospective short-circuit current I_a, the faster the fuse operates hence it can operate at a higher dc voltage level.
Typically, the fuse dc rating is 70% of its ac voltage rating for time constants between 10ms to 20ms, and the dc rating decreases as the time constant increases. No voltage derating is necessary for time constants less than 2½ms.
Published fuse characteristics and performance data generally concentrate on ac at 50 Hz or 60 Hz. values. The design monograph in figure 10.23 can be used to select a suitable ac high-speed fuse for dc application. The design requires the fault time constant $\tau = L/R$, which will specify the maximum allowable dc voltage $\hat{E}$, whence the maximum dc arcing voltage $\hat{V}_{arc}$. The fault time constant also specifies the pre-arcing I^2t derating factor *k*, used to specify the minimum prospective fault current $\hat{I}_a$ to ensure enough energy for the fuse to melt, thence clear.

$$\hat{I}_a = k\sqrt{I^2 t} \qquad (10.56)$$

This minimum current must be less than the prospective peak dc fault current given by

$$I_a = \frac{E}{R} \qquad (10.57)$$

That is, $\hat{I}_a < I_a$ is a fuse link requirement.

The instantaneous fault current is $i(t) = I_a(1-e^{-t/\tau})$, while the instantaneous rms current is

$$i_{rms}(t) = I_a\sqrt{1 + \frac{2e^{-n}}{n} - \frac{e^{-2n}}{2n} - \frac{1.5}{n}} \quad \text{where} \quad n = t/\tau \text{, the number of time constants} \qquad (10.58)$$

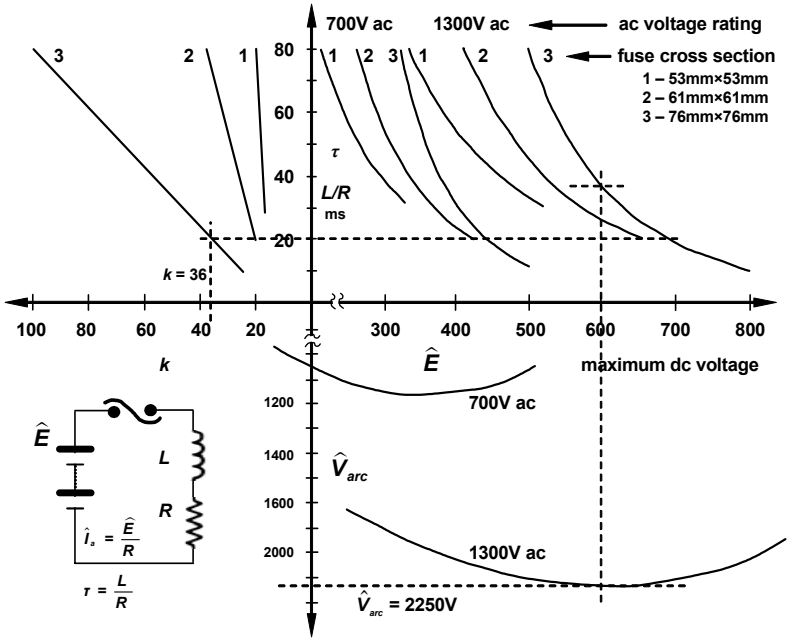


Figure 10.23. *Design curves for an ac fused used in dc applications.*

Example 10.8: ***DC circuit fuse link design***

A traction 600V dc supply has an equivalent source impedance of 20mΩ and 0.4mH, and a nominal dc load current of 600A.

i. Validate the suitability of the following ac fuse in being able to safely clear a dc fault current.
ii. Estimate the fuse losses at 20°C ambient.
iii. What is the maximum nominal current allowable in an air still 80°C ambient?
iv. Estimate the fuse losses in the 80°C ambient.

FUSE: High speed 900A, 1300V ac, with a pre-acing I^2t of 505,00A^{2}s at room temperature, in a case size 3 of cross section 75mm×76mm, giving 125W of losses at 20°C.
Figure 10.23 is applicable to this fuse link element.

Solution

The maximum applied voltage is $\hat{E}$ = 600V dc

The short circuit fault time constant is $\tau = L/R = 0.4\text{mH}/20\text{m}\Omega = 20\text{ms}$

i. From figure 10.23, a size 3 fuse will offer better voltage and current overheads than a type 2 fuse. The data yields *k* = 36, an arcing voltage maximum of 1920V dc, and would allow fault time constants of up to 36ms or peak dc supply voltages of up to 700V dc.
The prospective short circuit fault current from equation (10.57) is

$$I_a = E/R = 600\text{V}/20\text{m}\Omega = 30\text{kA}\ .$$

From equation (10.56), the minimum allowable fault current to ensure enough energy to melt and clear the fuse is

$$\breve{I}_a = k\times\sqrt{I^2t}$$
$$= 36\times\sqrt{505{,}00} = 25.6\text{kA}$$

Since $\breve{I}_a < I_a$, that is, 25.6kA < 30kA, the fuse will reliably and predictably melt, thence clear.

ii. The 125W fuse loss at rated current of 900A is reduced if the nominal load current is 600A. From equation (10.55):

$$P_{n\%} = \left(\frac{I_{load}}{I_n}\right)^2 \times P_{100\%}$$
$$= \left(\frac{600\text{A}}{900\text{A}}\right)^2 \times 125\text{W} = 55\tfrac{1}{2}\text{W}$$

iii. At ambient temperatures above 20°C, the fuse nominal current rating is decreased according to equation (10.54):

$$\hat{I} \le I_n\times\left(1-0.005\times\left(T_{amb}-20°\text{C}\right)\right)\times\left(1+0.05v\right)\times K_b$$
$$\le 900\text{A}\times\left(1-0.005\times\left(80°\text{C}-20°\text{C}\right)\right)\times\left(1+0.05\times 0\right)\times 1$$
$$\le 900\text{A}\times\left(1-0.005\times 60°\text{C}\right) = 630\text{A}$$

Thus the fuse would be satisfactory at 80°C with the nominal load current of 600A dc.

iv. The fuse losses at 600A in an 80°C would be approximately

$$P = \left(\frac{I_{load}\ @80°\text{C}}{\hat{I}\ @80°\text{C}}\right)^2 \times P_{100\%}$$
$$= \left(\frac{600\text{A}}{630\text{A}}\right)^2 \times 125\text{W} = 113\text{W}$$

♣

10.3.1viii – Alternatives to dc fuse operation

It may be possible in some applications to use an ac fuse in a dc circuit, before the rectification stage. Generally low voltage fuses are more effective than high voltage fuses. In high voltage transformer applications satisfactory protection may be afforded by transferring the fuse to the low voltage side. The fuse I^2t rating is transferred as with impedance transferring, that is, in the turns ratio squared.

$$I^2t_{fuse}^{primary} = \left(\frac{V_s}{V_p}\right)^2 \times I^2t_{semiconductor}^{secondary} \qquad (10.59)$$

Alternatively an mcb (miniature circuit breaker) may offer better protection in cases when the ac fault is more of an overload such that the current magnitude is limited. On overload, the mcb takes a longer time to clear than a fuse, thus the mcb is less prone to nuisance tripping. Aspects of both dc and ac relays and contactors are presented in Chapter 27.
Fuse protection is mainly applicable to more robust devices such as thyristors and diodes. Transistors (MOSFETs more readily than IGBTs, even when avalanche rated) usually fail as a result of over-current before any fuse link can clear the fault.

10.3.2 Protection with resettable fuses

Resettable fuses are basically thermistors. Thermistors are thermally sensitive resistors and have, according to type, a negative (NTC), or positive (PTC) resistance/temperature coefficient.
PTC (positive temperature coefficient) thermistors are ceramic or polymeric crystalline protection components whose electrical resistance rapidly increases as a certain temperature is exceeded.
Over-current circuit protection can be accomplished with the use of either a traditional fuse-link or PTC device. PTC devices are typically used in a wide variety of electronics applications where over-current events are common and automatic resettability is desired. This ability of a PTC to reset itself after experiencing a fault current makes it ideal within circuits that are not readily accessible or where a constant uptime is required.
There are two types of PTC thermistors based on different underlying materials: polymer and ceramic, as summarised in Table 10.2. Generally the device cross-sectional area determines the surge current capability, and the device thickness determines the surge voltage capability.

Thermal Properties

The operation of all PTC devices is based on an overall energy balance described by equation (10.60), which assumes a uniform temperature distribution within the device:

$$H\frac{\Delta T}{\Delta t} = I^2R - U\times(T - T_A) \qquad (\text{W}) \qquad (10.60)$$

where

I = current flowing through the device, A
R = resistance of the device, Ω
Δt = change in time, s
H = *heat capacity,* $H = m \times c_p$, J/ K
m = mass of the device, kg
c_p = heat capacity of the PTC thermistor device, J/kg K
ΔT = change in device temperature, K
T = device temperature, K
T_A = ambient temperature, K
U = effective heat-transfer coefficient, heat dissipation factor, $U=h\times A$, W/K

In equation (10.60), the current flowing through the device generates heat at a rate equal to I^2R. All or some of this heat is lost to the environment, at a rate described by the term $U\times(T - T_A)$. Any heat not lost to the environment raises the device temperature at a rate described by the term $mc_p\times(\Delta T/\Delta t)$.
If the heat generated by the device and the heat lost to its environment balance, then in this steady-state equilibrium condition, $\Delta T/\Delta t$ tends to zero and equation (10.60) reduces to:

$$I^2R = U\times(T - T_A) \qquad (\text{W}) \qquad (10.61)$$

Under normal operating conditions, this thermal steady-state is at a relatively low temperature and in the low resistance region, as indicated by operating *Point 1* in Figure 10.24.

The thermal characteristics of PTC devices are similar to those of the NTC devices, and can be described by the following terms:

- heat capacity, *H,* in J/K
- heat dissipation constant, *D,* in W/K
- thermal time constant, *τ,* in seconds

Heat capacity

The product of the specific heat and mass of the thermistor, heat capacity is the amount of heat required to produce a 1K change in the body temperature of the thermistor.

$$H = m\times c_p$$

Ceramic PTC thermistors have a heat capacity of about 3J/cm^3/K.

Heat dissipation constant
The heat dissipation factor is the amount of heat which is lost, over a unit of time, based on a 1K temperature difference between the heating element and ambient temperature.
It is the ratio of the change in the power applied to the thermistor to the resulting change in body temperature due to self-heating. The factors that affect the dissipation constant including: lead-wire materials, method of mounting, ambient temperature, conduction or convection paths between the device and its surroundings, and the structure, shape, and material of the PTC device itself.

$$P = IxV = U(T - T_A) = U\Delta T \qquad \text{(W)}$$

T: temperature of heating element, K
U: heat dissipation factor, W/K

Thermal time constant
The time required for the thermistor temperature to change 63.2% of the difference between the self-heated temperature and the ambient after the power is disconnected. The thermal time constant is also influenced by the same environmental factors as those that affect the dissipation constant.
The thermal time constant τ is the time to reach 0.632 times the temperature difference between T_o and T_A.

$$\tau = \frac{H}{U} \qquad \text{(s)} \qquad (10.62)$$

U: Heat dissipation factor, W/K
H: Heat capacity, J/K

10.3.2i Polymeric PTC devices

Polymeric PTC materials
Polymeric positive temperature coefficient circuit protectors are made from a conductive plastic formed into thin sheets, with electrodes attached to either side of the compressed stacked sheets. The conductive plastic matrix is manufactured from a nonconductive crystalline polymer and dispersed highly-conductive carbon black particles, typically high density polyethylene mixed with graphite. The plate electrodes ensure even distribution of power loss through the device, and provide a surface for leads to be attached or for surface mounting. The phenomenon that allows conductive plastic materials to be used for resettable over-current protection devices is that they exhibit a large non-linear positive temperature coefficient effect when heated. A PTC is a thermal characteristic that many materials exhibit whereby resistance increases with temperature. What makes a PTC conductive plastic material unique is the magnitude of its resistance increase. At a specific transition temperature, the increase in resistance is so large that it is characterised on a logarithmic scale.

Resettable over-current polymeric PTC protector physics
The conductive carbon black filler material in the PTC device is dispersed in a polymer that has a crystalline structure. The crystalline structure densely packs the carbon particles into its crystalline boundary so they are close enough together (beyond a level called the percolation threshold) to allow current to flow through the polymer insulator via the created carbon 'chains', due to a tunnelling effect. When the conductive plastic is at room temperature, there are numerous carbon chains forming parallel conductive paths through the mostly crystalline material.
Under electrical fault conditions, excessive current flows through the PTC device. Internal I^2R joule heating causes the conductive plastic material's temperature to rise. As this self-heating continues, the material's temperature continues to rise until it exceeds its phase transformation temperature. As the material passes through this phase transformation temperature, the densely packed crystalline polymer matrix changes to an amorphous structure, disrupting the network of conductive carbon paths. This phase change is accompanied by a small volumetric expansion. As the conductive particles move apart from each other, most of them no longer contact to conduct current and the resistance of the device increases sharply.
The material will stay 'hot', remaining in this high resistance state as long as the power is applied. This latched state provides continuous fold-back protection, until the electrical fault is cleared and the power is diminished. Reversing the phase transformation, by cooling, allows the carbon chains to re-form as the polymer re-crystallizes. The resistance quickly reduces toward its original low value.

Principle of operation
Both polymeric positive temperature coefficient thermistor protectors and traditional fuse-link devices react to internal I^2R joule heat generated by an excessive current flow in a circuit. Whereas a fuse-link melts open, interrupting the current flow, a PTC device restricts current flow as its bulk rises in temperature, changing from a low to a high resistance state. In both cases, this transitional condition is termed *tripping*. The characteristic curve in figure 10.24 shows the typical response of a PTC device to temperature.

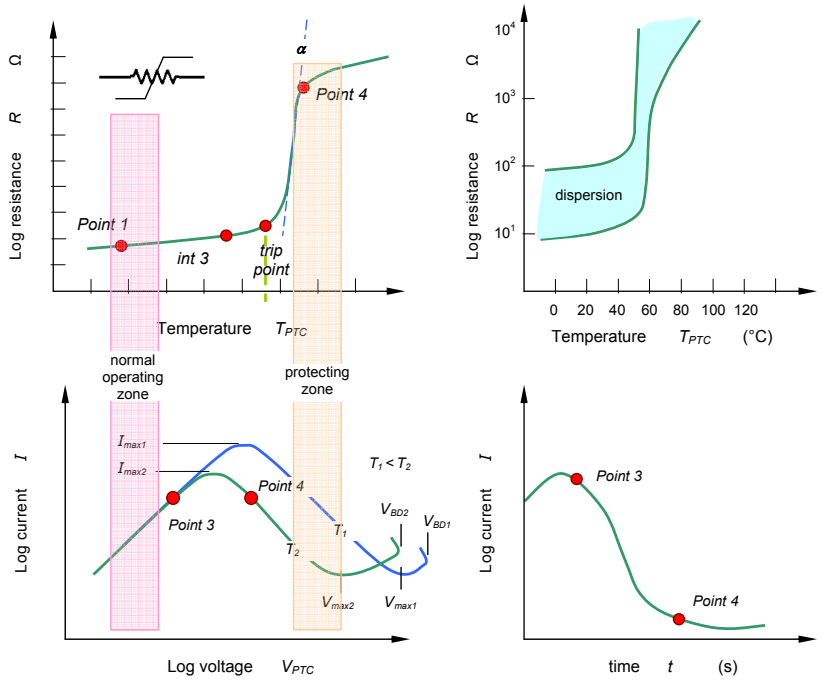


Figure 10.24. *Polymeric PTC thermistor operating R-V-I-t curves and typical tripping dispersion.*

If the current through the device is increased while the ambient temperature is maintained constant, the heat generated within the device increases and the temperature of the device also increases. Whilst the increase in current is modest, if the generated heat can be lost to the environment, the device will stabilize according to equation (10.61) at a higher temperature, such as *Point 2* in Figure 10.24.
Alternatively, instead of the current being increased, the ambient temperature is raised, the device will stabilize according to equation (10.61) at a higher temperature, possibly again at *Point 2*. *Point 2* could also be attained by a combination of both a current increase and an ambient temperature increase.
Further increase in either current, ambient temperature, or both will cause the device to reach a temperature where the resistance begins to rapidly increases, such as at *Point 3* in Figure 10.24. Any further increase in current or ambient temperature will cause the device to generate heat at a rate greater than the rate at which heat can be transferred to the environment, thus causing the device to heat up rapidly. A large increase in resistance occurs for a small change in temperature. In Figure 10.24, this region of large change in resistance for a small change in temperature occurs between points 3 and 4, and this operating region is termed the tripped state. This large increase in resistance causes a corresponding decrease in the current flowing in the circuit.
The resultant current reduction reduces the likelihood of circuit damage. Since the temperature change between operating points 3 and 4 is small, the term $(T - T_A)$ in equation (10.61) can be replaced by the constant $(T_{op} - T_A)$, where T_{op} is the operating temperature of the device. Then equation (10.60) becomes:

$$I^2R = \frac{V^2}{R} = U \times (T_{op} - T_A) \qquad \text{(W)} \qquad (10.63)$$

Since both U and $(T_{op} - T_A)$ are now constants, equation (10.63) reduces to a constant, I^2R = constant; that is, the device now operates in a constant power state. Expressing this constant power as V^2/R emphasizes that, in the tripped state, the device resistance is proportional to the square of the applied voltage. This relation holds until the device resistance reaches the upper knee of the curve, *Point 4* in Figure 10.24.
For a device that has tripped, as long as the applied voltage is high enough for the resulting V^2/R power to maintain the $U\times(T_{op} - T_A)$ loss, the device remains in the tripped state, that is, the device will remain latched in its protective high-resistance state. When the voltage is decreased to the point where the $U\times(T_{op} - T_A)$ loss can no longer be supplied, the device begins to reset to a lower resistance state, by traversing back along the R-T characteristic towards *Point 1*.

Electrical Properties
The electrical characteristics describing PTC devices (ceramic and polymeric) include the following:

- current-time characteristic
- resistance-temperature characteristic
- voltage-current characteristic
- power and minimum resistance
- temperature coefficient of resistance
- transition temperature
- voltage and frequency dependence
- voltage rating

Hold and trip current
Figure 10.25 illustrates the hold-current and trip-current behaviour of PTC devices as a function of device bulk temperature.
Region A represents the combinations of current and temperature at which the PTC device will trip (go into the high-resistance state) and protect the circuit. *Region B* describes the combinations of current and temperature at which the PTC device will allow for normal operation of the circuit, a low resistance state. In *Region C*, it is possible for the device to either trip or remain in the low-resistance state, depending on the individual device resistance. The boundaries between these regions are defined as the hold and trip currents.

Hold current, I_{Hold}, at a given temperature, is the highest steady-state current that a device will hold for an indefinite time without transitional tripping from the low resistance to the high resistance state.
Trip current, I_{Trip}, is the minimum current at which the device will switch from the low resistance to the high resistance state.

The trip current is typically greater than the normal operating current. Unlike time-to-trip, the hold current of a device is a steady-state condition that can be fairly accurately defined by the heat transfer environment. Under steady-state conditions, equation (10.63) is valid and the I^2R heat generated equals the heat lost to the environment. Therefore, if U increases, the hold current increases, with the approximate dependency:

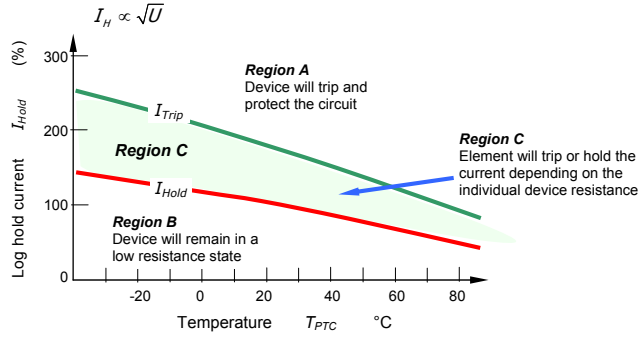


Figure 10.25. *Polymeric PTC device hold and trip currents as a function of device temperature.*

Since PTC devices are thermally activated, any change in the surrounding temperature will affect device performance. As the surrounding temperature increases, less energy is required to trip the device, thus the hold current decreases. Consequently the I_{Trip} and I_{Hold} curves both have negative slopes in Figure

10.25. Thermal derating curves and I_{Hold} versus temperature tables enable in-circuit design over a wide range of temperatures, as illustrated in example 10.8.
The heat transfer for PTC devices is affected by several factors, for example:

- An increase in the ambient temperature surrounding the device results in a reduction in heat transfer. This can be caused by a general increase in the ambient temperature, or the device being in proximity to a heat-generating source such as a power switching device, resistor or transformer. The hold current, power dissipation, and time-to trip of the device are all reduced.
- By changing the width/area of pcb copper pads or increasing the device lead lengths, which are in thermal contact with the device. A surface mount device on an increased copper pad area, results in an increase in the heat transfer. This results in a higher hold current and power dissipation, and a slower time-to trip.
- If the airflow around the device is increased, heat transfer is increased.

Time-to-trip
The time-to-trip of a PTC device is the time it takes for the voltage drop across the device to rise to greater than 80 percent of the voltage of the power source, or when the resistance of the device increases substantially relative to the load resistance. A trip event is caused when the rate of heat lost to the environment is less than the rate of heat generated, causing the device temperature to increase. The rate of temperature rise and the total energy required to make the device trip, depend on the fault current and the heat transfer environment.
For low-fault currents, for example two-to-three times the hold current, devices trip slowly since a substantial amount of the I^2R generated heat is lost to the environment, therefore only slowly increases the device temperature. This type of trip event can be considered as a non-adiabatic trip event. Under these conditions, the heat transfer to the environment plays a significant role in determining the time-to-trip of the device. The greater the heat transfer, the longer the time-to-trip.
At high-fault currents, for example ten times the hold current, the time-to-trip is reduced because most of the I^2R energy generated is retained in the device, thereby rapidly increasing its temperature. A trip event of this kind can be regarded as an adiabatic trip event. Under these conditions, the heat transfer to the environment is less significant in determining the time-to-trip of the device.
As tripping is a dynamic event, it is difficult to precisely predicted the change in the time-to-trip since a change in the heat transfer coefficient is often accompanied by a change in the thermal mass around the device. If for example the device uses a metal heatsink, not only will the heat transfer increase, but the device will also need to heat some fraction of the metal (due to the intimate thermal contact) before the device will trip. Therefore, not only is the thermal conductivity of the metal important, but the heat capacity of the metal is also a factor in determining the time-to-trip.
The switching time or time-to-trip t_s can be approximated, in an adiabatic condition, by:

$$t_s = \frac{c_p Vol}{P}(T_{ref} - T_A) \qquad (10.64)$$

T_{ref} reference temperature of PTC thermistor
Vol PTC thermistor volume
P switch-on power of the PTC thermistor

This equation shows that the switching time is influenced by the size of the PTC thermistor, its reference temperature and the power supplied. Switching times are lengthened by increasing the volume or the reference temperature; while a high power consumption by the PTC thermistor results in short switching times.

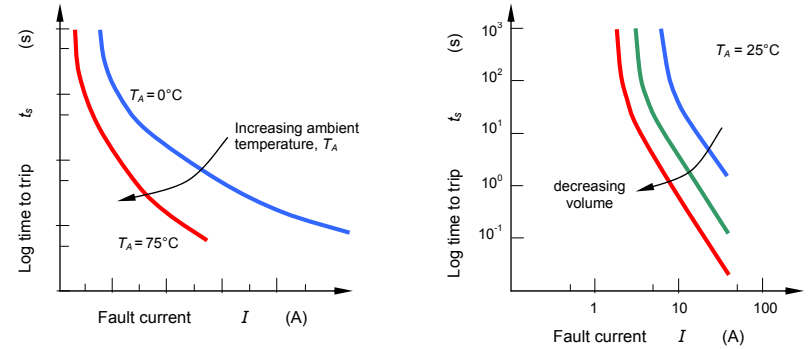


Figure 10.26. *Polymeric PTC curves for a PTC rated at 72V, 40A, with a 1.1A to 3.75A hold current.*

Figure 10.26 shows a typical pair of operating curves for a PTC device in still air at 0°C and 75°C. The curves are separate because the heat required to trip the device comes both from electrical I^2R heating and from the device environment. At 75°C the heat input from the environment is substantially greater than at 0°C, so the additional I^2R needed to trip the device is correspondingly less, resulting in a lower trip current at a given trip time, or a faster trip at a given trip current.

Device reset time

In Figure 10.27, after a trip event (when the current fault condition has been alleviated), the resistance recovery to a quasi-stable low value is rapid, with most of the recovery typically occurring within a few minutes. Figure 10.27 shows the resistance recovery curve and associated power dissipation for a family of leaded PTC fuse devices.

As with other electrical properties, the resistance recovery time depends upon both device design and the thermal environment. Since resistance recovery is related to device cooling, the greater the heat transfer, the quicker the recovery.

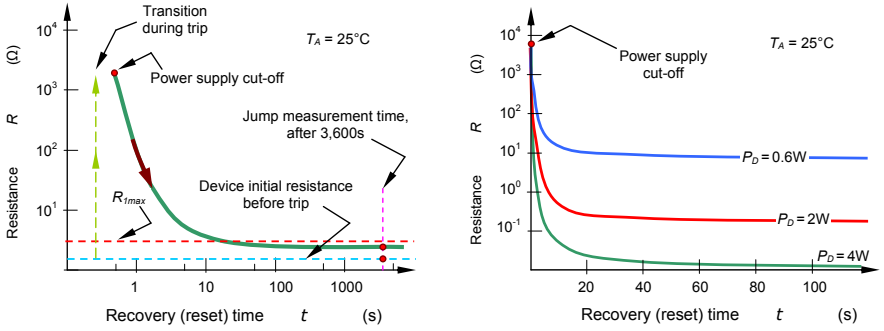


Figure 10.27. *Polymeric PTC typical resistance recovery after a trip event.*

Typical recovery resistance after a trip event: trip jump, R_{1MAX}

PTC devices exhibit resistance hysteresis after tripping, either through an electrical trip event or through a thermal event such as reflow soldering.

Figure 10.27 shows typical polymeric PTC device behaviour after tripping and when cooling. It can be seen that even after a number of hours, the device resistance is still greater than the initial pre-trip resistance. Over an extended period, device resistance will continue to fall and will eventually approach the initial resistance. Therefore, when PTC devices are being used, this 'trip jump' or 'reflow jump' is taken into consideration when determining the hold current. This increase in resistance is defined as R_{1MAX} and the jump is measured one hour after the thermal fault event. The long-term cold resistance of polymeric PTC thermistors increases with successive trip events.

10.3.2ii Ceramic PTC devices

Unlike PTC thermistors made of plastic materials, that is, polymeric materials, ceramic PTC thermistors always return to their initial resistance value, even after frequent heating/cooling cycles.

The thermal properties and many of the electrical properties are characterised the same for both PTC material types.

Ceramic PTC thermistors

Mixtures of barium carbonate, titanium oxide and other materials (become doped polycrystalline ceramic, containing $BaTiO_3$ - 69% plus titanates of Pb - 15%, Sr - 10%, Ca - 5% and 1% dopants), whose composition produces the desired electrical and thermal characteristics are ground, mixed and compressed into various shapes. These blank parts are then sintered, at temperatures just below 1400°C, and after cooling, they are contacted, provided with connection elements, and finally coated or encased. Multilayer or bulk ceramic types are available.

Generally, ceramic is a good insulating material with a high resistance. Semi-conduction and thus a low resistance are achieved by doping the ceramic with materials of a higher valency than that of the crystal lattice. Some of the barium and titanate ions in the crystal lattice are replaced by these higher valencies to obtain a specified number of free electrons which make the ceramic conductive.

The material structure is composed of many individual crystallites. At the edge of these monocrystallites, the so-called grain boundaries, potential barriers are formed. They prevent free electrons from diffusing into adjacent areas. The result is high resistance at the grain boundaries. However, this effect is neutralized at low temperatures. High dielectric constants and sudden polarization at the grain boundaries prevent the formation of potential barriers at low temperatures enabling a flow of free electrons.

Above the ferroelectric Curie temperature, dielectric constant and polarization decline so far that there is a rapid increase of the potential barrier heights and it becomes difficult for electrons to pass the potential barrier whence the resistivity of the corresponding material rises dramatically. In a specific temperature range above the Curie temperature T_c, the resistance of the PTC thermistor rises exponentially. Beyond the range of the positive temperature coefficient, the number of free charge carriers is increased by thermal activation. The resistance then decreases and exhibits a negative temperature characteristic (NTC) typical of semiconductors, as shown in figure 10.28.

Figure 10.28c shows that over the majority of the PTC thermistor operating temperature range, it exhibits a slight negative temperature coefficient, similar to most semiconductors. However, as the temperature approaches the switch temperature, T_s, or Curie temperature, the resistance of the element begins to rise rapidly. This steep increase in resistance continues as the temperature rises but eventually levels off and the temperature coefficient becomes negative again at higher temperatures.

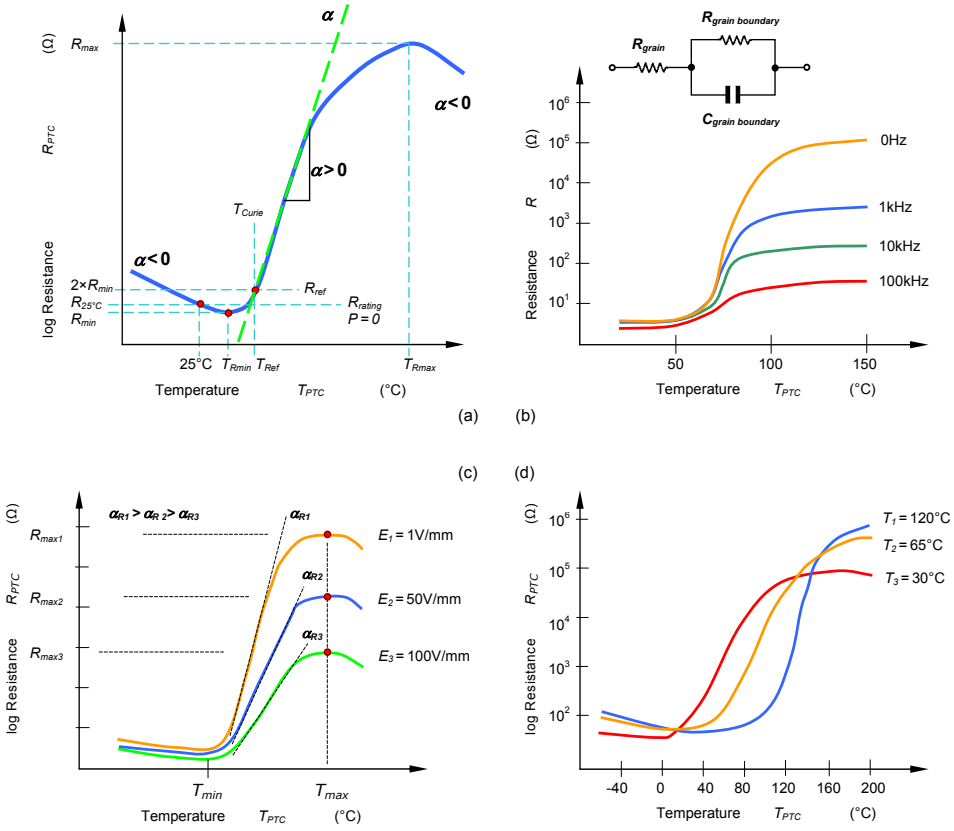


Figure 10.28. *PTC ceramic thermistor: (a) R-T characteristics; (b) influence of frequency on R-T characteristics and ac equivalent circuit; (c) influence of electric field strength E (varistor effect) on R-T characteristics; and (d) reference temperature effect on ceramic R-T characteristics.*

Switch temperature, T_s
The switch temperature of a ceramic PTC is the temperature at which the resistance of the PTC thermistor begins to increase rapidly. The switch temperature is usually defined as the temperature where the resistance of the element is twice the minimum resistance value R_{min}, $T_s = T(2 \times R_{min})$.

Transitional temperature coefficient, α
The temperature coefficient of resistance α is defined as the relative change in resistance referred to the change in temperature and is calculated for each point on the resistance versus temperature curve by:

$$\alpha = \frac{1}{R}\frac{dR}{dT} = \frac{d\,\ell n\,R}{dT} = \ell n 10 \times \frac{d\,\ell g R}{dT} = 2.3 \times \frac{d\,\ell g R}{dT}$$

In the range of the steep rise in resistance above *Point 3* in figure 10.24, R_{ref}, α is approximately constant. The following relation then applies:

$$R_2 \le R_{PTC} \le R_1 \rightarrow \alpha = \frac{\ell n R_2 - \ell n R_1}{T_2 - T_1} = \frac{\ell n \frac{R_2}{R_1}}{T_2 - T_1}$$

Within this temperature range, the reverse relation gives:

$$R_2 = R_1 e^{\alpha (T_2 - T_1)} \qquad (10.65)$$

The value of α for the individual types relates only to the temperature range in the steep region of the resistance curve, which is the region of primary interest for many applications.

Voltage dependence of resistance
Higher voltages applied to the ceramic PTC thermistor drop primarily at the grain boundaries with the result that the high field strengths dominating in these regions break-down the potential barriers, thus producing a lower resistance. The higher the potential barriers, the greater the influence of this 'varistor effect' on resistance. Below the reference temperature, most of the applied voltage is supported across the grain resistance. Thus the field strength at the grain boundaries decreases and the varistor effect is quite weak.
These mechanisms result in the increase of alpha and decreases the pre and post trip resistance as the field strength increases as shown in figure 10.28c.

Frequency dependence of resistance
Due to the structure of the PTC thermistor ceramic material, on ac voltages it is not a pure ohmic resistor. It acts as a capacitive resistor because of the grain boundary junction depletion layers. The impedance measured with ac voltages decreases with increasing frequency, as shown in figure 10.28b. The dc tripped resistance is reduced by a factor of over 50 when the element is used at 1kHz, so use of the PTC is generally restricted to DC and line frequency operation.

Protection circuit operation
Figure 10.29 illustrates the two operating states of a PTC fuse. During rated operation of the load the PTC resistance remains low, operating *Point 1* in figure 10.24. Upon overloading or shorting of the load, however, the power consumption in the PTC thermistor increases so much that it heats up, its resistance increases dramatically, and this reduces the current flow to the load to an admissible low level, operating *Point 4* in figure 10.24. Most of the source voltage V_s is then impresses across the PTC thermistor. Although the current is reduced it is sufficient to maintain the PTC in high-resistance mode, ensuring protection until the cause of the over-current has been removed.

Figure 10.29 illustrates the load-line operating principle of a PTC thermistor designed to operate as a resettable fuse. The region indicated as 'A' represents the normal range of current operation. When current exceeds I_{max}, the device self-heating increases its resistance and causes the circuit to operate in the region indicated by B.
The position of the circuit load-line can be designed such that the over-current protection is either automatically reset or requires a manual reset. In the automatic reset mode, the load line intercepts the *V-I* characteristic at the point F. Stable operation can only occur at this point for normal loads.
In the manual reset mode, the load line intercepts the *V-I* characteristic at three points in figure 10.29; C, D, and E. Point D is unstable so, in practice, stable operation only occurs at points C and E.

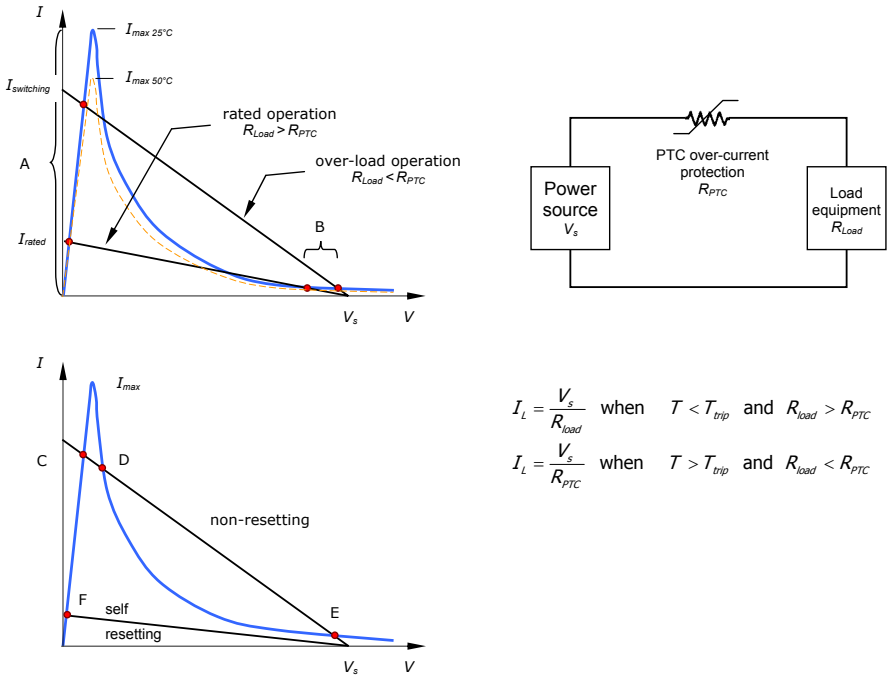


Figure 10.29. *Polymeric PTC thermistor circuit operating load line, showing the operating states of a PTC thermistor for over-current protection.*

PTC device application
Some of the types of applications that utilize the self-heated characteristics of the PTC thermistor include:

- Self-regulating heaters
- Over-current protection
- Liquid level sensing
- Constant current
- Time delay
- Motor starting
- Arc suppression

Generally the device cross-sectional area determines the surge current capability, and the device thickness determines the surge voltage capability. Polymer PTC devices typically have a lower resistance than ceramic PTCs which are stable with respect to voltage and temperature. After experiencing a fault condition, a change in initial resistance occurs with the polymeric PTC.
In balanced systems with a PTC thermistor in each conductor, resistance change may degrade line balance. Including additional series resistance such as a line-feed resistor, LFR, can reduce the effect of the R1 jump. In addition, some PTC thermistors are available in resistance bands to minimize R1 effects. Polymer types are also commonly used singly to protect domestic equipment.
Ceramic PTC devices do not exhibit an R1 jump (because of the reversible ferroelectric Curie temperature mechanism), and their higher resistance avoids the need for installing an LFR. While this reduces component count, the resistance does vary with applied voltage and frequency. Since this change can be substantial (for example, decreasing by a factor of about 3 at 1kV), it is essential that any secondary overvoltage protection be correctly rated to handle the resulting surge current, which can be three times larger than predicted by the nominal resistance of the ceramic PTC. In a typical line application, line balance is critical.

Table 10.2: Characteristics of polymeric and ceramic PTC thermistor fuse devices

material	nominal Ohms	Maximum voltage current trip	resistance stability (with voltage and temperature)	resistance change after surge	typical application
Polymer PTC Thermistor	0.01 - 20		Good	10 - 20%	industrial equipment
Ceramic PTC Thermistor	10 - 50	600V, 13A	*R* decreases with temperature and under impulse	small	balanced line

Example 10.9: ***Resettable ceramic fuse design***

A 24V transformer, operating in an ambient temperature range of 20˚C to 60˚C, is to be PTC thermistor protected under the following conditions:

Normal current = 80mA
Fault current = 300mA

Determine if a 50V, 20Ω ceramic device with the following characteristics, is suitable.

The trip current, the minimum must-switch current, is given by

$$I_{Trip} = \sqrt{\frac{\delta\left(107 - 0.85 \times T_A^{\min}\right)}{0.8 \times R_{25°C}}}$$

The hold current, the maximum no-switch current, is given by

$$I_{Hold} = \sqrt{\frac{\delta\left(93 - 0.85 \times T_A^{\max}\right)}{1.2 \times R_{25°C}}}$$

where:

$\delta = 0.008$ is the dissipation factor
$R_{25°C} = 20\Omega$ is the nominal resistance at 25˚C

Solution

For this application, the requirements are, a PTC element rated for at least 24V, 50/60Hz, can carry 80mA in a 60˚C ambient, and will switch when conducting less than 300mA at 20˚C.

i. The device maximum rated rms voltage must be greater than the application operational voltage:

$$V_{\max} > V_{operational}$$
$$50V\ ac > 28V\ ac$$

ii. The trip current must be less than the fault current, 300mA:

$$I_{Trip} < I_{fault}$$
$$I_{Trip} = \sqrt{\frac{\delta\left(107 - 0.85 \times T_A^{\min}\right)}{0.8 \times R_{25°C}}} = \sqrt{\frac{0.008 \times (107 - 0.85 \times 20°C)}{0.8 \times 20\Omega}}$$
$$= 0.21A < 0.30A$$

iii. The hold current (current without switching) must be greater than the normal operating current, 80mA:

$$I_{Hold} > I_{operational}$$
$$I_{Hold} = \sqrt{\frac{\delta\left(93 - 0.85 \times T_A^{\max}\right)}{1.2 \times R_{25°C}}} = \sqrt{\frac{0.008 \times (93 - 0.85 \times 60°C)}{1.2 \times 20\Omega}}$$
$$= 0.12A > 0.08A$$

The selected PTC fuse is suitable for this transformer protection case.

♣

Traditional Fuses versus PTCs

Fuses and PTC devices are both over-current protection devices, though each offer their own unique operating characteristics and benefits. Understanding the differences between the two technologies should make the choice in selection easier, depending on the application. The most obvious difference

is that PTCs are automatically resettable whereas traditional Fuses need to be replaced after they are tripped. Whereas a fuse will completely stop the flow of current (which may be desired in critical applications) after most similar over-current event, PTCs continue to enable the equipment to function, except in extreme cases.

Because they reset automatically, many circuit designers choose PTCs in instances where over-current events are expected to occur often, and where maintaining low warranty and service costs, constant system uptime, and/ or user transparency are at a premium. They are also often chosen in circuits that are difficult to access in or remote locations, were fuse replacement would be difficult. There are several other operating characteristics to be considered that distinguish PTCs and fuses, and it is also best to test and verify device performance before use within the end application.

- General use PTCs are not rated above 240V while LV fuses are rated up to 600V ac.
- Reviewing product specifications indicates that similarly rated PTCs have about twice (sometimes more) the resistance of fuses.
- The hold (operating) current rating for PTCs can be up to 14A, while the maximum level for fuses can exceed 30A.
- The useful upper limit for a PTC is generally 85ºC, while the maximum operating temperature for fuses is 125ºC. Ambient temperature effects are in addition to the normal derating. PTCs hold and trip rating must be derated when applied at conditions other than room ambient. For example, any rise in ambient temperature will decrease the hold current rating as well as the trip current. A reduction in ambient temperature will increase the trip current as well as the hold current.
- Comparing the time-current curves of PTCs to time-current curves of fuses show that the speed of response for a PTC is similar to the time delay of a Slow-Blow fuse.
- When a PTC is in a 'tripped state' it protects the circuitry by limiting the current flow to a low leakage level. Leakage current can range from less than a hundred milliamps at rated voltage up to a few hundred milliamps at lower voltages. Fuses on the other hand completely interrupt the current flow when tripped, and this open circuit results in no leakage current when subjected to an overload current.
- PTCs are rated for a maximum short circuit current at rated voltage also known as 'breaking capacity' or I_{max}. This fault current level is the maximum current that the device can withstand safely, keeping in mind that the PTC will not actually interrupt the current flow; it has a Leakage Current. A typical PTC short circuit rating is 40A; or for the battery strap PTCs, this value can reach 100A. Fuses do in fact interrupt the current flow in response to the overload and the range of interrupting ratings, vary from tens of amperes up to 10,000A at rated voltage.
- A PCT resettable fuse has better defined characteristics in low voltage dc applications, than traditional fuses, in terms of arcing and resultant circuit voltages with inductive circuits.

10.3.3 Summary of over-current limiting devices

Over-current protection technologies are summarized in Tale 10.3, and as follows:

- PTC thermistors provide self-resetting protection.
- Fuses provide good overload capability and low resistance.
- Heat coils protect against lower level 'sneak currents'.
- LFRs provide the most fundamental level of protection, combined with the precision resistance values needed for balanced lines and are often combined with other devices.

Table 10.3: Summary of over-current limiters

type		performance						
action	Technology	line at post operation	speed	accuracy	resistance stability	operating current	series resistance	current rating
Reducing series	Polymer PTC thermistor	reset	fastest	good	poor	low	medium - low	low
	Ceramic PTC thermistor	reset	fast	good	low	low	high	Low
interrupting series	Fuse	disconnected	slow	fair	good	medium	low	medium - high
	Line feed resistor	both lines disconnected	poor	poor	good	high	high	low
diverting series/shunt	Heat coil	shorted or open	slow	poor	medium	low	medium	low
diverting series	Thermal switch	shorted	poor	poor	good	high	low	high

10.4 Overvoltage

Voltage transients in electrical circuits result from the sudden release of previously stored energy, such as with insulation breakdown arcing, fuses, contactors, freewheeling diode current snap, switches, and transformer energising and de-energising. These induced transients may be repetitive or random impulses. Repetitive voltage spikes are observable but random transients are elusively, unpredictable in time and location. A spike is usually brief but may result in high instantaneous power dissipation. A voltage spike in excess of a semiconductor rating for just a few microseconds usually results in catastrophic device failure. Extensive noise may be injected into low-level control logic causing spurious faults. Generally, high-frequency noise components can be filtered, but low-frequency noise is difficult to attenuate.

Overvoltage devices are placed in parallel with a load or circuitry to be over-voltage protected, to limit the amount of voltage that can appear across the input to a circuit. The overvoltage device appears as a very high-impedance (virtually an open circuit) under normal operating conditions. When an overvoltage event occurs, however, the overvoltage device changes its impedance to divert current through itself, around the protected circuit.

Overvoltage protection devices are designed to protect circuits and additional, they must:

- Not interfere with normal circuit operation.
- Provide maintenance-free operation.
- Reduce long-term cost of the installation by minimizing maintenance time and system downtime.
- Allow the designer to meet industry standards.

Effective transient overvoltage protection requires that the impulse energy be dissipated in the parallel added transient absorption circuit at a voltage low enough to afford circuit survival.

Clamping and Crowbar Devices

Over-voltage protection devices can be classified as either clamping or fold-back (or crowbar). Zener diodes and metal oxide varistors are clamping devices, since they attempt to clamp the voltage at a defined voltage during a stress event. A crowbar device, such as gas discharge tubes and thyristor surge suppressors attempts to create a short circuit when a trigger voltage is reached, with both cases illustrated in Figure 10.30.

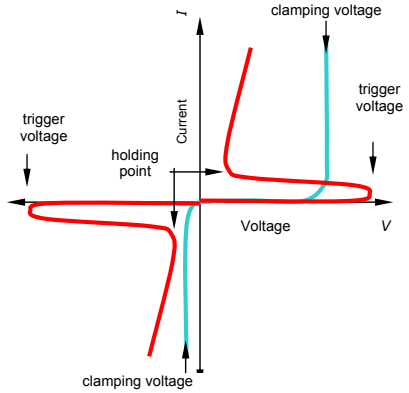


Figure 10.30. *I-V characteristics of a bidirectional crowbar device (black) and a unidirectional clamping device (red).*

Crowbar devices with low on-state voltage can keep voltage levels well below the critical values for sensitive electronic elements and carry considerable current without self-damage due to power dissipation. The lowest current and voltage point that can sustain the on-state of the crowbar device is an important parameter and is often called the holding point, as seen Figure 10.30. If the electrical node being protected can supply the voltage and current levels of the holding point, a crowbar device may not turn off after the electrical stress has been removed. The crowbar device must insure the protection turns off when the electrical stress is removed and does not turn on during normal operation.

Voltage clamp devices do not have the problem of not turning off after a stress. Clamping devices protecting dissipate considerable power, which is dissipated internally. Clamping devices need a low dynamic resistance in the on state to insure that while carrying large currents the voltage does not exceed the allowed levels for the sensitive circuit elements.

Voltage protection can be classified as either *unidirectional or bidirectional*, as shown in figure 10.31. The clamping device has asymmetrical *I-V* characteristics, so is classified as unidirectional, while the crow bar device with symmetrical *I-V* characteristics, performs bidirectional clamping. AC circuits generally require suppression which is symmetrical, that is bidirectional.

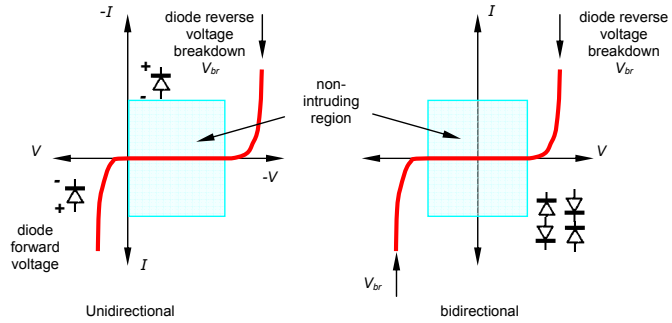


Figure 10.31. *I-V characteristics of: (a) a unidirectional device and (b) a bidirectional device.*

10.4.1 Transient voltage suppression clamping devices

Two voltage transient suppression techniques can be employed.

- **Transient voltage attenuation**
  Low pass filters, such as an *L-C* filter, can be used to attenuate high frequencies and allow the low-frequency power to flow.
- **Diverter (to limit the residual voltage)**
  Voltage clamps such as crowbars or snubbers are usually slow to respond. The crowbar is considered in section 10.2.3 while the snubber, which is for low-energy applications, is considered in sections 8.2 and 8.3.

The voltage-limiting function may be performed by a number of non-linear impedance devices such as reverse selenium rectifiers, avalanche (commonly called Zener) diodes, and varistors made of various materials such as silicon carbide or zinc oxide.

The relationship between the current in the non-linear device, *I*, and the voltage across its terminals, *V*, is typically described by the power law

$$I = kV^{\alpha} \quad \text{(A)} \tag{10.66}$$

k is an element constant dependent on device geometry and material in the case of the varistor, and the non-linear exponent *α* is defined as

$$\alpha = \frac{\log I_2 - \log I_1}{\log V_2 - \log V_1} = \frac{\log I_2 / I_1}{\log V_2 / V_1} \left(= \frac{1}{\log V_2 / V_1} \right) \tag{10.67}$$

where I_1 and I_2 are taken a decade apart, $I_2 / I_1 = 10$. The term alpha (α) represents the degree of non-linearity of the conduction. The higher the value of alpha, the better the clamp and therefore alpha may be used as a figure of merit. Linear resistance has an alpha of 1 and a conductance of $k = 1/R$ ($I = \frac{1}{k} V^{+1}$).

The non-linear voltage-dependent static and dynamic resistances are given by

$$R = \frac{V}{I} = \frac{V}{kV^{\alpha}} = \frac{1}{k} V^{1-\alpha} \quad (\Omega) \tag{10.68}$$

$$R_{dyn} = \frac{dv}{di} = \frac{1}{\alpha k V^{\alpha-1}} = \frac{V}{\alpha I} = \frac{R}{\alpha} \quad (\Omega) \tag{10.69}$$

and the power dissipation is

$$P = VI = V\, kV^{\alpha} = kV^{\alpha+1} \quad \text{(W)} \tag{10.70}$$

The most useful transient suppressors are the Zener diode and the varistor. They are compact devices which offer nanosecond response time and high energy absorption capability.

1 - The Zener diode, usually called a *Transient Voltage Suppressor*, **TVS** in voltage suppression applications, is an effective clamp and comes the closest to being a constant voltage clamp, having an

alpha of 35. Since the avalanche junction area is small and not highly uniform, substantial heating occurs in a small volume. The energy dissipation of the Zener diode is limited, although transient absorption Zener devices with peak instantaneous powers of 50 kW are available. These peak power levels are obtained by:

- Using diffusion technology, which leads to low metallisation contact resistance, narrow base width, and minimises the temperature coefficient.
- Achieving void-free soldering and thermal matching of the chip and the large area electrodes of copper or silver. Molybdenum buffer electrodes are used.
- Using bulk silicon compatible glass passivation which is alkali metal contamination free, and is cut without glass cracking.

Voltage ratings are limited to 280V but devices can be series connected for higher voltage application. This high-voltage clamping function is unipolar and back-to-back series connected Zener diodes can provide high-voltage bipolar symmetrical or asymmetrical voltage clamping.

2 - The varistor (variable res**istor -** voltage-dependant resistance inversely related to voltage) is a ceramic, bipolar, non-linear semiconductor utilising silicon carbide for continuous transient suppression or sintered zinc oxide for intermittent dissipation. Approximately 90 per cent by weight of zinc oxide and suitable additives such as oxides of bismuth, cobalt, manganese and other metal oxides, when pressed, can give varistors with alphas better than 25. The micro-structure of the plate capacitor like body consists of a matrix of highly conductive (and high thermal conductivity) zinc oxide grains separated by highly resistive inter-granular grain boundaries of the additive oxides. Micro-varistors are only produced where the sintered zinc oxide grains meet, providing pn junction semiconductor-type characteristics, as shown in figure 10.32a. The grain sizes vary from approximately 100µm in diameter for low-voltage varistors down to 10µm for high voltage components, producing 30 to 250V/mm (typically 2V to 3V per grain boundary junction). The junctions block conduction at low voltage and provide non-linear electrical characteristics at high voltage. Effectively pn junctions are distributed in parallel and series throughout the structure volume, giving more uniformly distributed heat dissipation than the plane structure Zener diode. The diameter (parallel conduction paths over the area) determines current capability, hence maximum power dissipation, while thickness (number series connected micro-varistors) specifies voltage, as indicted by the I-V characteristics in figure 10.32b. A greater number of adjacent boundaries in series and parallel (that is, the volume of the device) leads to higher energy absorption capability. The structure gives high terminal capacitance values (which decreases with voltage rating according to V -1) depending on area, thickness, and material processing. The varistor may therefore be limited in high-frequency applications (>1kHz), due to $CV^2 f$ related losses. Functionally the varistor is similar to two identical Zener diodes connected back-to-back, in series.

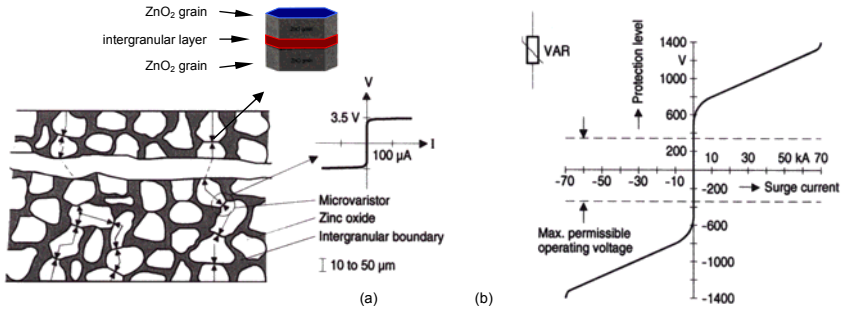


Figure 10.32. *Varistor: (a) conduction mechanisms and (b) I-V linear characteristics.*

Figure 10.33a shows the general equivalent circuit models for the varistor, which consists of the inter-granular boundary resistance R_{IG}, ($\rho \approx 10^{12}$ to 10^{13} Ωcm) the ohmic bulk resistance R_B of the zinc oxide ($\rho \approx 1$ to 10Ωcm), and the non-linear varistor resistance R_{VAR},(0 to ∞Ω).

Leakage current region, $I < 10^{-4}$

Figure 10.33b show the model when the inter-granular boundary resistance R_{IG} dominates the resistance $R_B << R_{IG}$, giving $\alpha = 1$, as shown in figure 10.34a. R_{IG} is temperature (negative) dependant, decreasing with temperature, producing increased leakage current, hence higher steady-state standby losses.

Normal operating region, $I > 10^3$

In figure 10.33c, with $R_{VAR} << R_{IG}$ and $R_B << R_{VAR}$, R_{VAR} dominates electrical behaviour, giving $\alpha > 30$, as shown in figure 10.34a.

High current clamping region, $10^3 > I > 10^{-5}$

In figure 10.33d, the resistance is low as $R_{VAR} << R_{IG}$ and $R_{VAR} < R_B$, giving $\alpha = 1$, as shown in figure 10.34a, with the ohmic bulk resistance R_B of the zinc oxide dominating.

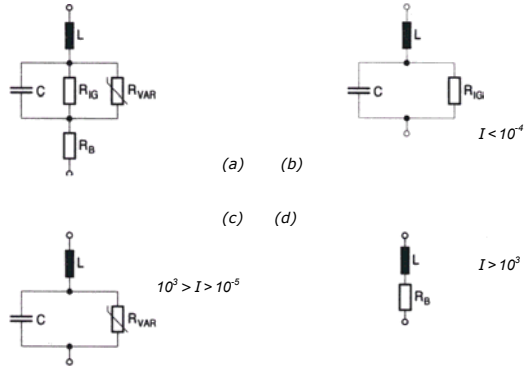


Figure 10.33. *Varistor equivalent circuit models:*
(a) complete model; (b) low current; (c) normal operating region model; and (d)high current model.

The inter-granular capacitance *C,* measured at 1kHz and has a positive temperature coefficient, <0.1%/K, increases with increased thickness (increased voltage rating) and decreases with increase area (increased current/power rating). The capacitance acts as a high pass filter, but restricts the operating frequency limit due to $½CV^2f$ transferred losses.
The lead inductance (≈1nH/mm) *L* limits the element transient response (*L/R*), hence lead length should be minimised.
The varistor voltage is the voltage drop across the element when the current is 1mA, at 25°C.

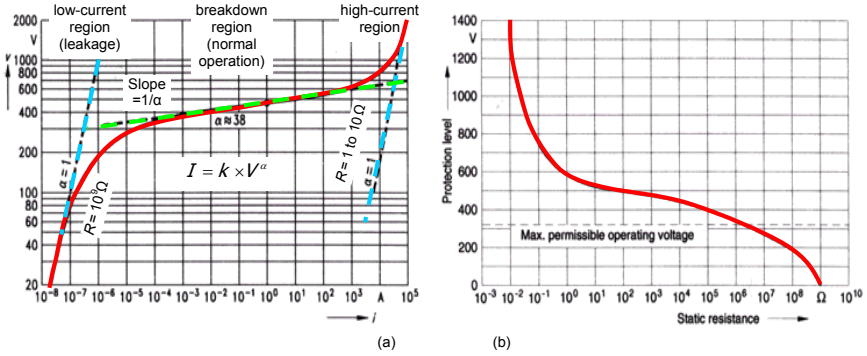


Figure 10.34. *Varistor: (a) I-V linear and (b) static resistance characteristics.*

10.4.1i - Comparison between Zener diodes and varistors

Figure 10.35a illustrates the *I-V* characteristics of various voltage clamping devices suitable for 240 V ac application. The resistor with alpha equal to 1 is shown for reference. It is seen that the higher the exponent alpha, the nearer an ideal constant voltage characteristic is attained, and that the Zener diode performs best on these grounds. When considering device energy absorption and peak current and voltage clamping level capabilities, the Zener diode loses significant ground to the varistor.

The higher the alpha, the lower will be the standby power dissipated. Figure 10.35b shows the dependence of standby power dissipation variation on withstand voltage for various transient absorbers. A small increase in Zener diode withstand voltage produces a very large increase in standby power dissipation. Various device compromises are borne out by the comparison in Table 10.4.
The current, power, and energy ratings of varistors typically are rated values up to 85°C, then linearly derated to zero at a case temperature of 125°C. Voltage-limiting diodes are typically linearly derated from rated values at 75°C to zero at 175°C. Reliability depends on the ambient temperature and applied voltage, and lifetime decreases with increased voltage or temperature. In the case of the varistor, an 8 per cent increase in applied voltage halves the mean time between failures, mtbf, for applied voltages less than 0.71 times the nominal voltage. Below 40°C ambient, the mtbf for a varistor is better than 7 x 10^8 hours (0.7 fit).
The voltage temperature coefficient for the varistor is - 0.05 per cent/K while +0.1 per cent/K is typical for the power Zener.

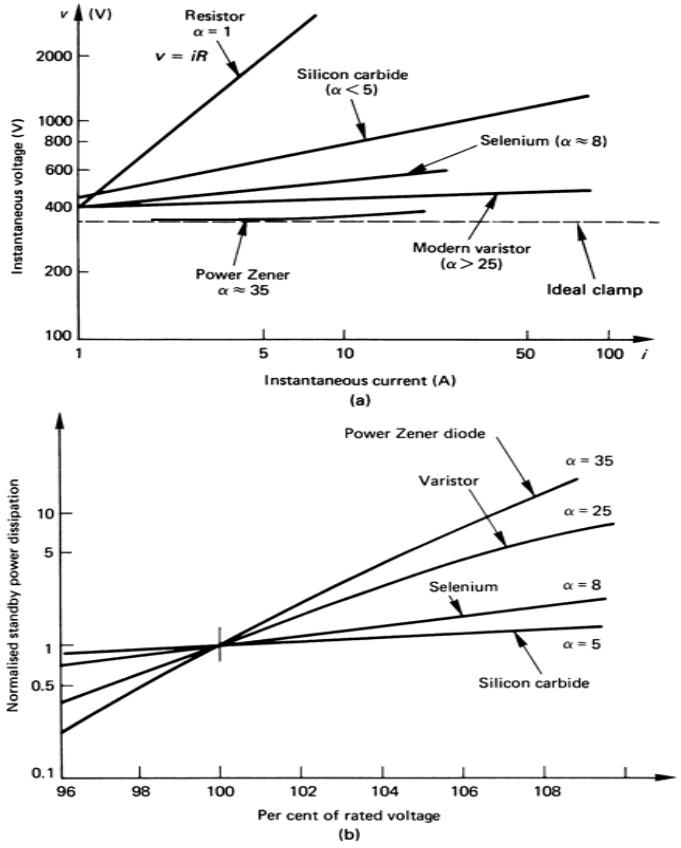


Figure 10.35. *(a) The I-V characteristics of four transient voltage suppressor devices, with resistance characteristics for reference and (b) standby power dissipation characteristics showing the higher the alpha the lower the standby power dissipation.*

The following design points will specify whether a Zener diode or varistor clamp is applicable and the characteristics of the required device.

- Determine the necessary steady-state voltage rating.
- Establish the transient energy to be absorbed by the clamp.

- Calculate the peak transient current through the clamp.
- Determine power dissipation requirements.
- Determine the clamping voltage to which the transient is to be suppressed.
- Estimate the number of fault cycles during the lifetime.

In order to meet higher power ratings, higher voltage levels or intermediate voltage levels, Zener diodes or varistors can be series-connected. The only requirement is that each series device has the same peak current rating. In the case of the varistor this implies the same disc diameter. Then the *I-V* characteristics, energy rating, and maximum clamping voltages are all determined by summing the respective characteristics and ratings of the individual devices.
Parallel operation is difficult and matched *I-V* characteristics are necessary.
A feature of varistors often overlooked is deterioration, which is not applicable to TVS diodes. Figure 10.36a shows that at relatively low energy levels an infinite number of transients can be absorbed, while at rated absorbed energy only one fault is allowed. This single fault, lifetime, is defined as that energy level that causes a 10 per cent increase in clamping voltage level, for a specified current density.

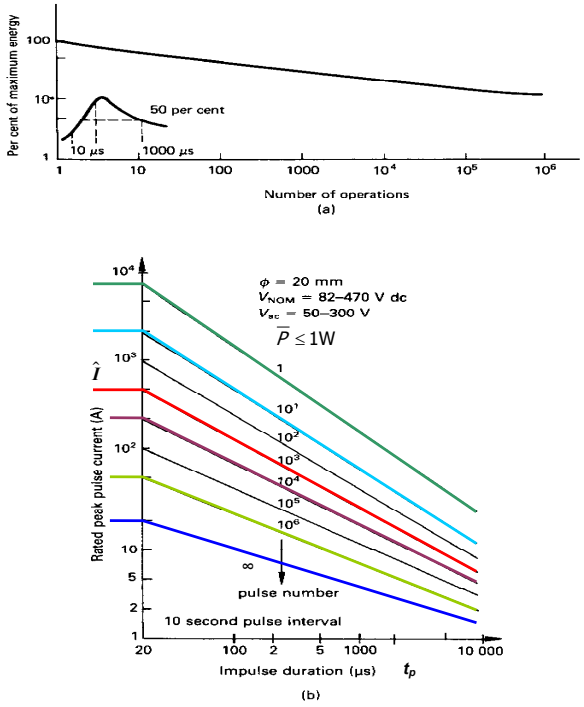


Figure 10.36. *Pulse lifetime ratings for a Zinc oxide varistor: (a) lifetime for fixed 10/100μs pulses and (b) lifetime number for variable-duration square-wave pulses.*

Figure 10.36b shows that high currents can be tolerated for short intervals. The lower the pulse repetition number, the higher the allowable current. The absorbed energy rating is given by

$$W = k\hat{I}\,V_c\,t_p \quad \text{(J)} \tag{10.71}$$

where k = 1 for a rectangular pulse and k = √2 for impulse waveforms: 10μs/1000μs shown in figure 10.36a and 8μs/20μs. The maximum allowable energy pulse is usually based on a 2ms current pulse of magnitude such that a 10% variation in clamping voltage results. Varistors rated for 1000V ac, 1280V dc at 1mA standby, are capable of clamping once, 2870J associated with an 80kA, 2ms pulse, and have a typical capacitance of 2nF at 25°C and 1kHz.
The failure mode of the Zener diode and varistor is a short circuit. Subsequent high current flow may cause an explosion and disintegration of contacts, forming an open circuit. This catastrophic condition can be avoided by fuse protection.

Table 10.4: Comparison of typical transient suppressor characteristics

Suppressor type	Standby current	Peak current at 1ms exp.	Peak power at 1ms	Peak energy	Voltage clamping ratio at 10A	Voltage range	Capacitance at 1MHz
	mA	A	kW	J		dc	nF
Silicon carbide varistor	5	-	-	50	4.6	15-300	-
Selenium	12	30	9	9	2.3	35-700	-
Metal oxide varistor	1	120	40	70	1.7	14-1200	2
Zener diode (5W)	0.005	5.5	1.5	2	1.4	1.8-280	1

Selenium suppressors

Selenium, a naturally occurring substance, has been used as a semiconductor in rectifiers and suppressors. Although its popularity as a rectifier has virtually ceased in favour of its silicon equivalent, demand for selenium suppressors continues.

Depositing the elements on a metal substrate's surface produces selenium cells. This provides the cells with good thermal mass and energy dissipation as well as 'self-healing' characteristics, allowing the device to survive energy discharges in excess of the rated value. Selenium's crystalline structure gives it the ability to continue functioning after a burst of energy in excess of its short pulse width rating. Its suppressor operation is comparable to a pressure relief valve – when the pressure rises, the relief valve opens, releases the pressure, and then resets itself.

Because of its unique properties, the selenium suppressor remains viable in many applications. Its transient voltage clamping characteristic, its ability to continuously dissipate power and handle long surges, make it better than MOVs or silicon suppressors for some applications.

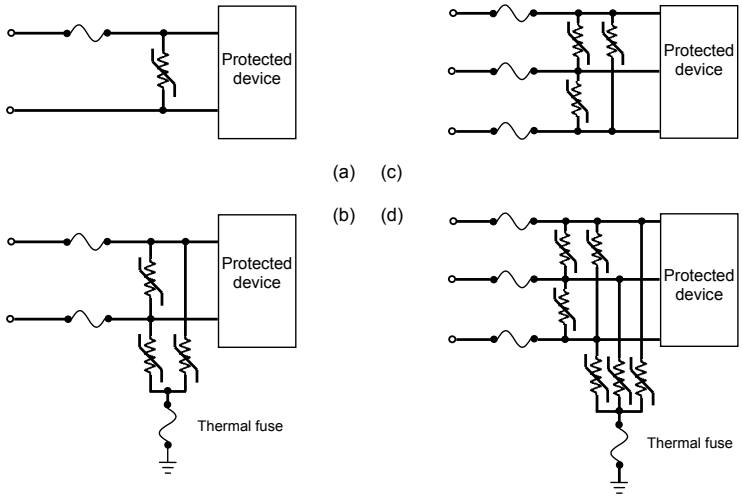


Figure 10.37. *Thermistor circuit protection: (a) ac single-phase or dc circuit protection; (b) ac single-phase circuit with line-line and line-ground protection; (c) ac three-phase circuit line-line protection; and (d) ac three-phase circuit with line-line and thermally fused line-ground protection.*

The selenium suppressor can absorb energy levels in excess of its rated capability while maintaining its clamping characteristics subsequent cycles. The layering of the suppressor onto the aluminium plate allows the suppressor's energy capabilities to follow that of a heat sink thermal curve. This heat sink capability allows steady-state power dissipation up to 40 times that of an MOV. For a 130V suppressor, the selenium suppressor allows steady-state dissipation of 2.5W to 80W, compared with an MOV that allows only 0.1W to 2.5W.

Selenium suppressor cell plates are available in sizes varying from less than 20mm x 20mm to in excess of 30mm x 30mm that can function at a temperature of 0°C to 55°C ambient without any derating. The voltage of a selenium suppressor cell starts at 26Vrms or 22.5Vdc per cell plate, with a 75V maximum due to the dielectric ceiling of the cell. The capacitor plate nature allows placement in series to attain higher voltage levels.

Other suppressors can handle high current, short pulse widths in the microsecond range, but the selenium suppressor can handle milli-second pulse width currents, making it a slower but a more robust suppressor than silicon devices. It has a typical response time of less than 1ms and is capable of handling pulses with long decay times as experienced with the shunt fields of large DC motors or any inductive loads with L/R ratios in the 100ms range, such as with power conditioning systems (that is, from power strips to service entrance)., generators, AC controllers, on the DC side of a rectified generator output, across SCRs on large controllers, and on transformers for line-to-line transient suppression.

Fundamentals of overvoltage protection theory

Electronic equipment and components have been designed to function properly when used within their specified current and voltage ratings. When these ratings are exceeded during operation, the equipment or components may sustain permanent damage and they may cease to operate. Common sources of overvoltage conditions are lightning, ac power contact and power induction. Other electrical components may be susceptible to shifts in system ground potential, increasing the need for overvoltage protection. Voltage protection devices may be installed in parallel with the equipment or components to be protected. In the event of an overvoltage condition, protection devices switch rapidly from a high to a low impedance state, thus clamping the transient voltage across the components to a safe operating level. Under normal operating conditions, the overvoltage device appears as a high impedance device (virtually open circuit, with minimal leakage current) and does not affect normal system operation.

Example 10.10: ***Non-linear voltage clamp***

Evaluate the current of a 1mA @ 250V Zener diode when used to clamp at 340V dc. At 340V dc, calculate the percentage decrease in voltage-dependent resistance and the per unit increase in power dissipation, assuming $\alpha = 30$.

Solution

i. From $I = kV^{\alpha}$, equation (10.66)

$I_2 = I_1(V_2/V_1)^{\alpha} = 1$ mA $(340\text{V}/250\text{V})^{30} = 10.14$A

The Zener diode will conduct 10.14A when clamping at 340V (a 10,140 increase on the standby current of 1mA)

ii. From equation (10.68), $R = V^{1-\alpha}/k$ therefore

$$1-\frac{R_2}{R_1}=1-\left(\frac{V_2}{V_1}\right)^{1-\alpha}=1-\left(\frac{340\text{V}}{250\text{V}}\right)^{-29}=0.99987$$

The percentage decrease in resistance is 99.987 per cent.

The static resistance decreases from (250V / 1mA) 250kΩ to (340V / 10.1A) 33.5Ω.

By differentiating equation (10.66), the incremental resistance (*dv/di*) reduces to 1.12Ω (33.5Ω/30).

iii. $P = kV^{\alpha+1}$ (equation (10.70))

$$\frac{P_2}{P_1}-1=\left(\frac{V_2}{V_1}\right)^{31}-1=\left(\frac{340\text{V}}{250\text{V}}\right)^{31}-1=13793.5$$

The per unit power increase is 13,800.

The power increases from (250V × 1mA) 0.2 W at 250V standby to (340V ×10.14A) 3447.6 W when clamping at 340V dc.

♣

10.4.2 Transient voltage fold-back devices

A fold-back device is normally in a high-resistance state for voltages below the break-over voltage. In this state little current flows through the device. When the voltage exceeds the break-over voltage, the device *folds back* or goes into a low-impedance state, allowing the device to conduct large currents away from sensitive parallel connected electronics. The device will continue to remain in this low impedance state until the current through the device is decreased below its holding current.

Fold-back devices have an advantage over clamping devices because in the fold-back state little voltage appears across the load while the device conducts harmful surges away from the load, whereas clamping devices remain at the clamping voltage. The power dissipated in the fold-back device is therefore much lower than in a clamping device, allowing a much smaller device to be used to conduct the same amount of surge current. In addition to its smaller size and lower power dissipation, a fold-back device offers lower capacitance and cost for a given silicon die area.

10.4.2i The surge arrester

A surge arrestor (or gas discharge tube, GDT) is a high-current, two terminal, hermetically sealed-gas (usually neon and argon) discharge element, as shown in figure 10.38. GDTs apply a short circuit under surge conditions, returning to a high impedance state after the surge. The sealing shields the device from external impregnation, hence ensuring stable gas-physics properties. The internal electrodes are especially electron emission promoted coated for stability and are displaced by about 1mm. The inner cylindrical surface of the insulator is ignition-aid coated to speed-up and stabilise the gas discharge, by distorting the electric field. These features define the electrical characteristics such as spark-over voltage, low capacitance, pulsed and ac discharge current handling capability, as specified in figure 10.39.
Unlike the varistor or Zener diode, the voltage collapses to near zero when the external surge voltage exceeds the device internal electric field strength, eventually creating an low voltage (10V) sustained ionised arc, which is only extinguished when the external energy is reduced to zero, as resulting from a voltage reversal in an ac circuit.

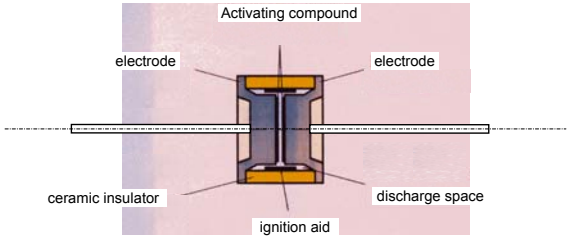


Figure 10.38. *Construction of an inert gas surge arrestor.*

A number of transitional stages occur during surge voltage suppression, as shown in figure 10.39.

- When inactive, the surge arrestor appears as a low capacitance (<1pF) in parallel with a high resistance, typically 1GΩ, where virtually no current flows.
- When the element *spark-over voltage* is reached, V_s (devices ranging between 70V and The operating mechanisms are such that the surge arrestor is not normally suited to dc-circuit operation (or highly inductive ac loads), since to revert to a high impedance mode, the current must drop below the arc discharge mode minimum level of a few 100mA. For this reason, a fail-safe mechanism is incorporated to react the resultant high heating losses that occur with continuous arcing. A spring tension-loaded thermal fuse type mechanism is incorporated to short the two electrodes after melting the separating insulating spacer. Figure 10.40b show the typical short-circuit reaction characteristics as a function of the current flowing through the arrestor.5kV), the voltage rapidly falls to the *glow voltage level* V_{gl}, which is between 70 to 200V with a low current of 10mA, gradually increasing to about 1.5A – region G in figure 10.39.
- As the arrestor current increases, transition to the *arc voltage* V_a mode occurs, where the voltage falls to 10V to 35V, independent of the subsequent current – region A. The transition time between the glow and arc region is dependent on the available current of the impulse, the distance and shape of the electrodes, the gas composition, gas pressure and the proprietary emission coatings.
- As the over-voltage decreases, the arrestor current decreases to a level where the arc cannot be sustained. The arc ceased suddenly, passing briefly through the glow region, and finally extinguishing at the voltage V_e, termed the *extinguishing voltage*.

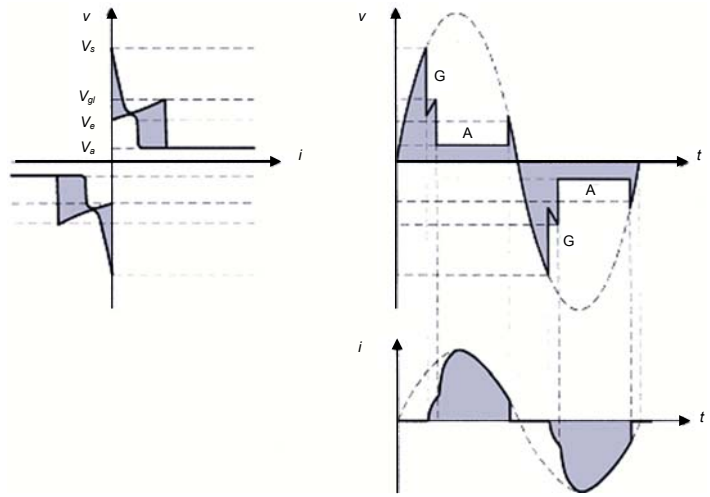


Figure 10.39. *Over-voltage limiting characteristics of an inert gas surge arrestor.*

Response behaviour
The rate of rise of terminal voltage affects the electrical performance, as shown in figure 10.40a. At low *dv/dt*'s (<1V/μs), the dc spark-over voltage V_{sdc} of ignition is determined by the electrode spacing, the gas type and pressure, and the degree of pre-ionization of the noble gas.
At high *dv/dt*'s, the spark-over voltage exceeds the lower steady-state value, V_{sdc}. The ignition-aid coating on the inner cylindrical surface reduces the voltage spread of the resultant impulse spark-over voltage V_{si}. GDTs have no *di/dt* sensitivity.
The operating mechanisms are such that the surge arrestor is not normally suited to dc-circuit operation (or highly inductive ac loads), since to revert to a high impedance mode, the current must drop below the arc discharge mode minimum level of a few 100mA. For this reason, a fail-safe mechanism is incorporated to react the resultant high heating losses that occur with continuous arcing. A spring tension-loaded thermal fuse type mechanism is incorporated to short the two electrodes after melting the separating insulating spacer. Figure 10.40b show the typical short-circuit reaction characteristics as a function of the current flowing through the arrestor.

Switching spark gaps
The gas discharge principle used in the voltage surge arrestor is also applicable to the three-terminal switching spark gap. The device is deliberately ignited, by the build-up of the terminal voltage, (devices from a few hundred volts, up to 6kV) to produce extremely fast (<50ns) high current (>1kA) switching operations (>2M operations), over a very wide temperature range, virtually without loss when conducting and a high insulating resistance (>100MΩ) when non-conducting.

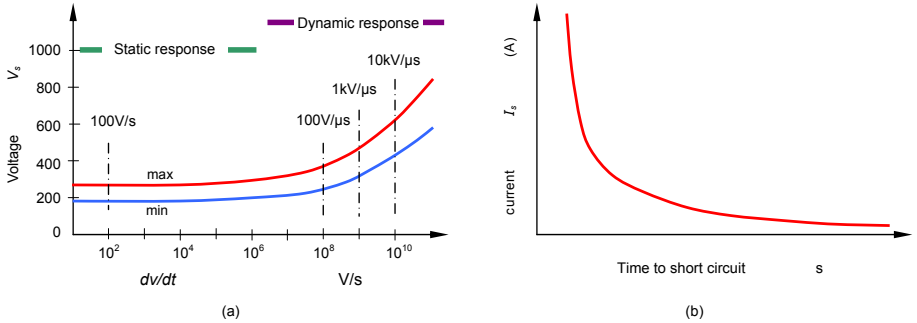


Figure 10.40. *Surge arrestor dynamic characteristics: (a) dv/dt response and (b) fusing time.*

The Dark Effect
The first surge on the GDT tube results in a higher breakdown than subsequent successive surges. As the GDT is normally housed in a plastic module and deployed in a dark cabinet, the term was called the dark effect. The initial strike ionizes the gas to make it settle into a consistent breakdown voltage specification. The impact has been reduced by the design geometry and emission coating composition of the gas tube. The first surge impulse is typically 10% higher than the average impulse let-through voltage. Surge at very high impulse current levels do not experience the phenomenon, which indicates the dark effect is dependent on the surge current and source impedance.

The Spark Effect
The spark effect is due to the arc being of a high enough energy density to cause contaminants (impurities) to be released from the internal materials into the GDT gas atmosphere under a single surge. These contaminants in the gas cause the increase of the dc breakdown voltage by more than 10% between the first two surges. Subsequent surges trigger the “getter” effect of the emission coating that will attract the impurities (contaminants) and reduce the breakdown voltage to the original level. Contaminants suspended in the gas change the gas composition and decrease or increase the breakdown voltage according to the Paschen curve of the particular gas mixture.
Life Cycle of a GDT
The GDT does wear out due to particulates being dislodged from the electrodes during tube arcing. The impact of the arc across the tube is dependent on the energy strike, so the life of the GDT tube is dependent on the impulse applied to it.
The surge ionizing effect charges the tube and therefore attracts the particulates to one end of the tube. This has the effect of changing the electrical properties such as the dc breakdown voltage.
The end of life shorting of the GDT is caused by the rapid breakdown of the emission coating and the electrode material (metal) that further increases internal contaminates. The free materials in the tube attach themselves to the side of the ceramic body between the two electrodes, thereby causing a ‘virtual short’ between the electrodes.

10.4.2ii Thyristor voltage fold-back devices

Thyristor-based devices initially clamp the line voltage, then switch to a low-voltage on-state. After the surge, when the device current drops below its ‘holding current’, the protecting device returns to its original high impedance state.
Figure 10.41 shows the protection action difference between a device that voltage clamps (diode avalanche breakdown action, figure 10.41a), and a device that initially clamps then voltage folds back to a low impedance state (thyristor action, figure 10.41b). The main benefits of thyristor type protection are lower voltage overshoot and an ability to handle moderate currents without a device wear-out or deterioration mechanism. The disadvantages of thyristor protectors are relatively high capacitance, which is a limitation in high-speed digital applications, and low tolerance of excessive current. Thyristor circuit protectors can act either as secondary protection in conjunction with gas discharge tubes, GDTs, or as primary protection for more controlled environments of lower surge magnitudes.

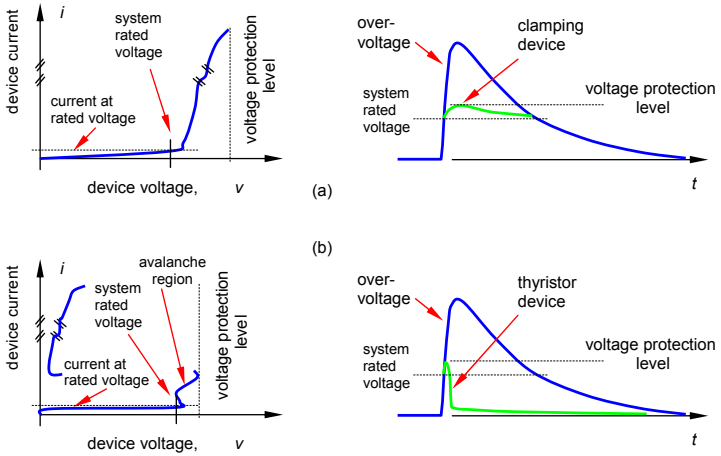


Figure 10.41. *Semiconductor I-V characteristics and switching voltage performance: (a) clamping and (b) fold-back devices.*

Thyristors are multilevel sandwiches of n and p doped silicon which form intertwined bipolar transistors. When the bipolar transistors are triggered they can enter into a self-sustaining low-resistance state. Thyristors, specifically the SCR, is inherently a unidirectional crowbar device. Modifications of the basic SCR have produced a variety of bidirectional and unidirectional options, specifically the triac and diac (diode for alternating current), as shown in figure 10.42.

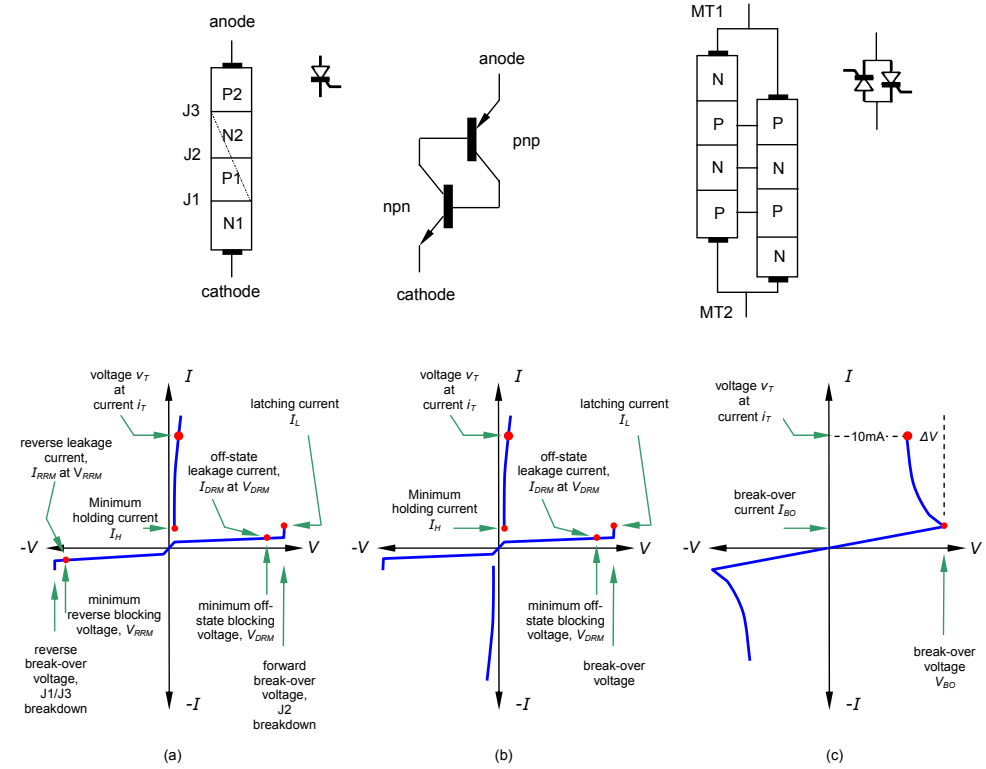


Figure 10.42. *Thyristor physical structure, equivalent circuit and I-V curves for thyristors: (a) an SCR; (b) of a pair of anti-parallel SCRs, the triac; and (c) the diac.*

The protection capability for an SRC is asymmetrical as shown in Figure 10.42a. In the positive direction, turn on of the thyristor results in a dramatic decrease in resistance while in the negative direction the thyristor provides voltage clamping action, similar to a diode based TVS device. For protection in both voltage polarities, to provide symmetrical crowbar behaviour, it is necessary to use two anti parallel SCRs. This can be achieved with a pair of discrete SCRs, or with an integrated structure in a single silicon die that has five doped regions, as illustrated in Figure 10.42. The integrated device is usually called a Thyristor Surge Protection Device (TSPD) and its *I-V* characteristic is shown in Figure 10.43a. The clamping voltage level of fixed voltage thyristors is set during the manufacturing process. Gated thyristors have their protective level set by the voltage applied to the gate terminal.

In response to a transient surge, the thyristor voltage folds back to provide a low-impedance path to ground. The circuit must have enough impedance to limit the fault current below the peak pulse current (I_{PP}) rating of the thyristor. The over-current protector typically does not operate during a lightning pulse. Two voltage triggered fold-back silicon semiconductor devices are commonly used for circuit voltage protection, the thyristor surge protection device, TSPD, and the SIDAC (silicon thyristor (diode, misnomer) device for alternating current). Both are voltage triggered switches but the TSPD is used to reliably protect telecom lines from high current levels and over-voltage occurrences while a SIDAC (a more electrically robust DIAC) is intended for use as a triggering device.

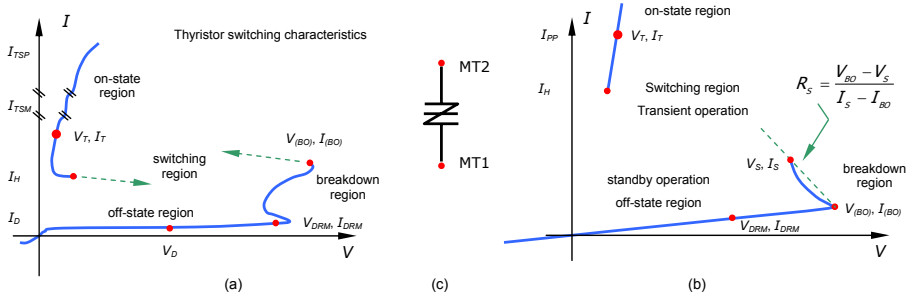


Figure 10.43. *Thyristor fold-back I-V operation: (a) TSPD; (b) SIDAC; and (c) circuit symbol.*

The TSPD

The TSPD is a silicon structure device typically manufactured on an n-type substrate. It is equivalent of two SCR's 'connected' in anti-parallel, which allows the flow of the electric current in both directions. The TSPD is capable of sinking a surge current pulse to ground when transient voltage appears across its two terminals, occurring when the break-over voltage of the device is reached. The device typically operates symmetrically, protecting in the positive and negative direction. The TSPD turns from the off-state to the on-state based on the breakdown and break-over voltage levels that appear between its two terminals, MT1 and MT2. The devices have a current and voltage curve that has a 'fold-back' affect, where the break-over is high, while the clamping voltage is low, basically a short, after the device turns-on giving it high surge abilities. Figure 10.43c shows the symbol for both the TSPD and the SIDAC.

The TSPD is a crowbar device, meaning it has two states of functionality: open circuit and short circuit. It is transparent during normal circuit operation, in that it is an open circuit across its two terminals. Most TSPDs are symmetrical bidirectional designs but there are also unidirectional devices with a built in diode, or asymmetric bidirectional TSPDs are available with a reduced break-over trigger voltage in one polarity.

Typical TSPD surge current capabilities are up to 200A for a 10/1000µs surge voltage. Operating voltages typically cover a broad range, from 12V up through several hundred volts. They have good *dv/dt* sensitivity but poor *di/dt* sensitivity.

The main features of TSPDs are:

Advantages:

- There is no wear–out (aging) mechanism present as with Gas Discharge Tubes and MOVs
- Very fast turn-on switching
- Electrical parameter consistency (V_{BO}, V_{BR}, I_H)
- High immunity to *dv/dt* conditions (>2kV/ms)
- Compared with the MOV, the total energy dissipated is lower, since the crowbar characteristic is not possessed by MOV devices
- Similar current surge capabilities as the GDT
- Short circuit mechanism for protection of the equipment

Disadvantages:

- Very high current surge pulse limitation, where more silicon is needed
- Temperature dependency of the electrical parameters
- Surge performance limited at low temperatures (<-20°C)
- Capacitance is dependent of the die size, but lower than TVS

The SIDAC

The SIDAC is a multi-layer silicon semiconductor usually manufactured on a p-type substrate. Being a bilateral device, it switches from a blocking state to a conducting state when the applied voltage of either polarity exceeds the break-over voltage. As with other trigger devices, the SIDAC switches through a negative resistance region to the low voltage on-state and will remain on until the main terminal current is interrupted or falls below the holding current. When the SIDAC switches to the on state, the voltage across the device drops to less than 3V, depending on magnitude of the main terminal current flow. The main application for the SIDAC is ignition circuits or inexpensive high voltage power supplies.

The difference between a TSPD and a SIDAC is that the SIDAC is intended to be used as a triggering device. The TSPD is intended to withstand surge current levels which involves high levels of peak power, such as required by telecommunication protection standards. Most of the applications for the SIDAC's are related to capacitor discharge circuitry, as part of a RLC circuit; commonly as lamp starters,

strobes and flasher, a stove igniter, etc. The key features of the SIDAC are similar to those of the TSPD. When comparing a similar TSPD with a SIDAC device, the surge current abilities of the TSPD are much larger than the SIDAC. Other key parameters that TSPDs have advantage over SIDACs are lower leakage current (I_{DRM}) and *dv/dt* immunity.

The *I-V* curve in figure 10.43b shows the electrical characteristics of a SIDAC. Typical devices are rated at 1A, 220V with junction operating temperatures up to 125°C. Commutation times are better than 100µs and the switching resistance R_s in figure 10.43b, is typically 100Ω.

10.4.2iii Polymeric voltage variable material technologies

Polymer Electrostatic Discharge (ESD) suppressor devices consist of a polymer embedded with conducting particles as shown in Figure 11.26Xa. At high voltage, arcs between the particles create a low resistance path resulting in a drop in voltage. Additionly, the polymeric suppressors can be manufactured with a gap in an electrode that connects two end terminations. The gap causes the two terminations to be electrically discontinuous (current cannot flow). Into the gap, a polymer-based material is back-filled. This voltage variable material (VVM) has similar electrical characteristics to the zinc-oxide material. Under normal circuit conditions, the VVM acts like an insulator, but when an ESD transient occurs, the VVM transitions to a conductor and shunts the ESD to ground. Polymer devices are bidirectional crowbar devices as shown in Figure 10.44b.

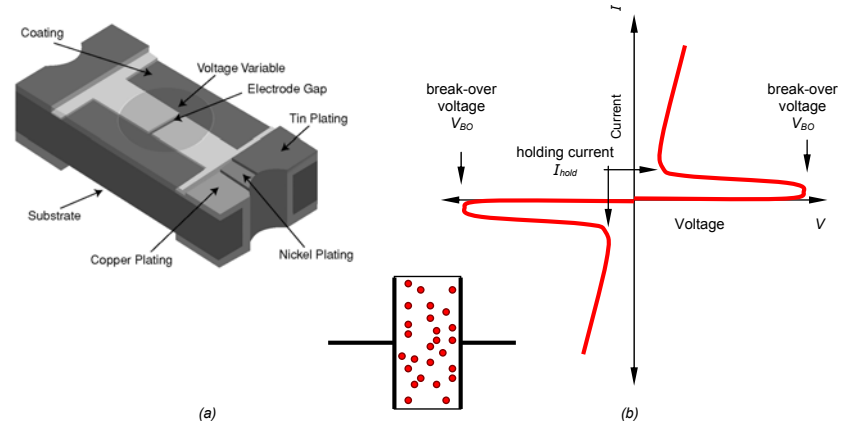


Figure 10.44. *Polymer ESD suppressor: (a) construction and (b) I-V curve of a polymer device.*

Polymer ESD suppressor devices are specifically for electrostatic discharge protection of sensitive low voltage technology. ESD is the transfer of electrical charge between any two objects. ESD is different from other, common overvoltage events (switching and surge transients) in that the time it takes ESD to transition from zero to maximum current and voltage is very short. The rise time of an ESD event is less than a nanosecond, while the other transients take longer than a microsecond to reach their peaks. Since polymeric suppressors are generally specifically designed only for ESD protection, they are not capable of withstanding the higher energy levels of surge transients. On the other hand, polymeric products have the lowest capacitance, 0.050pF, of the suppressor technologies and are used to protect high-speed communication lines.

Polymer devices have high bipolar turn-on voltages, usually over 100V, but turn-on quickly, limiting the exposure to high voltage. The working voltage ranges up to 24V dc, with a leakage current of less than 1nA. The operating temperature range is typically from -65°C to +125°C.

10.4.2iv The crowbar

A *crowbar* can be used for overvoltage and/or over-current protection in both ac and dc circuits. Figure 10.45 illustrates how an SCR can be used to provide fault protection for sensitive dc power electronic circuits and loads. Whenever a fault condition occurs the crowbar SCR is triggered, shorting the supply. The resultant high supply current flowing blows the fuse, or initiates a fast-acting circuit breaker/mcb, thereby isolating the load from the supply. The diode D_c provides a current path for inductive load energy.

The load current is measured by the voltage across the sense resistor *R*. When this voltage reaches a preset limit, that is the load current has reached the fault level, the SCR is triggered. The load or dc link voltage is measured from the resistor divider R_2 - R_3. When this voltage exceeds the pre-determined limit the SCR is triggered and the fuse is blown by the crowbar short-circuit current, isolating the sensitive load from the supply. The load voltage is safely clamped to zero by the conducting SCR or diode D_c.

A judiciously selected crowbar SCR can conduct many times its average current rating. For the few milliseconds in which the fuse is isolating, the SCR I^2t surge current feature can be exploited. The SCR I^2t rating must be larger than the fuse total I^2t rating. If the SCR crowbar is fuse-link protected then the total I^2t of the dc-link fuse link must be less than the pre-arcing I^2t of the SCR crowbar fuse link.

An ac crowbar can comprise two antiparallel-connected SCR's across the fuse-protected ac line, or alternatively one SCR in a four-diode full-wave rectifying bridge.

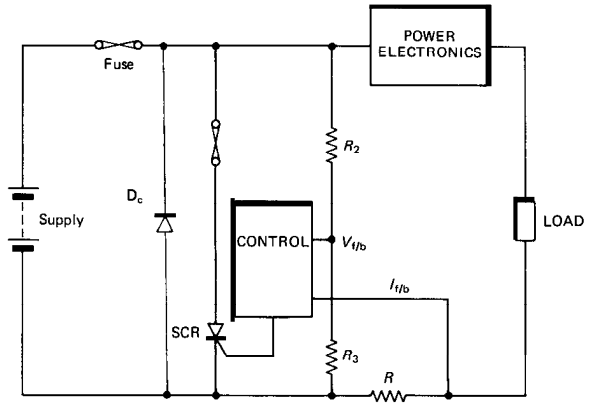


Figure 10.45. *An SCR crowbar for overvoltage and over-current protection.*

10.4.3 Coordination protection

Some primary protective devices such as semiconductor-based devices are fast enough to react in time; however these devices tend to have limited current handling capability. Also, semiconductor devices rated for the primary protection task tend to capacitively load a circuit (due to their physically large size) resulting in bandwidth limitations. Non-semiconductor surge protectors, such as the Gas Discharge Tube (GDT), do not capacitively load circuits and can handle very large currents (tens of kA); however, these devices are slower to react and may not keep the voltages sufficiently low to provide successful protection by themselves. Therefore, conventional protection must be based on a number of stages of such devices. These stages typically start with a GDT as the primary protector for its current handling capability, followed by a semiconductor thyristor protector for speed – the secondary protector.

When GDT and thyristor shunt protectors are used as primary and secondary protectors, the protection coordination between them is complicated in practice. When a surge event occurs, the fast secondary protector will act to limit voltage within the system first (due to its speed). Often this protector will be rated to keep the circuit voltages quite low in order to protect the equipment. Thus, its action can prevent the high energy primary protector (GDT), which requires a higher voltage to operate, from working. In this circumstance, damage is likely to occur to the secondary protector before the GDT operates.

This problem is solved by the complex process of inter-stage coordination. Coordination is the process of placing impedance between the primary and secondary protectors to ensure that sufficient voltage is generated across the primary protector, resulting from current flowing in the secondary protector, to trigger the primary device. Coordination is engineered properly when the primary protector operates after the secondary protector operates, yet before the secondary protector is damaged. The coordinating impedance can be resistive, capacitive, inductive, non-linear or a combination of all of these; proper selection is critical.

Large resistance is the easiest choice to ensure that only a small current in the secondary device causes a significant voltage across the primary device causing it to operate. However, large resistance introduces considerable loss within the data transmission path which is often unacceptable. Capacitance and inductance are also useful, but these impedances are frequency related and so circuits coordinated with such will work only for a band of surge frequencies. Non-linear resistance can also be used to

create different coordinating arrangements based on the duration of the surge - low level surges for example which last a long time causing the coordinating impedance to change to a high resistance state choking further current flow and triggering the primary protector (the basis of operation of the PTC).

10.4.4 Summary of voltage protection devices

There is a variety of devices available to provide shunt electrical over-voltage protection to electronic systems and components. Each device has its own characteristics as outlined in figure 10.46 and Table 10.5.

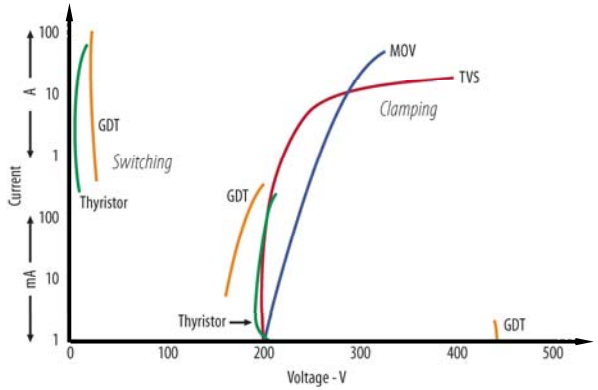


Figure 10.46. *The different I-V characteristics of over-voltage protectors.*

Overvoltage protection technologies may be summarized as follows:

- GDTs offer the best AC power and high surge current capability. For high speed systems, the low capacitance makes GDTs the preferred choice.
- Thyristors provide better impulse protection, but at a lower current.
- MOVs are low cost components.
- TVS offers better performance in low dissipation applications.

Table 10.5: Features of various protection device technologies. All reset to normal after operation

Type	Protection Mechanism	Polarity	Clamp or Crowbar	Speed	Voltage Accuracy	Current Capability	Size / Capacitance	Lowest Trigger Voltage	deterioration
High Power Surge Events – 8/20 µs, 10/1000 µs, etc.									
Gas Discharge Tube GDT	Breakdown of a gas at high voltage	Bidirectional	Crowbar	Slow	Fair	Very high	Large / low	75V	No, depends on severity
Thyristor Surge Protection Devices TSPD	Turn on of coupled bipolar transistors - thyristor	Unidirectional or Bidirectional	Crowbar	Fair	Good	Medium to high	Small / medium	80V	No
Transient Voltage Suppressor TVS	Non-linear resistance of ceramic of zinc oxide grains	Bidirectional	Clamp	Fast	Good	Low	Small / high	NA	No
Metal Oxide Varistor MOV	Forward bias and reverse bias diode conduction	Bidirectional or Unidirectional	Clamp	Fair	Poor	Medium to High	Small / medium	NA	Yes
Very Fast Surge Events – ESD									
Polymer ESD Device PESD	Arcing between particles in polymer	Bidirectional	Crowbar	Fast	Poor	Low	Small / low	~100V	Yes

10.5 Interference

Electromagnetic phenomenon, whether intentional or unintentional by-products, tend to result in undesirable consequences in power electronic circuits and equipment, in terms of generated noise and susceptibility.

- EMC - *Electromagnetic Compatibility*
  The ability of a component or its associated system to operate and function correctly in its intended electromagnetic environment.

- EMI - *Electromagnetic Interference*
  Electromagnetic emissions from a component or its associated system that interfere with the normal operation of another component or system, or the emitting component or system itself.

10.5.1 Noise

RFI noise (electromagnetic interference, EMI) and the resultant equipment interaction is an area of power electronic design that is often fraught, under-estimated or overlooked.
EMI is due to the effects of undesired energy transfer caused by radiated electromagnetic fields or conducted voltages and currents. The interference is produced by a source emitter and is detected by a susceptible victim via a coupling path. The source itself may be a self-inflicted victim. The effects of this interference can vary from simple intermittent reset conditions to a catastrophic failure.
The coupling path may involve one or more of the following four coupling mechanisms.

- *Conduction* - electric current, I
- *Radiation* - electromagnetic field, Z_o
- *Capacitive coupling* - electric field, E
- *Inductive coupling* - magnetic field, H

10.5.1i - Conducted noise is coupled between components through interconnecting wiring such as through power supply (both ac and dc supplies) and ground wiring. This common impedance coupling is caused when currents from two or more circuits flow through the same wiring impedance. Coupling can also result because of common mode and differential (symmetrical) currents, which are illustrated in figure 10.47. Two forms of common mode currents exist. When the conducting currents are equal such that $V_{cm1} = V_{cm2}$, then the common mode currents are termed asymmetrical, while if $V_{cm1} \neq V_{cm2}$, then the currents are termed non-symmetrical.

10.5.1ii - Radiated electromagnetic field coupling can be considered as two cases, namely

- near field, $r \ll \lambda/2\pi$, where radiation due to electric fields, E, and magnetic fields, H, are considered separate
- far field, $r \gg \lambda/2\pi$, where the coupling is treated as a plane wave.

The boundary between the near and far field is given by $r = \lambda/2\pi$ where λ is the noise wavelength and r is distance from the source.
As a reference impedance, the characteristic impedance of free space in the far field Z_o, is given by E/H, which is constant, $\sqrt{\mu_o/\varepsilon_o} = 120\pi = 377\Omega$.

In the ***near field*** region, the r^{-3} (as opposed to r^{-2} and r^{-1}) term dominates field strength.
- A wire currying current produces $E \alpha r^{-3}$ and $H \alpha r^{-2}$,
  - thus the electric field E dominates and the wave impedance $Z > Z_o$.
- A wire loop carrying current produces $H \alpha r^{-3}$ and $E \alpha r^{-2}$,
  - thus the magnetic field H dominates and the wave impedance $Z < Z_o$.

In the near field, interference is dominated by the effective input impedance, Z_{in}, of the susceptible equipment and the source impedance R_s of its input drive.
- Electric coupling increases with increased input impedance, while
- magnetic coupling decreases with increased input impedance.

That is, electric fields, E, are a problem with high input impedance, because the induced current results in a high voltage similar to that given by equation (10.72)

$$v = i_c \times R_s // Z_{in} = C_c \frac{dv}{dt} \times R_s // Z_{in} \qquad (10.72)$$

while magnetic fields, H, are a problem with low input impedance, because the induced voltage results in a high current similar to that given by equation (10.73)

$$i = \frac{v_c}{R_s // Z_{in}} = \frac{M\frac{di}{dt}}{R_s // Z_{in}} \qquad (10.73)$$

In the ***far field*** the r^{-1} term dominates.
In the far field region both the E and H fields are in phase and at right angles. Importantly their magnitudes both decrease, inversely proportionally with distance r, so their magnitude ratio remains constant. That is, in the far field the characteristic impedance $Z_o = E/H = \sqrt{\mu_o/\varepsilon_o} = 120\pi = 377\Omega$ is constant. The far field radiation wave with this constant impedance is termed a *plane wave*. The electric field component of the plane wave tends to dominate interference problems in the far field region.

10.5.1iii - Electric field coupling is caused by changing voltage differences, dv/dt, between conductors. This coupling is usually modelled by capacitance.
The changing electric field produces a current according to $i = C_c\, dv/dt$, where coupling capacitance C_c is dependant on distance of separation, area, and the permittivity of the media. The effect of the produced current is dependant on the source impedance R_s and the effective input impedance, Z_{in}, of the victim equipment as given by equation (10.72).

10.5.1iv - Magnetic field coupling is due to changing currents, di/dt, flowing in conductors. This coupling mechanism is usually modelled by a magnetically coupled circuit, or a transformer, according to $v = M di/dt$, where the resultant current is given by equation (10.73). The mutual inductance M is related to loop area, orientation, separation distance, and screening and its permeability. This induced voltage is independent of any ground connection or electrical connection between the coupled circuits. Magnetic field problems tend to be at low frequencies. Below 100kHz effective screen materials (due to the skin effect) are steel, mu-metal (μ_r = 20,000), and permalloy, while at higher frequencies the good electrical conduction properties of copper and aluminium are more effective despite there much lower permeabilities.

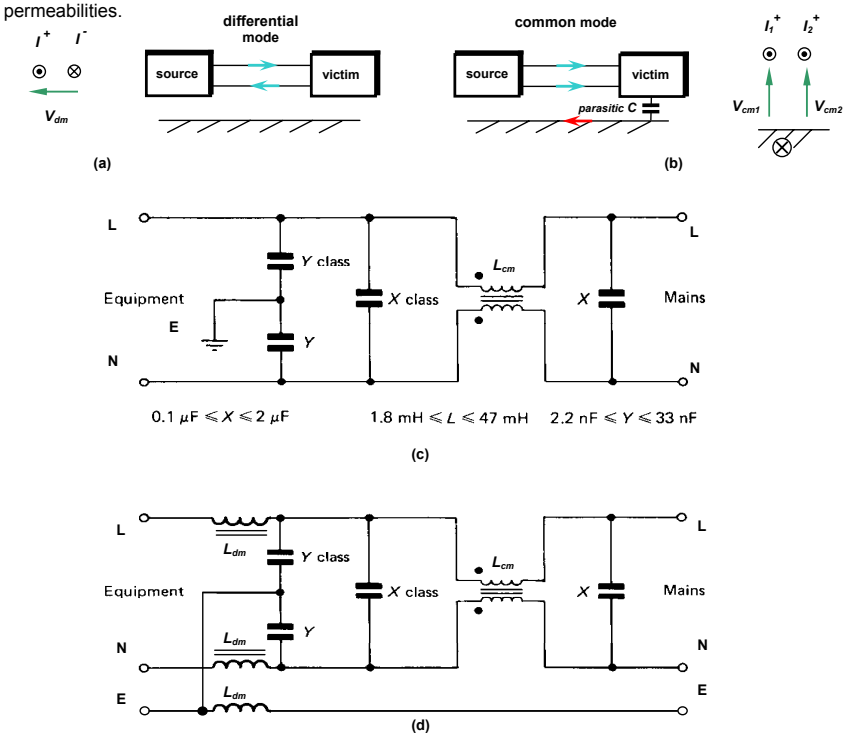


Figure 10.47. *Common mode & differential mode mains supply noise filtering: (a) differential mode noise paths; (b) common mode noise paths; (c) simple L-C mains filter; and (d) high specification mains filter.*

10.5.2 Mains filters

The conducted ac mains borne noise can be attenuated to safe levels by filtering. The simplest type of filter is an inductor in series with the load in order to reduce any current *di/dt* changes. It is usual practice to use *L-C* filtering, which gives second-order attenuation. The typical circuit diagram of an ac mains voltage filter, with common mode noise filtering, is shown in figure 10.47c. The core inductance is only presented to any ampere turn imbalance (common mode current), not the much larger principle throughput (go and return) ac current, hence the core dimensional requirements can be modest. Extra non-coupled inductance is needed for differential mode filtering, as shown in figure 10.47d. Only the higher frequency noise components can be effectively attenuated since the filter must not attenuate the 50/60 Hz ac mains component.

10.5.3 Noise filtering precautions

For power electronics, circuit noise suppression and interaction is ultimately based on a try-it and see approach. Logic and experience do not necessarily prevail. The noise reduction precautions to follow are orientated towards power electronics applications.

Good circuit layout and construction (incorporated at the initial design stage) can greatly reduce the radiated noise, both transmitted and received. Obvious starting points are minimising wire loop lengths, using ground planes, capacitor decoupling, twisted wire pairs, and judicious placement of magnetic components. Use opto-couplers, not only to isolate signals but to allow flexible signal grounding that can bypass ground power noise around sensitive circuitry. Sensitive electronic circuitry should be rfi radiation protected by copper (electric and high frequency magnetic) or mild steel (low frequency magnetic) sheeting, depending on the type of radiation and frequency. Shielding, including electrically isolated heatsinks, should be electrically connected to a point that minimises interference. This may involve connection to supply rails (one of positive, zero, negative) or ground.

An *R-C* snubber across a diode decreases *dv/dt* while a series inductive snubber will limit *di/dt*. Mains ac supply series input inductors for bridge rectifiers (plus diode *R-C* snubbers) decrease the amount of diode recovery noise injected back into the mains and into the equipment. Most effective are common mode transformers in all input and output connection cabling. Although differential mode line inductors may be effective in decoupling input power lines, stability issues can arise when used in output cables. Figure 10.48 outlines the frequency bands where the various interference modes can be expected, and the techniques commonly used to suppression that interference.

In ac circuit applications, zero-voltage turn-on and zero-current turn-off minimise any rapid changes in current, thus reducing radiation. To minimise freewheel diode recovery noise, slow down switch turn-on.

To minimise interactive noise effects, high noise immune circuit designs can be employed which utilise mos technology. The high-voltage input thresholds of cmos logic (4000 series), 74AC (not ACT) logic series, and power MOSFETs and IGBTs (high gate threshold and capacitance), offer circuit noise immunity. Gates with Schmitt trigger (hysteresis) inputs are preferable, for example, 4093, 74AC132, etc. Since noise possesses both magnitude and duration, the much slower response times (along with high input thresholds) of 4000 HEF series cmos may result in better noise immunity in applications requiring clock frequencies below a few megahertz.

DSP core operating voltages below a few volts necessitate: the use of a multilayer pcbs with ground planes, carefully layout separating analogue and digital circuitry, low inductance ceramic chip decoupling, watchdog circuitry, etc. Do not avoid using analogue circuitry (±12V), if it is applicable.

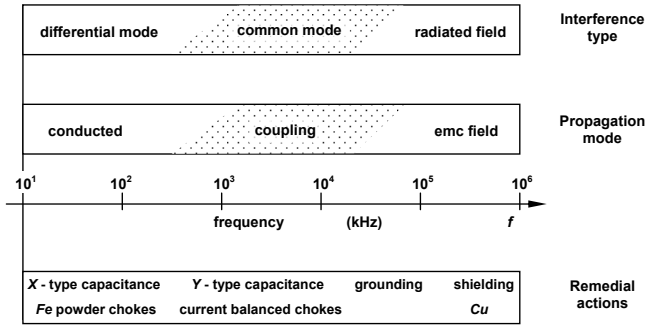


Figure 10.48. *Expected interference types, mode of propagation, and remedial techniques depending on the interference frequency.*

10.6 Earthing

The planet earth is electrically neutral. This means that it has the same number of electrons and protons, so their charges cancel out overall. Thus the Earth has an electric potential of zero. The earth wire of a mains plug is connected to the actual earth, terra firma. Because of the size of the earth it is not possible to charge up anything wired to earth.

This inability to charge equipment connected to the earth is the reason that many systems have their metal boxes wired electrically to the earth. This means that any fault inside the equipment cannot produce a dangerous voltage on its enclosing metal box, so no electric shock is possible from touching the outside of the box even if internally there is an electrical fault.

Figure 10.49. *Three different grounding symbols.*

As shown in figure 10.49, various symbols are used on circuit diagrams to represent earth or ground potential. It is usual assumed that they all mean 'zero volts', that is, the place from which all other voltages in the circuit are referenced or measured. In practice, the meanings of the symbols are slightly different, specifically:

- The earth symbol indicates a place actually wired to terra firma via the mains wires provided or using a wire to a non-corroding metal plate buried in the earth.
- The ground symbol usually indicates a connection back to a place in the power supply, which provides the energy required by the circuit in order to work. It is usually assumed that this place in the power supply is connected to the earth.
- The chassis symbol means a connection to a metal box enclosing the circuit. As far as the circuit is concerned, this metal box is as good a place as the earth for referencing voltages. From the point of view of most electronic circuits, this functions just like an earth connection, however it need not actually be connected to the earth. Hence the chassis of some equipment can potentially be charged up to a high voltage, with respect to the earth.

In most cases, the ground and chassis connections are just indirect paths to earth. However, in some cases, for example, a portable radio using batteries, or the electrics in a car, the ground or chassis represent a sort of 'local' or 'floating' version of the earth used as the zero volts reference point. In most cars, the electrical equipment is powered from a 12V battery. This provides 12V (positive or negative) with respect to the metal bodywork (the chassis). So far as all the car electronics is concerned, it experiences only 12V. However this does not prevent an electric shock when stepping in or out of the car. This is because the chassis may sometimes become charged up to a high voltage with respect to the Earth due to movement of the insulating rubber tyres.

Reading list

General Electric Company, *Transient Voltage Suppression,* 400.3, 1982.

Grafham, D.R. *et al., SCR Manual,* General Electric Company, 6th Edition, 1979.

Williams, T., *EMC for Product Designers*, Newnes, 2nd Edition, 1998.

Fuse manufacturers
Eaton Electrical's Cutler Hammer formerly Westinghouse, Ferraz Shawmut, S&C Electric, Efen, Siba, Bussmann, Littelfuse, Cooper Power Systems, General Electric, and Fusetek.

Problems

10.1. Derive an expression for the worst case maximum allowable voltage-sharing resistance for n series devices each of voltage rating V_D and maximum leakage I_m across a supply V_s. The resistance tolerance is ± 100a per cent and the supply tolerance is ± 100b per cent.
If V_s = 1500 V, V_D = 200 V, I_m = 10 mA, n = 10 and tolerances are ±10 per cent, calculate resistance and maximum total power losses if
i. tolerances are neglected
ii. only one tolerance is considered
iii. both tolerances are included.
[i. R <5.5 kΩ, 63.8 W; ii. R <2.1 kΩ, 185 W; R <3.9 kΩ, 91 W; iii. R <280 Ω, 1234 W].

10.2. Derive a power loss expression for a voltage-sharing resistance network in which both supply and resistance tolerances are included. Assume a dc reverse bias of duty cycle δ.

10.3. Derive the power loss expression for an SCR string with voltage-sharing resistance and an ac supply.

10.4. Two diodes modelled as in figure 2.4a having characteristics approximated in the forward direction by
Diode D_1: $V_F = 1.0 + 0.01\ I_F$ (V)
Diode D_2: $V_F = 0.95 + 0.011\ I_F$ (V)
are connected in parallel. Derive general expressions for the voltage across and the current in each diode if the total current is 200 A.
At what total current and voltage will the diodes equally share?
[102.4 A, 97.6 A, 2.02 V; 100 A, 1.5 V]

10.5. In problem 10.4, what single value of resistance in series with each parallel connected diode match the currents to within 1 per cent of equal sharing? Calculate the resistor maximum power loss.
How will the current share at I_T = 100 A and I_T = 500A with the balancing resistors.
[14.5 mΩ, 148 W; 50 A, 50 A; 254 A, 246 A]

10.6. A Zener diode has an I-V characteristic described by $I = kV^{30}$. What percentage increase in voltage will increase the power dissipation by a factor of 1000?
[25 per cent]

10.7. What is the percentage decrease in the dynamic resistance of the Zener diode in question 10.6?
[99.845 per cent]

10.8. A string of three 2,600 V thyristors connected in series is designed to withstand an off-state voltage of 7.2 kV. If the compensating circuit consists of a series 33 Ω, 0.01 μF snubber in parallel with a 24 kΩ resistor, across each thyristor, and the leakage currents for the thyristors are 20 mA, 25 mA, and 15 mA, at 125°C, calculate the voltage across each thyristor, then the discharge current of each capacitor at turn-on.
[2400 V, 2280 V, 2520 V, 72.73 A, 69.09 A, 76.36 A]

10.9. The reverse leakage current characteristics of two series connected diodes are
Diode D_1: $I_1 = -10^{-4}\ V_1 + 0.14$ (A) for $V_1 < -1400$ V
Diode D_2: $I_2 = -10^{-4}\ V_2 + 0.16$ (A) for $V_2 < -1600$V
If the resistance across diode D_1 is 100 kΩ and $V_{D1} = V_{D2} = -2000$ V, what is the leakage current in each diode and what resistance is required across diode D_2?
[0.34 mA, 0.36 mA, ∞]

10.10. Two high voltage diodes are connected in series as shown in figure 10.5a. The dc input voltage is 5 kV and 10 kΩ dc sharing resistors are used. If the reverse leakage current of each diode is 25mA and 75mA respectively, determine the voltage across each diode and the resistor power loss.
[2750 V, 2250 V, 756.25 W, 506.25 W]

10.11. The forward characteristics of two parallel connected diodes are
Diode D_1: $I_1 = 200\ V_1 - 100$ (A) for $V_1 \geq 0.5$ V
Diode D_2: $I_2 = 200\ V_2 - 200$ (A) for $V_2 \geq 1$V
If the forward voltage of the parallel combination is 1.5V, determine the forward current through each diode.
[200 A, 100 A]

10.12. Two diodes are connected in parallel and with current sharing resistances as shown in figure 10.7. The forward I-V characteristics are as given in problem 10.11. The voltage across the parallel combination is 2V and the balancing resistors are equal in value. Calculate each diode voltage and current. Calculate resistor maximum power loss. Let I_{tot} = 400 A.

11

Naturally Commutating AC to DC Converters

- Uncontrolled Rectifiers

The rectifier converter circuits considered in this chapter have in common an ac voltage supply input and a dc load output. The function of the converter circuit is to convert the ac source energy into fix dc load voltage. Turn-off of converter semiconductor devices is brought about by the ac supply voltage reversal, a process called *line commutation* or *natural commutation.*
Converter circuits employing only diodes are termed *uncontrolled* (or *rectifiers*) while the incorporation of only thyristors results in a (fully) *controlled converter.* The functional difference is that the diode conducts when forward-biased whereas the turn-on of the forward-biased thyristor can be controlled from its gate. An uncontrolled converter provides a fixed output voltage for a given ac supply and load.
Thyristor converters allow an adjustable output voltage by controlling the phase angle at which the forward biased thyristors are turned on. With diodes, converters can only transfer power from the ac source to the dc load, termed rectification and can therefore be described as *unidirectional converters.*
Although rectifiers provide a dc output, they differ in characteristics such as output ripple and mean voltage as well as efficiency and ac supply current harmonics.

An important rectifier characteristic is that of pulse number, which is defined as the repetition rate in the direct output voltage during one complete cycle of the input ac supply.

A useful way to judge the quality of the required dc output, is by the contribution of its superimposed ac harmonics. The harmonic or ripple factor *RF* is defined by

$$RF_v = \frac{V_{ac}}{V_{dc}} = \sqrt{\frac{V_{rms}^2 - V_{dc}^2}{V_{dc}^2}} = \sqrt{\frac{V_{rms}^2}{V_{dc}^2} - 1} = \sqrt{FF^2 - 1}$$

where *FF* is termed the form factor. RF_v is a measure of the voltage harmonics in the output voltage while if currents are used in the equation, RF_i gives a measure of the current harmonics in the output current. Both *FF* and *RF* are applicable to the input and output, and are defined in section 11.6.

The general analysis in this chapter is concerned with single and three phase ac rectifier supplies feeding inductive and resistive dc loads. Purely resistive load equations generally can be derived by setting inductance *L* to zero in the *L-R* load equations.

11.1 Single-phase uncontrolled converter circuits – ac rectifiers

11.1.1 Half-wave circuit with a resistive load, R

The simplest meaningful single-phase half-wave load to analyse is the resistive load. The ac supply *V* is impressed across the load every second ac cycle half period, when load current flows.
The load voltage and current shown in figure 11.1a are defined by

$$v_o(\omega t) = i_o R = \begin{cases} \sqrt{2}V \sin\omega t & 0 \le \omega t \le \pi \\ 0 & \pi \le \omega t \le 2\pi \end{cases} \qquad (11.1)$$

The circuit voltage and current equations can be found by substituting $L = 0$, $\beta = \pi$ and $\varphi = 0$ in the generalised equations (11.18) to (11.20) in section 11.1.3. The average dc output current I_o and voltage V_o are given by

$$V_o = I_o R = \frac{1}{2\pi}\int_0^{\pi} \sqrt{2}V \sin\omega t \, d\omega t = \frac{\sqrt{2}}{\pi}V = 0.45V \qquad (11.2)$$

The rms voltage across the load $V_{o\,rms}$, and rms load current $I_{o\,rms}$, are

$$V_{o\,rms} = \left[\frac{1}{2\pi}\int_0^{\pi} 2V^2 \sin^2\omega t \, d\omega t\right]^{½} = I_{o\,rms}R = \frac{1}{\sqrt{2}}V \qquad (11.3)$$

and the power dissipated in the load, specifically the load resistor, is

$$P_o = I_{o\,rms}^2 R = ½\frac{V^2}{R} \qquad (11.4)$$

The ac current in the load is

$$I_{ac} = \sqrt{I_{o\,rms}^2 - I_o^2} = \frac{V}{R}\left[½ - \frac{2}{\pi^2}\right]^{½} \qquad (11.5)$$

The load voltage harmonics are

$$v_o(\omega t) = \frac{\sqrt{2}V}{\pi} + \frac{\sqrt{2}V}{2}\sin\omega t - \frac{2\sqrt{2}V}{\pi}\left[\frac{1}{1\times 3}\cos 2\omega t + \frac{1}{3\times 5}\cos 4\omega t .. + \frac{1}{n^2-1}\cos n\omega t \ .....\right] \qquad (11.6)$$

for $n = 2, 4, 6, \ldots$
For a resistive load, the load voltage and current ripple factors are both $\sqrt{(½\pi)^2 - 1}$. $FF = ½\pi$. The poor output voltage form factor can be improved by using a capacitor across the output, the load resistor.

11.1.2 Half-wave circuit with a resistive and back emf R-E load

With an opposing emf *E* in series with the resistive load, the load current and voltage waveforms are as shown in figure 11.1b. Load current commences when

$$\omega t = \alpha = \sin^{-1}\frac{E}{\sqrt{2}V} \qquad (11.7)$$

and ceases when

$$\omega t = \pi - \alpha = \pi - \sin^{-1}\frac{E}{\sqrt{2}V} \qquad (11.8)$$

The diode conducts for a period $\theta = \pi - 2\alpha$, during which energy is delivered to both the load resistor *R* and load back emf *E*.

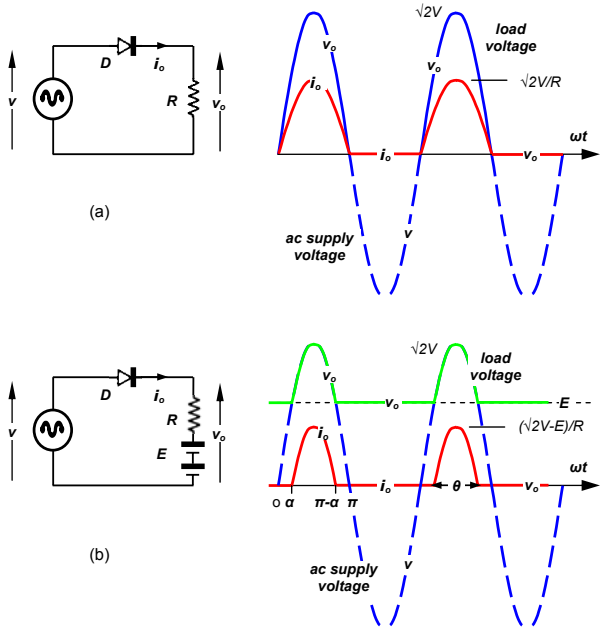


Figure 11.1. *Single-phase half-wave rectifiers:*
(a) purely resistive load, R and (b) resistive load R with back emf, E.

The load average and rms voltages are

$$V_o = \left(½ + \frac{\alpha}{\pi}\right)E + \frac{1}{2\pi}\int_{\alpha}^{\pi-\alpha}\sqrt{2}V\sin\omega t\, d\omega t$$
$$= \left(½ + \frac{\alpha}{\pi}\right)E + \frac{1}{\pi}\sqrt{2}V\cos\alpha \qquad (11.9)$$

$$V_{o\,rms} = \left[E^2\left(½ + \frac{\alpha}{\pi}\right) + V^2\left(½ - \frac{\alpha}{\pi} + \frac{1}{2\pi}\sin 2\alpha\right)^2\right]^{½} \qquad (11.10)$$

The load average and rms currents are

$$I_o = \frac{1}{R}\left[\frac{\sqrt{2}V}{\pi}\cos\alpha - E\left(½ - \frac{\alpha}{\pi}\right)\right] = \frac{1}{R}\left[\frac{\sqrt{2}V}{\pi}\sin ½\theta - E\frac{\theta}{2\pi}\right] \qquad (11.11)$$

$$I_{o\,rms} = \frac{1}{R}\left[\frac{V^2}{2\pi}\sin\theta - \frac{2\sqrt{2}}{\pi}V\,E\sin ½\theta + \left(V^2 + E^2\right)\frac{\theta}{2\pi}\right]^{½} \qquad (11.12)$$

The total power delivered to the *R-E* load is

$$P_o = P_R + P_E = I_{o\,rms}^2 R + E\,I_o \qquad (11.13)$$

Example 11.1: ***Half-wave rectifier with resistive and back emf load***

A dc motor has series armature resistance of 10Ω and is fed via a half-wave rectifier, from the single-phase 230V 50Hz ac mains. Calculate

i. The rectifier diode peak current
ii. The motor average starting current

If at full speed, the motor back emf is 100V dc, calculate

iii. The average and rms motor voltages and currents
iv. The motor electrical losses
v. The power converted to rotational energy
vi. The supply power factor and motor efficiency
vii. Diode approximate loss if modelled by $v_D = 0.8 + 0.025\times i_D$.

Solution

Worst case conditions are at standstill when the motor back emf is zero and the circuit and waveforms in figure 11.1a are applicable.

i. The peak supply and peak load voltage is √2×V = √2×230 = 325.3V.
The peak diode and load current is

$$\hat{i}_D = \hat{i}_o = \frac{\hat{v}_o}{R} = \frac{325.3\text{V}}{10\Omega} = 32.5\text{A}$$

ii. The motor average current, at starting, is given by equation (11.2)

$$V_o = I_o R = 0.45\times 230\text{V} = 103.5\text{V}$$
$$I_o = \frac{V_o}{R} = \frac{103.5\text{V}}{10\Omega} = 10.35\text{A}$$

With a 100V back emf, the circuit and waveforms in figure 11.1b are applicable.
The current starts conducting when

$$\omega t = \alpha = \sin^{-1}\frac{E}{\sqrt{2}V} = \sin^{-1}\frac{100\text{V}}{\sqrt{2}\times 230\text{V}} = 17.9°$$

The current conducts for a period $\theta = \pi - 2\alpha = 180° - 2\times 17.9 = 144.2°$, ceasing at $\omega t = \pi - \alpha = 162.1°$.

iii. The average and rms load currents and voltages are given by equations (11.9) to (11.12).

$$V_o = \left(½ + \frac{\alpha}{\pi}\right)E + \frac{1}{\pi}\sqrt{2}V\cos\alpha$$
$$= \left(½ + \frac{17.9°}{180°}\right)\times 100\text{V} + \frac{1}{\pi}\sqrt{2}\times 230\text{V}\times\cos 17.9° = 158.5\text{V}$$

$$I_o = \frac{1}{R}\left[\frac{\sqrt{2}V}{\pi}\sin ½\theta - E\frac{\theta}{2\pi}\right]$$
$$= \frac{1}{10\Omega}\left[\frac{\sqrt{2}\times 230\text{V}}{\pi}\sin ½\times 144.2° - 100\text{V}\times\frac{144.2°}{360°}\right] = 5.85\text{A}$$

$$V_{o\,rms} = \left[E^2\left(½ + \frac{\alpha}{\pi}\right) + V^2\left(½ - \frac{\alpha}{\pi} + \frac{1}{2\pi}\sin 2\alpha\right)^2\right]^{½}$$
$$= \left[100^2\left(½ + \frac{17.9°}{180°}\right) + 230^2\left(½ - \frac{17.9°}{180°} + \frac{1}{2\pi}\sin 2\times 17.9°\right)^2\right]^{½} = 179.2\text{V}$$

$$I_{o\,rms} = \frac{1}{R}\left[\frac{V^2}{2\pi}\sin\theta - \frac{2\sqrt{2}}{\pi}V\,E\sin ½\theta + \left(V^2 + E^2\right)\frac{\theta}{2\pi}\right]^{½}$$
$$= \frac{1}{10\Omega}\left[\frac{230^2}{2\pi}\sin 144.2° - \frac{2\sqrt{2}}{\pi}\times 230\text{V}\times 100\text{V}\times\sin ½\times 144.2° + \left(230^2 + 100^2\right)\frac{144.2°}{360°}\right]^{½}$$
$$= 10.2\text{A}$$

iv. The motor loss is the loss in the 10Ω resistance in the dc motor equivalent circuit

$$P_R = I_{o\,rms}^2 R = 10.2^2\times 10\Omega = 1041.5\text{W}$$

v. The back emf represents the source of electrical energy converted to mechanical energy

$$P_E = E\times I_o = 100\text{V}\times 5.85\text{A} = 585\text{W}$$

vi. The supply power factor is defined as the ratio of the supply power delivered, *P*, to apparent supply power, *S*

$$pf = \frac{P}{S} = \frac{P_R + P_E}{V\times I_{o\,rms}} = \frac{1041.5\text{W} + 585\text{W}}{230\text{V}\times 10.2\text{A}} = 0.69$$

The motor efficiency is

$$\eta = \frac{P_E}{P_R + P_E} = \frac{585\text{W}}{1041.5\text{W} + 585\text{W}}\times 100 = 40.0\%$$

vii. By assuming the diode voltage drop is insignificant in magnitude compared to the 230V ac supply, then the currents and voltages previously calculated involve minimal error. The rectifying diode power loss is

$$P_D = 0.8\times I_o + 0.025\Omega\times I_{o\,rms}^2$$
$$= 0.8\times 5.85\text{A} + 0.025\Omega\times 10.2^2 = 7.3\text{W}$$

♣

11.1.3 Single-phase half-wave rectifier circuit with an R-L load

A single-phase half-wave diode rectifying circuit with an R-L load is shown in figure 11.2a, while various circuit electrical waveforms are shown in figure 11.2b. Load current commences when the supply voltage goes positive at $\omega t = 0$. It will be seen that load current flows not only during the positive half of the ac supply voltage, $0 \le \omega t \le \pi$, but also during a portion of the negative supply voltage, $\pi \le \omega t \le \beta$. The load inductor stored energy maintains the load current and the inductor's terminal voltage reverses and is able to overcome the negative supply and keep the diode forward-biased and conducting. This current continues until all the inductor energy, $½Li^2$, is released ($i = 0$) at the *current extinction angle* (or *cut-off angle*), $\omega t = \beta$.

During diode conduction the circuit is defined by the Kirchhoff voltage equation

$$L\frac{di}{dt} + Ri = v_R + v_L = v = \sqrt{2}\,V\,\sin\,\omega t \qquad (\text{V}) \qquad (11.14)$$

where *V* is the rms ac supply voltage. Solving equation (11.14) yields the load (and diode) current

$$i(\omega t) = \frac{\sqrt{2}\,V}{Z}\left\{\sin\,(\omega t - \phi) + \sin\phi\; e^{-\omega t/\tan\phi}\right\} \qquad (\text{A}) \qquad (11.15)$$
$$0 \le \omega t \le \beta \ge \pi \qquad (\text{rad})$$

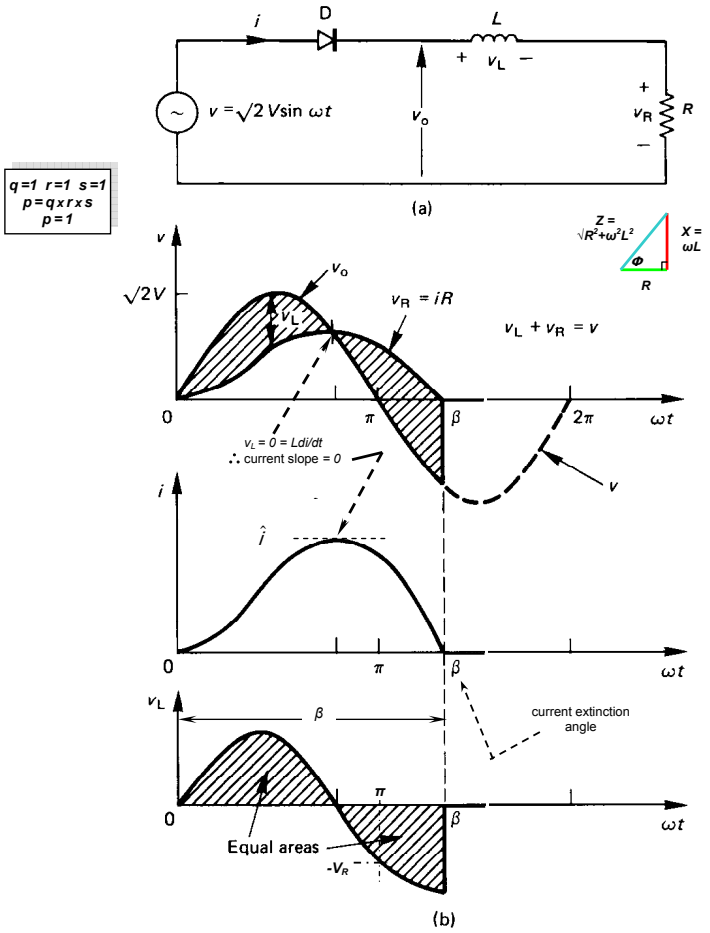


Figure 11.2. *Half-wave rectifier with an R-L load:*
(a) circuit diagram and (b) waveforms, illustrating the equal area and zero current slope criteria.

where $Z = \sqrt{(R^2 + \omega^2 L^2)}$ (ohms)

$$\tan\phi = \omega L / R = Q \text{ and } R = Z\cos\phi$$

$$i(\omega t) = 0 \quad \text{(A)} \qquad \beta \le \omega t \le 2\pi \quad \text{(rad)} \tag{11.16}$$

The current extinction angle β is determined solely by the load impedance Z and can be solved from equation (11.15) when the current, $i = 0$ with $\omega t = \beta$, such that $\beta > \pi$, that is

$$\sin(\beta - \phi) + \sin\phi\, e^{-\beta/\tan\phi} = 0 \tag{11.17}$$

This is a transcendental equation which can be solved by iterative techniques. Figure 11.3a can be used to determine the extinction angle β, given any load impedance (power factor) angle $\phi = \tan^{-1}\omega L/R$.

The mean value of the rectified current, the output current, $\bar{I}_o$, is given by integration of equation (11.15)

$$\bar{I}_o = \frac{1}{2\pi}\int_0^{\beta} i(\omega t)\, d\omega t \quad \text{(A)}$$

$$\bar{I}_o = \frac{\sqrt{2}V}{2\pi R}(1-\cos\beta) \quad \text{(A)} \tag{11.18}$$

while the mean output voltage V_o is given by

$$V_o = \frac{1}{2\pi}\int_0^{\beta}\sqrt{2}\,V\sin\omega t\; d\omega t = \bar{I}_o R = \frac{\sqrt{2}V}{2\pi}(1-\cos\beta) \quad \text{(V)} \tag{11.19}$$

Since the mean voltage across the load inductance is zero, $V_o = \bar{I}_o R$ (see the equal area criterion below). Figure 11.3b shows the normalised output voltage V_o/V as a function of $\omega L/R$.

The rms output (load) voltage and current are given by

$$V_{rms} = \left[\tfrac{1}{2\pi}\int_0^{\beta}\left(\sqrt{2}V\right)^2\sin^2\omega t\, d\omega t\right]^{\frac{1}{2}} = V\left[\tfrac{1}{2\pi}\{\beta - \tfrac{1}{2}\sin 2\beta\}\right]^{\frac{1}{2}}$$

$$i_{rms} = \frac{V\cos\phi}{R}\left[\frac{1}{2\pi}\left\{\beta - \frac{\sin\beta\cos(\beta+\phi)}{\cos\phi}\right\}\right]^{\frac{1}{2}} = \frac{V}{Z}\left[\frac{1}{2\pi}\left\{\beta - \frac{\sin\beta\cos(\beta+\phi)}{\cos\phi}\right\}\right]^{\frac{1}{2}} \tag{11.20}$$

From equations (11.19) and (11.20) the harmonic content in the output voltage is indicated by the voltage form factor.

$$FF_v = \frac{V_{rms}}{V_o} = \frac{\left[\pi\{\beta - \tfrac{1}{2}\sin 2\beta\}\right]^{\frac{1}{2}}}{1-\cos\beta} \tag{11.21}$$

For a resistive load, when $\beta = \pi$, the form factor reduces to a value of 1.57. The ripple factor is therefore $\sqrt{FF_v^2 - 1} = 1.21$. For a purely resistive load the voltage and current form factors are equal.

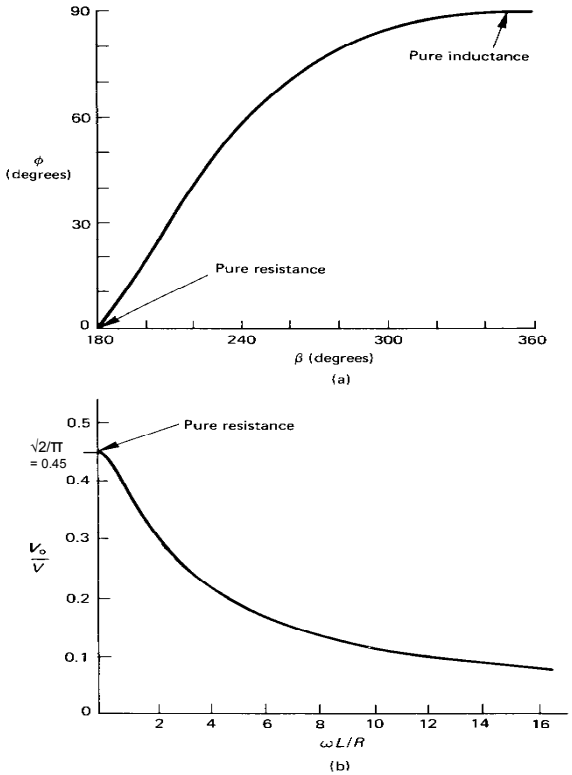


Figure 11.3. *Single-phase half-wave converter characteristics: (a) load impedance angle ø versus current extinction angle β and (b) variation in normalised mean output voltage V_o/V versus ωL/R.*

The power delivered to the load, which is the power delivered to the load resistance R, is

$$P_L = I_{rms}^2 R \qquad (11.22)$$

The supply power factor, using the rms current in equation (11.20), is

$$pf = \frac{\text{power, } P_L}{\text{apparent power}} = \frac{I_{rms}^2 R}{I_{rms} V} = \frac{I_{rms} R}{V} = \frac{V_{R\,rms}}{V} = \left[\frac{1}{2\pi}\left\{\beta - \frac{\sin\beta\ \cos(\beta+\phi)}{\cos\phi}\right\}\right]^{1/2} \cos\phi = \mu \times \cos\phi \qquad (11.23)$$

The characteristics for an R-L-E load can be determined by using $\alpha = 0$ in the case of the half-wave controlled converter in section 12.2.1iii.
For a purely inductive load, L, $\beta = 2\pi$ is substituted into the appropriate equations. The average output voltage tends to zero and the current is given by

$$i(\omega t) = \frac{\sqrt{2}V}{\omega L}\{1 - \cos\omega t\} \qquad \text{(A)}$$

which has a mean current value of $\sqrt{2}\,V/\omega L$.

11.1.3i – Inductor equal voltage area criterion

The average output voltage V_o, given by equation (11.19), is based on the fact that the average voltage across the load inductance, in steady state, is zero. The inductor voltage is given by

$$v_L = L\,di/dt \qquad \text{(V)}$$

which for the circuit in figure 11.2a can be expressed as

$$\int_0^{\beta/\omega} v_L(t)\,dt = \int_{i_o}^{i_\beta} L\,di = L(i_\beta - i_0) \qquad (11.24)$$

If the load current is in steady state then $i_\beta = i_o$, which is zero here, and in general

$$\int v_L\,dt = 0 \qquad \text{(Vs)} \qquad (11.25)$$

The inductor voltage waveform for the circuit in figure 11.2a is shown in the last plot in figure 11.2b. The inductor equal voltage area criterion implies that the shaded positive area must equal the shaded negative area, in order to satisfy equation (11.25). The net inductor energy at the end of the cycle is zero (specifically, unchanged since $i_o = i_\beta$), that is, the energy into the inductor equals the energy transferred from the inductor. This area aspect is a useful aid in predicting and drawing the load current waveform.
It is useful to superimpose the supply voltage v, the load voltage v_o, and the resistor voltage v_R waveforms on the same time axis, ωt. The load resistor voltage, $v_R = Ri$, is directly related to the load current, i. The inductor voltage v_L will be the difference between the load voltage and the resistor voltage, and this bounded net area must be zero. Thus the average output voltage is $V_o = I_o R$. The equal voltage areas associated with the load inductance are shown shaded in two plots in figure 11.2b.

11.1.3ii - Load current zero slope criterion

The load inductance voltage polarity changes from positive to negative as energy initially transferred into the inductor, is released. The stored energy in the inductor allows current to continue to flow after the input ac voltage has reversed. At the instant when the inductor voltage reverses, its terminal voltage is zero, and

$$v_L = L\,di/dt = 0$$
$$\text{that is} \quad di/dt = 0 \qquad (11.26)$$

The current slope changes from positive to negative, whence the voltage across the load resistance ceases to increase and starts to decrease, as shown in figure 11.2b. That is, the Ri waveform crosses the supply voltage waveform with zero slope, whence when the inductor voltage is zero, the current begins to decrease. The fact that the resistor voltage slope is zero when $v_L = 0$, aids prediction and sketching of the various circuit waveforms in figure 11.2b, and subsequent waveforms in this chapter.

11.1.4x Half-wave rectifier circuit with a R load and capacitor filter

The output voltage ripple factor of a half-wave rectifier with a resistive can be improved by adding decoupling capacitance across the load output, as shown in figure 11.4X.

In the period $\alpha \le \omega t \le \beta$

$$v_o(t) = \sqrt{2}V_s \sin\omega t$$

and

$$i_s = i_C + i_o = C\frac{dv_o}{dt} + \frac{v_o}{R_L}$$

Solving for the source current i_s

$$i_s(t) = \frac{\sqrt{2}V_s}{R_L}\left[\omega R_L C\cos\omega t + \sin\omega t\right]$$
$$= \frac{\sqrt{2}V_s}{R_L}\sqrt{1+(\omega R_L C)^2}\cos(\omega t - \theta) \qquad (11.27)$$

where $\theta = \tan^{-1}\dfrac{1}{\omega R_L C}$

Since $i_s = 0$ when $\omega t = \beta$, $\beta - \theta = \frac{1}{2}\pi$. That is

$$\beta = \tfrac{1}{2}\pi + \theta = \tfrac{1}{2}\pi + \tan^{-1}\frac{1}{\omega R_L C}$$

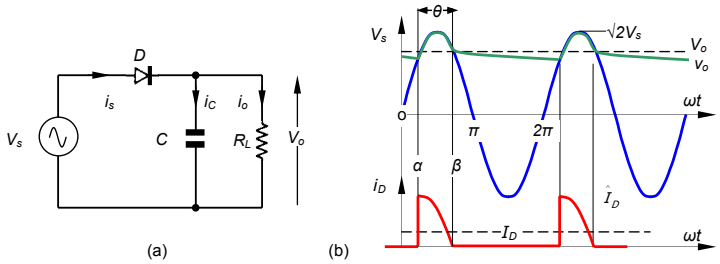


Figure 11.4x. *Single-phase half-wave rectifier:*
(a) circuit with a capacitively filtered resistive load and (b) waveforms.

For the period $\beta \le \omega t \le 2\pi + \alpha$

$$i_s = 0$$

and

$$C\frac{dv_o}{dt} + \frac{v_o}{R_L} = 0 \quad \text{where} \quad v_o(\omega t = \beta) = \sqrt{2}V_s\cos\theta$$

Therefore

$$v_o(t) = \sqrt{2}V_s e^{-(\omega t - \beta)\tan\theta}\cos\theta \qquad (11.28)$$

Since $v_o = \sqrt{2}V_s \sin\alpha$ when $\omega t = 2\pi + \alpha$

$$\sin\alpha = e^{-\alpha\tan\theta}e^{-(\frac{3}{2}\pi - \theta)\tan\theta}\cos\theta$$

where α can be solved iteratively.

The peak to peak ripple voltage is $2V_s(1-\sin\alpha)$, which decreases as C increases for which $\alpha \to \frac{1}{2}\pi$.
The full-wave rectified case is considered in section 11.1.7iv, where the period boundary $\omega t = \pi + \alpha$ is used.

Example 11.2: ***Half-wave rectifier with source resistance***

In the dc supply half-wave rectifier circuit of figure 11.4, the source voltage is $230\sqrt{2}\sin(2\pi\ 50t)$ V with an internal resistance $R_i = 1$ Ohm, $R_L = 10$ Ohms, and the filter capacitor C is very large. Calculate

i. the mean value of the load voltage, V_o
ii. the diode average and peak currents, I_D, $\hat{I}_D$
iii. the capacitor peak charging and discharging currents
iv. the diode reverse blocking voltage, V_{DR}

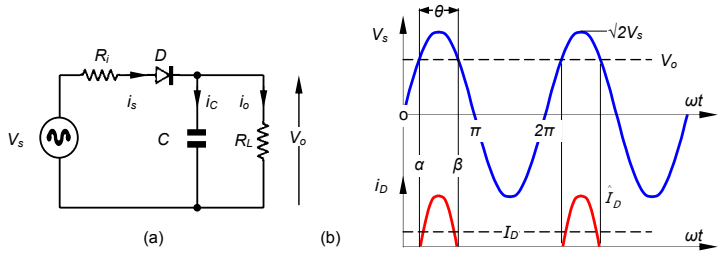


Figure 11.4. *Single-phase half-wave rectifier: (a) circuit with a resistive load and (b) waveforms.*

Solution

i. Because the load filter capacitor is large, it is assumed that the dc output voltage is ripple free and constant. The capacitor provides the load current when the ac supply level is less that the dc output. The load current and peak diode (hence supply) current are therefore

$$I_o = \frac{V_o}{R_L} \qquad \hat{I}_D = \frac{\sqrt{2}V_s - V_o}{R_i}$$

The ac supply provides current, through the rectifying diode, during the period

$$i_s = \frac{1}{R_i}\left(\sqrt{2}V_s \sin\omega t - V_o\right) \qquad \alpha \le \omega t \le \beta$$

If the capacitor voltage is to be maintained constant, the charge into the capacitor must equal the charge delivered by the capacitor when the rectifying diode is not conducting, that is

$$\int_{\alpha}^{\beta}\left(i_s - i_o\right)d\omega t = \int_{\beta}^{\alpha+2\pi} i_o\, d\omega t$$

Also

$$V_o = \sqrt{2}V_s \sin\alpha$$

$$\alpha = \frac{\pi - \theta}{2} \qquad \beta = \frac{\pi + \theta}{2}$$

Manipulation yields

$$\tan ½\theta - ½\theta = \pi\frac{R_i}{R_L} = \pi\frac{1\Omega}{10\Omega} = 0.1\pi$$

An iterative solution yields θ = 99.6º, that is, the diode conducts for a period of 5.53ms (10ms×99.6º/180º), every cycle of the ac supply, 20ms. The capacitor, hence output voltage, is

$$V_o = \sqrt{2}V_s \sin\alpha = \sqrt{2}V_s \sin\frac{\pi - \theta}{2}$$

$$= \sqrt{2}\times 230\text{V}\times\sin\frac{180° - 99.6°}{2} = 209.95\text{V}$$

ii. The average diode current is given by

$$\bar{I}_D = \frac{1}{2\pi}\int_{\alpha}^{\beta}\frac{1}{R_i}\left(\sqrt{2}V_s \sin\omega t - V_o\right)d\omega t = \frac{1}{2\pi R_i}\left(\sqrt{2}V_s \times 2\times cos\frac{\pi - \theta}{2} - V_o \times \theta\right)$$

$$= \frac{1}{2\pi 1\Omega_i}\left(\sqrt{2}\times 230\text{V}\times 2\times cos\frac{180° - 99.6°}{2} - 209.95\text{V}\times\pi\times\frac{99.6°}{180°}\right) = 21.0\text{A}$$

Alternatively, as would be expected, the average diode current is the average load current:

$$\bar{I}_D = I_o = \frac{V_o}{R_L} = \frac{209.95\text{V}}{10\Omega} = 21.0\text{A}$$

The peak diode current is

$$\hat{I}_D = \frac{\sqrt{2}V_s - V_o}{R_i} = \frac{\sqrt{2}\times 230\text{V} - 210\text{V}}{1\Omega} = 115.3\text{A}$$

iii. The capacitor peak charging current is the difference between the peak diode current and the load current, viz., 115A - 21A = 94A, while the peak discharging current is the average load current of 21A.

iv. The diode reverse voltage is the difference between the instantaneous supply voltage and the output voltage 210V. This is a maximum at the negative peak of the ac supply, when the diode voltage is √2×230V + 210V = 535.3V. During any period when the load is disrupted, the output capacitor can charge up to √2×230V, hence the diode can experience, worse case, 2×√2×230V = 650.5V.

♣

11.1.4 Half-wave circuit with an R-L load and freewheel diode

The circuit in figure 11.2a, which has an *R-L* load, is characterised by discontinuous current (*i* = 0) and high ripple current. Continuous load current can result when a diode D_f is added across the load as shown in figure 11.5a. This freewheel diode prevents the voltage across the load from reversing during the negative half-cycle of the ac supply voltage. The inductor energy is not returned to the ac supply, rather is retained in the load circuit. The stored energy in the inductor cannot reduce to zero instantaneously, so the current is forced to find an alternative path whilst decreasing towards zero. When the rectifier diode D_1 ceases to conduct at zero volts it blocks, and diode D_f provides an alternative load current freewheeling path, as indicated by the waveforms in figure 11.5b.

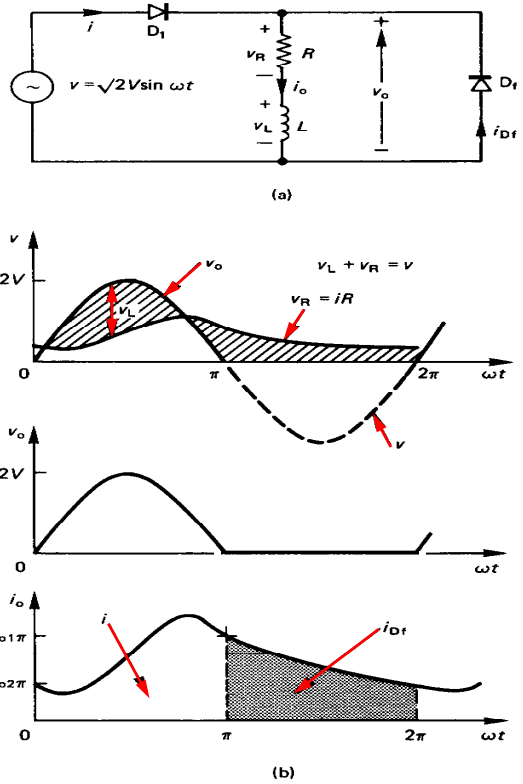


Figure 11.5. *Half-wave rectifier with a load freewheel diode and an R-L load: (a) circuit diagram and parameters and (b) circuit waveforms.*

The output voltage is the positive half of the sinusoidal input voltage. The mean output voltage (thence mean output current) is

$$V_o = \bar{I}_o R = \frac{1}{2\pi}\int_0^{\pi}\sqrt{2}V\sin\omega t\, d\omega t$$

$$V_o = \frac{\sqrt{2}V}{\pi} = 0.45\times V = \bar{I}_o R \qquad \text{(V)} \qquad (11.29)$$

The rms value of the load circuit voltage v_o is given by

$$V_{rms} = \sqrt{\frac{1}{2\pi}\int_0^{\pi}\left(\sqrt{2}V\sin\omega t\right)^2 d\omega t} = V/\sqrt{2} = 0.71\times V \quad \text{(V)} \tag{11.30}$$

The output ripple (ac) voltage is defined as

$$V_{Ri} \triangleq \sqrt{V_{rms}^2 - V_o^2} = \sqrt{\left(\sqrt{2}V/2\right)^2 - \left(\sqrt{2}V/\pi\right)^2} = V\sqrt{1/2 - 2/\pi^2} = 0.545\times V \tag{11.31}$$

hence the load voltage form and ripple factors are defined as

$$FF_v = V_{rms}/V_o = \pi/2 = 1.57$$
$$RF_v \triangleq V_{Ri}/V_o = \sqrt{\left(\frac{V_{rms}}{V_o}\right)^2 - 1} = \sqrt{FF_v^2 - 1} = \sqrt{\pi^2/4 - 1} = 1.211 \tag{11.32}$$

After a large number of ac supply cycles, steady-state load current conditions are established, and from Kirchhoff's voltage law, the load current is defined by

$$L\frac{di}{dt} + Ri = \sqrt{2}V\sin\omega t \quad \text{(A)} \qquad 0 \le \omega t \le \pi \tag{11.33}$$

and when the freewheel diode conducts

$$L\frac{di}{dt} + Ri = 0 \quad \text{(A)} \qquad \pi \le \omega t \le 2\pi \tag{11.34}$$

During the period $0 \le \omega t \le \pi$, when the freewheel diode current is given by $i_{Df} = 0$, the supply current, which is the load current, are given by

$$i(\omega t) = i_o(\omega t) = \frac{\sqrt{2}V}{Z}\sin(\omega t - \phi) + (I_{o2\pi} + \frac{\sqrt{2}V}{Z}\sin\phi)e^{-\omega t/\tan\phi} \quad \text{(A)} \qquad 0 \le \omega t \le \pi \tag{11.35}$$

for

$$I_{o2\pi} = \frac{\sqrt{2}V}{Z}\sin\phi\,\frac{1+e^{-\pi/\tan\phi}}{e^{\pi/\tan\phi} - e^{-\pi/\tan\phi}} \quad \text{(A)}$$

where $Z = \sqrt{R^2 + (\omega L)^2}$ (ohms)

$\tan\phi = \omega L/R$

During the period $\pi \le \omega t \le 2\pi$, when the supply current $i = 0$, the freewheel diode current and hence load current are given by

$$i_o(\omega t) = i_{Df}(\omega t) = I_{o1\pi}e^{-(\omega t - \pi)/\tan\phi} \quad \text{(A)} \qquad \pi \le \omega t \le 2\pi \tag{11.36}$$

for

$$I_{o1\pi} = I_{o2\pi}e^{\pi/\tan\phi} \quad \text{(A)}$$

For discontinuous load current (the freewheel diode current i_{Df} falls to zero before the rectifying diode D_1 recommences conduction), the appropriate integration gives the average diode currents as

$$\bar{I}_{D1} = \frac{V}{\sqrt{2}\pi R}\left(2 - \left(1 + e^{-\pi/\tan\phi}\right)\times\sin^2\phi\right)$$
$$\bar{I}_{Df} = \bar{I}_o - \bar{I}_{D1} = \frac{V}{\sqrt{2}\pi R}\left(1 + e^{-\pi/\tan\phi}\right)\times\sin^2\phi \tag{11.37}$$

In figure 11.5b it will be seen that although the load current can be continuous, the supply current is discontinuous and therefore has a high harmonic content.

The output voltage Fourier series ($V_o + V_1 + V_{n=2,4,6..}$) is (see equation (11.6))

$$v_o(t) = \frac{\sqrt{2}V}{\pi} + \frac{\sqrt{2}V}{2}\sin\omega t - \frac{\sqrt{2}V}{\pi}\sum_{n=2,4,6}^{\infty}\frac{2}{\left(n^2-1\right)}\cos n\omega t \tag{11.38}$$

Dividing each harmonic output voltage component by the corresponding load impedance at that frequency gives the harmonic output current, whence rms current. That is

$$I_n = \frac{V_n}{Z_n} = \frac{V_n}{|R + jn\omega L|} = \frac{V_n}{\sqrt{R^2 + (n\omega L)^2}} \tag{11.39}$$

and

$$I_{rms} = \sqrt{I_o^2 + \sum_{n=1,2,4,6..} 1/2\, I_n^2} \tag{11.40}$$

Example 11.3: ***Half-wave rectifier – with load freewheel diode***

In the circuit of figure 11.5, the source voltage is $240\sqrt{2}\sin(2\pi\ 50t)$ V, R = 10 ohms, and L = 50 mH. Calculate

i. the mean and rms values of the load voltage, V_o and V_{rms}
ii. the mean value of the load current, $\bar{I}_o$
iii. the current boundary conditions, namely $I_{o1\pi}$ and $I_{o2\pi}$
iv. the average freewheel diode current, hence average rectifier diode current
v. the rms load current, hence load power and supply rms current
vi. the supply power factor

If the freewheel diode is removed from across the load, determine

vii. an expression for the current hence the current extinction angle
viii. the average load voltage hence average load current
ix. the rms load voltage and current
x. the power delivered to the load and supply power factor

From the rms and average output voltages and currents, determine the load form and ripple factors.

Solution

i. From equation (11.29), the mean output voltage is given by

$$V_o = \frac{\sqrt{2}V}{\pi} = \frac{\sqrt{2}\times 240\text{V}}{\pi} = 108\text{V}$$

From equation (11.30) the load rms voltage is

$$V_{rms} = V/\sqrt{2} = 240\text{V}/\sqrt{2} = 169.7\text{V}$$

ii. The mean output current, equation (11.29), is

$$\bar{I}_o = \frac{V_o}{R} = \frac{\sqrt{2}V}{\pi R} = \frac{\sqrt{2}\times 240\text{V}}{\pi\times 10\Omega} = 10.8\text{A}$$

iii. The load impedance is characterised by

$$Z = \sqrt{R^2 + (\omega L)^2} = \sqrt{10^2 + (2\pi\times 50\text{Hz}\times 0.05)^2} = 18.62\ \Omega$$
$$\tan\phi = \omega L/R = 2\pi\times 50\text{Hz}\times 0.05\text{H}/10\Omega = 1.57 \text{ or } \phi = 57.5° \equiv 1\text{rad}$$

From section 11.1.4, equation (11.35)

$$I_{o2\pi} = \frac{\sqrt{2}V}{Z}\sin\phi\,\frac{1+e^{-\pi/\tan\phi}}{e^{\pi/\tan\phi} - e^{-\pi/\tan\phi}}$$
$$I_{o2\pi} = \frac{\sqrt{2}\times 240\text{V}}{18.62\Omega}\times\sin(\tan^{-1}1.57)\times\frac{1+e^{-\pi/1.57}}{e^{\pi/1.57} - e^{-\pi/1.57}} = 3.41\text{A}$$

Hence, from equation (11.36)

$$I_{o1\pi} = I_{o2\pi}e^{\pi/\tan\phi} = 3.41\times e^{\pi/1.57} = 25.22\text{A}$$

Since $I_{o2\pi} = 3.41\text{A} > 0$, continuous load current flows.

iv. Integration of the diode current given in equation (11.36) yields the average freewheel diode current.

$$\bar{I}_{Df} = \frac{1}{2\pi}\int_0^{\pi} i_{Df}(\omega t)\,d\omega t = \frac{1}{2\pi}\int_0^{\pi} I_{o1\pi}e^{-\omega t/\tan\phi}\,d\omega t$$
$$= \frac{1}{2\pi}\int_0^{\pi} 25.22\text{A}\times e^{-\omega t/1.57\text{rad}}\,d\omega t = \frac{25.22\text{A}}{2\pi}\times 1.57\text{rad}\times\left[1 - e^{-\frac{\pi}{1.57}}\right] = 5.46\text{A}$$

The average input current, which is the rectifying diode mean current, is given by

$$\bar{I}_s = \bar{I}_{D1} = \bar{I}_o - \bar{I}_{Df} = 10.8\text{A} - 5.46\text{A} = 5.34\text{A}$$

v. The load voltage harmonics given by equation (11.38) can be used to evaluate the load current at the load impedance for that frequency harmonic.

$$v(t) = \frac{\sqrt{2}V}{\pi} + \frac{\sqrt{2}V}{2}\sin\omega t - \sum_{n=2,4,6}^{\infty}\frac{2\sqrt{2}V}{\left(n^2-1\right)\pi}\cos\left(n\omega t\right)$$

The following table shows the calculations for each frequency component.

harmonic *n*	$V_n = \frac{2\sqrt{2}V}{(n^2-1)\pi}$ (V)	$Z_n = \sqrt{R^2+(n\omega L)^2}$ (Ω)	$I_n = \frac{V_n}{Z_n}$ (A)	$\frac{1}{2}I_n^2$
0	***(108.04)****	10.00	10.80	***(116.72)***
1	***(169.71)****	18.62	9.11	41.53
2	72.03	32.97	2.18	2.39
4	14.41	63.62	0.23	0.03
6	6.17	94.78	0.07	0.00
8	3.43	126.06	0.03	0.00
see equation (11.38) for first two terms			$I_o^2 + \sum \frac{1}{2}I_n^2 =$	160.67

The rms load current is

$$I_{rms} = \sqrt{I_o^2 + \sum_{n=1,2,4..} \tfrac{1}{2}I_n^2} = \sqrt{160.7} = 12.68\text{A}$$

The power dissipated in the load resistance is therefore

$$P_{10\Omega} = I_{rms}^2 R = 12.68\text{A}^2 \times 10\Omega = 1606.7\text{W}$$

The freewheel diode rms current is

$$I_{Df} = \sqrt{\frac{1}{2\pi}\int_0^{\pi}\left(I_{o1\pi}e^{-(\omega t)/\tan\phi}\right)^2 d\,\omega t}$$

$$= \sqrt{\frac{1}{2\pi}\int_0^{\pi}\left(25.22\text{A}\times e^{-(\omega t)/1.57\text{rad}}\right)^2 d\,\omega t} = 8.83\text{A}$$

Thus the input (and rectifying diode) rms current is given by

$$I_{D1_{rms}} = I_{s_{rms}} = \sqrt{I_{rms}^2 - I_{Df\,rms}^2}$$

$$= \sqrt{12.68^2 - 8.83^2} = 9.09\text{A}$$

vi. The input ac supply power factor is

$$pf = \frac{P_{out}}{V_{rms}I_{rms}} = \frac{1606.7\text{W}}{240\text{V}\times 9.09\text{A}} = 0.74$$

vii. If the freewheel diode D_f is removed, the current is given by equation (11.15), that is

$$i(\omega t) = \frac{\sqrt{2}V}{Z}\left\{\sin(\omega t - \phi) + \sin\phi\ e^{-\omega t/\tan\phi}\right\}$$

$$= \frac{\sqrt{2}\times 240\text{V}}{18.62\Omega}\left\{\sin(\omega t - 1.0) + 0.841\times e^{-\omega t/1.57}\right\}$$

$$= 18.23\times\left\{\sin(\omega t - 1.0) + 0.841\times e^{-\omega t/1.57}\right\} \quad \text{(A)} \qquad 0 \le \omega t \le \beta \quad \text{(rad)}$$

The current extinction angle β is found by setting $i = 0$ and solving iteratively for β. Figure 11.3a gives an initial estimate of 240° (4.19 rad) when ϕ = 57.5° (1 rad). That is

$$0 = \sin(\beta - 1.0) + 0.841\times e^{-\beta/1.57}$$

gives β = 4.08 rad or 233.8°, after iteration.

viii. The average load voltage from equation (11.19) is

$$V_o = \frac{\sqrt{2}V}{2\pi}(1-\cos\beta) = \frac{\sqrt{2}\times 240\text{V}}{2\pi}(1-\cos 4.08) = 86.0\text{V}$$

The average load current is

$$\bar{I}_o = V_o / R = 86.0\text{V}/10\Omega = 8.60\text{A}$$

ix. The load rms voltage is 169.7V with the freewheel diode and increases without the diode to, as given by equation (11.20),

$$V_{rms} = V\left[\tfrac{1}{2\pi}\{\beta - \tfrac{1}{2}\sin 2\beta\}\right]^{\frac{1}{2}}$$

$$= 240\text{V}\left[\tfrac{1}{2\pi}\{4.08 - \tfrac{1}{2}\sin 2\times 4.08\}\right]^{\frac{1}{2}} = 181.6\text{V}$$

The rms load current from equation (11.20) is decreased to

$$i_{rms} = \frac{V}{Z}\left[\frac{1}{2\pi}\left\{\beta - \frac{\sin\beta\cos(\alpha+\beta+\phi)}{\cos\phi}\right\}\right]^{\frac{1}{2}}$$

$$= \frac{240\text{V}}{18.62\Omega}\times\left[\frac{1}{2\pi}\left\{4.08 - \frac{\sin 4.08\cos(4.08+1.57)}{\cos 1.57}\right\}\right]^{\frac{1}{2}} = 9.68\text{A}$$

Removal of the freewheel diode decreases the rms load current from 12.68A to 9.68A.

x. The load power is reduced without a load freewheel diode, from 1606.7W with a load freewheel diode, to

$$P_{10\Omega} = i_{rms}^2 R = 9.68^2\times 10\Omega = 937\text{W}$$

The supply power factor is also reduced, from 0.74 to

$$pf = \frac{P_{out}}{V_{rms}I_{rms}} = \frac{937\text{W}}{240\text{V}\times 9.68\text{A}} = 0.40$$

Load factor	circuit with freewheel diode		circuit without freewheel diode	
	form factor	ripple factor	form factor	ripple factor
	FF = $^{rms}/_{ave}$	RF = $\sqrt{FF^2-1}$	FF = $^{rms}/_{ave}$	RF = $\sqrt{FF^2-1}$
Voltage factor	169.7V/108V = 1.57	1.21	181.6V/86V = 2.1	1.86
Current factor	12.68A/10.8A = 1.17	0.615	9.68A/8.60A = 1.12	0.517

♣

11.1.5 Single-phase full-wave bridge rectifier circuit with a resistive load, R

The simplest meaningful single-phase full-wave load to analyse is the resistive load. The supply is impressed across the load every ac cycle half period, when load current flows.
The load voltage and current shown in figure 11.6a are defined by

$$v_o(\omega t) = i_o R = \left|\sqrt{2}V\sin\omega t\right| \qquad 0 \le \omega t \le 2\pi \tag{11.41}$$

The average dc output current and voltage are double the half-wave case and are given by

$$V_o = I_o R = \frac{1}{\pi}\int_0^{\pi}\sqrt{2}V\sin\omega t\ d\,\omega t = \frac{2\sqrt{2}}{\pi}V = 0.90V \tag{11.42}$$

The rms voltage across the load, and rms load current, are √2 greater than the half-wave case, specifically

$$V_{orms} = \left[\frac{1}{\pi}\int_0^{\pi}2V^2\sin^2\omega t\ d\,\omega t\right]^{\frac{1}{2}} = I_{orms}R = V \tag{11.43}$$

and the power dissipated in the load, specifically the load resistor *R*, is

$$P_o = I_{orms}^2 R = \frac{V^2}{R} \tag{11.44}$$

The ac current in the load is

$$I_{ac} = \sqrt{I_{orms}^2 - I_o^2} = \frac{V}{R}\left[1 - \frac{8}{\pi^2}\right]^{\frac{1}{2}} \tag{11.45}$$

The load voltage harmonics are (twice the half-wave case, without the supply frequency component)

$$v_o(\omega t) = \frac{2\sqrt{2}V}{\pi} - \frac{4\sqrt{2}V}{\pi}\left[\frac{1}{1\times 3}\cos 2\omega t + \frac{1}{3\times 5}\cos 4\omega t.. + \frac{1}{n^2-1}\cos n\omega t.....\right] \tag{11.46}$$

for n = 2, 4, 6, …

11.1.6 Single-phase full-wave bridge rectifier circuit with a resistive and back emf load, R-E

With an opposing emf *E* in the load circuit, the load current and voltage waveforms are as shown in figure 11.6b. Load current commences when

$$\omega t = \alpha = \sin^{-1}\frac{E}{\sqrt{2}V} \tag{11.47}$$

and ceases when

$$\omega t = \pi - \alpha = \pi - \sin^{-1}\frac{E}{\sqrt{2}V} \tag{11.48}$$

Diodes conduct every ac half cycle for a period $\theta = \pi - 2\alpha$, during which energy is delivered to both the load resistor R and load back emf E.
The load average and rms voltages are

$$\begin{aligned} V_o &= 2E\frac{\alpha}{\pi} + \frac{1}{\pi}\int_{\alpha}^{\pi-\alpha}\sqrt{2}V\sin\omega t\, d\omega t \\ &= 2E\frac{\alpha}{\pi} + \frac{2}{\pi}\sqrt{2}V\cos\alpha \end{aligned} \tag{11.49}$$

$$V_{orms} = \left[2\frac{\alpha}{\pi}E^2 + V^2\left(1 - 2\frac{\alpha}{\pi} + \frac{1}{\pi}\sin 2\alpha\right)\right]^{1/2} \tag{11.50}$$

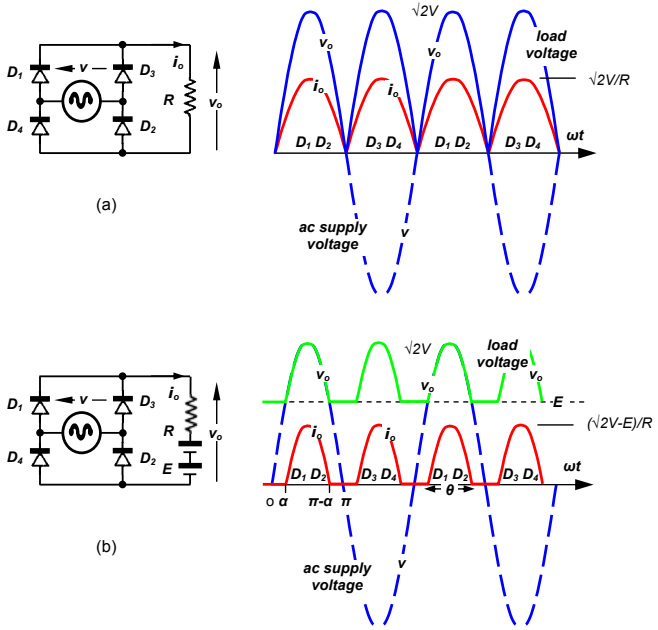


Figure 11.6. *Single-phase full-wave rectifiers:*
(a) purely resistive load, R and (b) resistive load R with back emf, E.

The load average and rms currents are

$$I_o = \frac{1}{R}\left[\frac{2\sqrt{2}V}{\pi}\sin\tfrac{1}{2}\theta - E\frac{\theta}{\pi}\right] \tag{11.51}$$

which is double the half-wave case and

$$I_{orms} = \frac{1}{R}\left[\frac{V^2}{\pi}\sin\theta - \frac{4\sqrt{2}}{\pi}VE\sin\tfrac{1}{2}\theta + \left(V^2 + E^2\right)\frac{\theta}{\pi}\right]^{1/2} \tag{11.52}$$

which is √2 greater than the half-wave case.
The total power delivered to the load is

$$P_o = P_R + P_E = I_{orms}^2 R + E\,I_o \tag{11.53}$$

Example 11.4: ***Full-wave rectifier with resistive and back emf load***

A dc motor, with series armature resistance of 10Ω and a back emf of 100V dc, is fed via a full-wave rectifier from the single-phase 230V 50Hz ac mains. Calculate

i. The average and rms motor voltages and currents, and diode maximum reverse voltage
ii. The supply power factor and motor efficiency

Solution

With a 100V back emf, the circuit and waveforms in figure 11.6b are applicable.
The current starts conducting when

$$\omega t = \alpha = \sin^{-1}\frac{E}{\sqrt{2}V} = \sin^{-1}\frac{100\text{V}}{\sqrt{2}\times 230\text{V}} = 17.9°$$

The current conducts for a period $\theta = \pi - 2\alpha = 180° - 2\times 17.9 = 144.2°$, ceasing at $\omega t = \pi - \alpha = 162.1°$.

i. The average and rms load currents and voltages are given by equations (11.49) to (11.52).

$$\begin{aligned} V_o &= 2E\frac{\alpha}{\pi} + \frac{2}{\pi}\sqrt{2}V\cos\alpha \\ &= 2\times 100\text{V}\frac{17.9°}{180°} + \frac{2}{\pi}\sqrt{2}\times 230\text{V}\times\cos 17.9° = 216.9\text{V} \end{aligned}$$

$$\begin{aligned} I_o &= \frac{1}{R}\left[\frac{2\sqrt{2}V}{\pi}\sin\tfrac{1}{2}\theta - E\frac{\theta}{\pi}\right] \\ &= \frac{1}{10\Omega}\left[\frac{2\sqrt{2}\times 230\,\text{V}}{\pi}\sin\tfrac{1}{2}\times 144.2° - 100\text{V}\times\frac{144.2°}{180°}\right] = 11.7\text{A} \end{aligned}$$

$$\begin{aligned} V_{orms} &= \left[2\frac{\alpha}{\pi}E^2 + V^2\left(1 - 2\frac{\alpha}{\pi} + \frac{1}{\pi}\sin 2\alpha\right)\right]^{1/2} \\ &= \left[2\times 100^2\frac{17.9°}{180°} + 230^2\left(1 - 2\times\frac{17.9°}{180°} + \frac{1}{\pi}\sin 2\times 17.9°\right)\right]^{1/2} = 231.4\text{V} \end{aligned}$$

$$\begin{aligned} I_{orms} &= \frac{1}{R}\left[\frac{V^2}{\pi}\sin\theta - \frac{4\sqrt{2}}{\pi}VE\sin\tfrac{1}{2}\theta + \left(V^2 + E^2\right)\frac{\theta}{\pi}\right]^{1/2} \\ &= \frac{1}{10\Omega}\left[\frac{230^2}{\pi}\sin 144.2° - \frac{4\sqrt{2}}{\pi}\times 230\text{V}\times 100\text{V}\times\sin\tfrac{1}{2}\times 144.2° + \left(230^2 + 100^2\right)\frac{144.2°}{180°}\right]^{1/2} \\ &= 14.43\text{A} \end{aligned}$$

The diode maximum reverse voltage is √2×230+100=425.3V.

ii. The motor loss is the loss in the 10Ω resistance in the dc motor equivalent circuit

$$P_R = I_{orms}^2 R = 14.43^2\times 10\Omega = 2082.2\text{W}$$

The back emf represents the source of electrical energy converted to mechanical energy

$$P_E = E\times I_o = 100\text{V}\times 11.7\text{A} = 1170\text{W}$$

The supply power factor is defined as the ratio: supply power delivered to apparent supply power

$$pf = \frac{P}{S} = \frac{P_R + P_E}{V\times I_{orms}} = \frac{2082.2\text{W} + 1170\text{W}}{230\text{V}\times 14.43\text{A}} = 0.98$$

The motor efficiency is

$$\eta = \frac{P_E}{P_R + P_E} = \frac{1170\text{W}}{2082.2\text{W} + 1170\text{W}}\times 100 = 36.0\%$$

♣

11.1.7 Single-phase, full-wave bridge rectifier circuit with an R-L load

Single-phase full-wave diode bridge circuits are shown in figures 11.7a and 11.7b. Both circuits appear identical as far as the load and supply are concerned. It will be seen in part b that two fewer diodes can be employed but this circuit requires a centre-tapped secondary transformer where each secondary has only a 50% copper utilisation factor. For the same output voltage, each of the secondary windings in figure 11.7b must have the same rms voltage rating as the single secondary winding of the transformer in figure 11.7a. The rectifying diodes in figure 11.7b experience twice the reverse voltage, (2√2 V), as that experienced by each of the four diodes in the circuit of figure 11.7a, (√2 V).

Figure 11.7c shows bridge circuit voltage and current waveforms. Assuming a 1:1(:1) transformer turns ratio, and with an inductive passive load, (no back emf) continuous load current flows, which is given by

$$i_o(\omega t) = \frac{\sqrt{2}V}{Z}\left[\sin(\omega t - \phi) + \frac{2\sin\phi}{1-e^{-\pi/\tan\phi}} \times e^{-\omega t/\tan\phi}\right] \qquad 0 \le \omega t \le \pi \qquad (11.54)$$

Appropriate integration of the load current squared, gives the rms load (and ac supply) current:

$$I_{rms} = \frac{V}{Z}\left[1 + 4\sin^2\phi\tan\phi \times \left(1 + e^{-\pi/\tan\phi}\right)\right]^{-\frac{1}{2}} = I_s \qquad (11.55)$$

The load experiences the transformer secondary rectified voltage which has a mean voltage (thence mean load current) of

$$V_o = \frac{1}{\pi}\int_0^{\pi}\sqrt{2}V\sin\omega t\, d\omega t = \bar{I}_o R = 2\sqrt{2}V/\pi = 0.90V \qquad \text{(V)} \qquad (11.56)$$

Since the average inductor voltage is zero, the average resistor voltage equals the average *R-L* voltage.

The rms value of the load circuit voltage v_0 is

$$V_{rms} = \sqrt{\frac{1}{2\pi}\int_0^{2\pi}\left(\sqrt{2}V\sin\omega t\right)^2 d\omega t} = V \qquad \text{(V)} \qquad (11.57)$$

From the load voltage definitions in section 11.4, the load voltage form factor is constant:

$$FF_v = \frac{V_{rms}}{V_o} = \frac{V}{2\sqrt{2}V/\pi} = \frac{\pi}{2\sqrt{2}} = 1.11 \qquad (11.58)$$

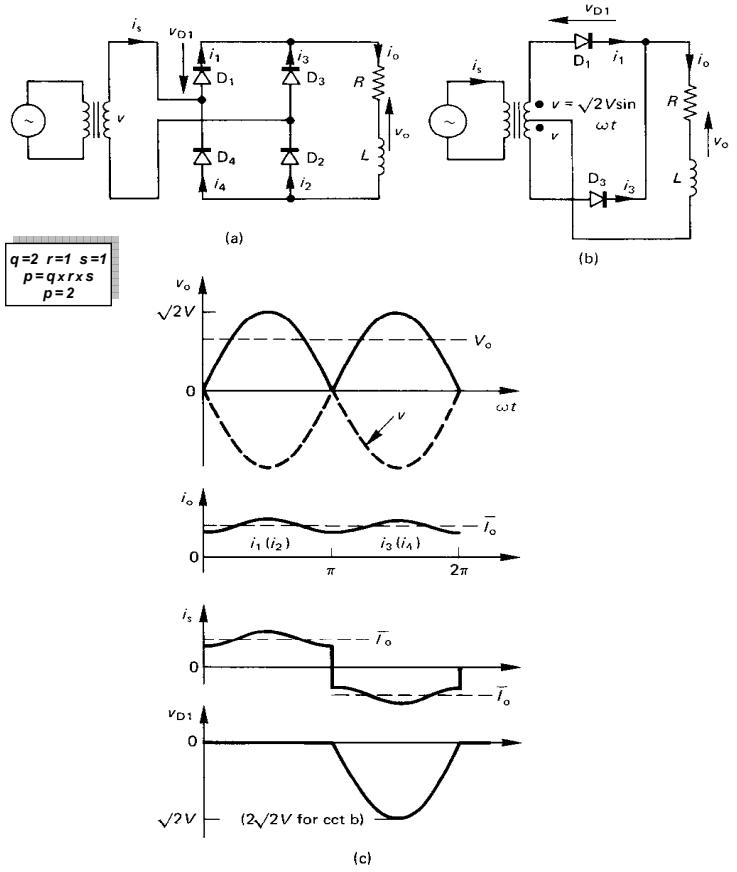


Figure 11.7. *Single-phase full-wave rectifier bridge: (a) circuit with four rectifying diodes; (b) circuit with two rectifying diodes; and (c) circuit waveforms.*

The load ripple voltage is

$$V_{RI} \triangleq \sqrt{V_{rms}^2 - V_o^2} = \sqrt{V^2 - \left(2\sqrt{2}/\pi\right)^2 V^2} = V\sqrt{1 - 8/\pi^2} = 0.435V \qquad \text{(V)} \qquad (11.59)$$

hence the load voltage ripple factor is

$$RF_v \triangleq V_{RI}/V_o = \sqrt{FF_v^2 - 1}$$
$$RF_v = \sqrt{1 - \left(2\sqrt{2}/\pi\right)^2} \;/\; 2\sqrt{2}/\pi = \pi^2/8 - 1 = 0.483 \qquad (11.60)$$

which is significantly less (better) than the half-wave rectified value of 1.211 from equation (11.32).

The output voltages and currents (rms and average) can be derived from the voltage Fourier expansion in equation (11.46):

$$v_o(\omega t) = \frac{2\sqrt{2}V}{\pi} + \frac{2\sqrt{2}V}{\pi}\sum_{n=2,4,6}^{\infty}\frac{2}{n^2-1}\cos n\omega t \qquad (11.61)$$

The first term is the average output voltage, as given by equation (11.56). Note the harmonic magnitudes decrease rapidly with increased order, namely $\frac{2}{3}:\frac{2}{15}:\frac{2}{35}:\frac{2}{63}:\cdots$. The output voltage is therefore dominated by the dc component and the harmonic at 2ω.

The output current can be derived by dividing each voltage component by the appropriate load impedance at that frequency. That is

$$\bar{I}_o = \frac{V_o}{R} = \frac{2\sqrt{2}V}{\pi R}$$
$$I_n = \frac{V_n}{Z_n} = \frac{2\sqrt{2}V}{\pi} \times \frac{\frac{2}{n^2-1}}{\sqrt{R^2 + (n\omega L)^2}} \qquad \text{for} \quad n = 2,4,6.. \qquad (11.62)$$

The load rms current whence load power, critical load inductance, and power factor, are given by

$$I_{rms} = \sqrt{I_o^2 + \sum_{n=2,4}^{\infty} \tfrac{1}{2} \times I_n^2} \qquad P_L = I_{rms}^2 R$$
$$pf = \frac{P_L}{V I_{rms}} = \frac{I_{rms}R}{V} \qquad L_{critical} = \frac{R}{3\times\omega} \quad \text{(see equation 11.65)} \qquad (11.63)$$

Each diode rms current is $I_{rms}/\sqrt{2}$. For the circuit in figure 11.7a, the transformer secondary winding rms current is I_{rms}, while for the centre-tapped transformer, for the same load voltage, each winding has an rms current rating of $I_{rms}/\sqrt{2}$. The primary current rating is the same for both transformers and is related to the secondary rms current rating by the turns ratio. Power factor is independent of turns ratio.

11.1.7i - Single-phase full-wave bridge rectifier circuit with an output L-C filter

A – with an output L-C filter and continuous inductor current

Table 11.1 shows three typical single-phase, full-wave rectifier output stages, where part c is a typical output filtering stage used to obtain a near constant dc output voltage.
If it is assumed that the load inductance is large and the load resistance small such that continuous load current flows, then the bridge average output voltage $\bar{V}_o$ is the same as the average voltage across the load resistor since the average voltage across the filter inductor is zero. From equation (11.61), the dominant load voltage harmonic is due to the second harmonic therefore the ac current is predominately the second harmonic current, $I_{o,ac} \approx I_{o,2}$. By neglecting the higher order harmonics, the various circuit currents and voltages can be readily obtained as shown in table 11.1. From equation (11.61) the output voltage is given by

$$\begin{aligned} v_o(\omega t) &= \bar{V}_o + V_{o,2}\cos 2\omega t \\ &= \frac{2\sqrt{2}V}{\pi} + \frac{2\sqrt{2}V}{\pi}\times\frac{2}{n^2-1}\cos n\omega t \quad \text{for } n = 2 \\ &= \frac{2\sqrt{2}V}{\pi} + \frac{2\sqrt{2}V}{\pi}\times\frac{2}{3}\cos 2\omega t \\ &= 0.90V + 0.60V\times\cos 2\omega t \end{aligned} \qquad (11.64)$$

With the filter capacitor across the load resistor, the average inductor current is equal to the average resistor current, since the average capacitor current is zero.

With continuous inductor current, the inductor current is

$$i_o(\omega t) = \bar{I}_o + I_{o,2}\cos 2\omega t = \frac{\bar{V}_o}{R} + \frac{2}{3}\frac{\bar{V}_o}{Z_2}\cos 2\omega t = \frac{0.90V}{R} + \frac{0.60V}{\sqrt{R^2+(2\omega L)^2}}\times\cos 2\omega t \quad (11.65)$$

From equation (11.65) for continuous inductor current, the average current must be larger than the peak second harmonic current magnitude, that is

$$\bar{I}_o > \left|I_{o,2}\right| \qquad \frac{\bar{V}_o}{R} > \frac{2}{3}\frac{\bar{V}_o}{Z_2} \quad (11.66)$$

Since the load resistance must be low enough to ensure continuous inductor current, then $2\omega L > R$ such that $Z_2 = \sqrt{R^2+(2\omega L)^2} \approx 2\omega L$. Equation (11.66) therefore gives the following load identity for continuous inductor current

$$\frac{1}{R} > \frac{2}{3}\frac{1}{Z_2} = \frac{1}{3\omega L} \text{ that is } L/R > 1/3\omega \quad (11.67)$$

The load and supply (peak) ac currents are $I_{o,ac} = I_{s,ac} = I_{o,2}$. The output and supply rms currents are

$$I_{o,rms} = I_{s,rms} = \sqrt{\bar{I}_o^2 + \tfrac{1}{2}I_{o,ac}^2} = \sqrt{\bar{I}_o^2 + \tfrac{1}{2}I_{o,2}^2} \quad (11.68)$$

and the power delivered to resistance *R* in the load is

$$P_R = I_{o,rms}^2 R \quad (11.69)$$

B – with an output L-C filter and discontinuous inductor current

If the inductor current reduces to zero, at angle β, all the load current is provided by the capacitor. Its voltage falls to V_o ($<\sqrt{2}\,V$) and inductor current recommences when

$$v_L = \sqrt{2}V\sin\omega t - V_o \quad (11.70)$$

at an angle

$$\alpha = \sin^{-1}\frac{V_o}{\sqrt{2}V} \quad (11.71)$$

By integrating $v = L\,di/dt$ for i, the inductor current is of the form

$$i_L(\omega t) = \frac{1}{\omega L}\left(\sqrt{2}V(\cos\alpha - \cos\omega t) - V_o(\omega t - \alpha)\right) \quad (11.72)$$

where $\alpha \le \omega t \le \beta$. The voltage V_o is found from equation (11.72) by iterative techniques.

11.1.7ii Single-phase, full-wave bridge rectifier circuit with an R-L-E load

An R-L load incorporating a back emf E, is shown in Table 11.1.

For **continuous** output current

When continuous load current flows, the rectified supply is continuously impressed across the series L-R-E load, therefore the average and rms output voltages respectively are

$$\bar{V}_o = \frac{1}{\pi}\int_o^\pi \sqrt{2}V_s\sin\omega t\,d\omega t = \frac{2\sqrt{2}}{\pi}V_s \qquad V_o = \sqrt{\frac{1}{\pi}\int_o^\pi\left(\sqrt{2}V_s\sin\omega t\right)^2 d\omega t} = V_s \quad (11.73)$$

Hence the output voltage form and ripple factors are

$$FF_v = \frac{V_o}{\bar{V}_o} = \frac{\pi}{2\sqrt{2}} \qquad RF_v = \sqrt{V_{FF}^2 - 1} = \sqrt{\frac{\pi^2}{8} - 1} \quad (11.74)$$

If the input current is approximated by its fundamental, $4\bar{I}_o/\pi$, the following input characteristics are realised

$$\text{Input displacement factor} = DPF = \cos\phi_i = \cos 0° = 1$$

$$\text{Distortion factor} = DF_{i1} = \frac{I_{i1}}{\bar{I}_o} = \frac{2\sqrt{2}}{\pi}$$

$$\text{power factor} = pf = DPF \times DF_{i1} = \frac{2\sqrt{2}}{\pi}$$

$$THD_{i1} = \sqrt{\frac{1-DF_{i1}^2}{DF_{i1}^2}}\times 100 = \sqrt{\frac{\pi^2}{8}-1}\times 100$$

The output current is found by solving

$$\sqrt{2}V_s\sin\omega t = Ri_o + L\frac{di_o}{dt} + E \quad (11.75)$$

which in steady state yields

$$i_o(t) = I_o e^{-\frac{\omega t}{\tan\theta}} + \sqrt{2}V_s\left[\sin(\omega t - \theta) - \frac{\sin\alpha}{\cos\theta}\right] \quad (11.76)$$

where $\tan\theta = \frac{\omega L}{R}$; $Z = \sqrt{R^2+\omega^2L^2}$; $\sin\alpha = \frac{E}{\sqrt{2}V_s}$

and the boundary conditions give

$$I_o = \frac{\sqrt{2}V_s}{Z}\frac{2\sin\theta}{1-e^{\frac{-\pi}{\tan\theta}}}$$

For continuous conduction, $i_o(\omega t = \theta) \ge 0$ in equation (11.76) gives the condition

$$\frac{2\sin\theta}{1-e^{\frac{-\pi}{\tan\theta}}}e^{\frac{-\alpha}{\tan\theta}} \ge \sin(\theta-\alpha) + \frac{\sin\alpha}{\cos\theta} \quad (11.77)$$

If the left hand side is less than the right hand side, **discontinuous** current flows in the load, and if the current extinction angle is β, then the average output voltage is given by

$$\bar{V}_o = \frac{1}{\pi}\left[\int_\alpha^\beta \sqrt{2}V_s\sin\omega t\,d\omega t + \int_\beta^{\pi+\alpha} E\,d\omega t\right]$$

$$\bar{V}_o = \frac{\sqrt{2}V}{\pi}\left[\cos\alpha - \cos\beta + (\pi+\alpha-\beta)\sin\alpha\right] \quad (11.78)$$

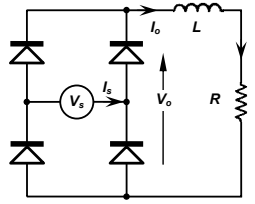
In the general solution to the circuit differential equation in equation (11.76), for discontinuous output current, (zero current boundary conditions), I_o for equation (11.76) becomes (during conduction)

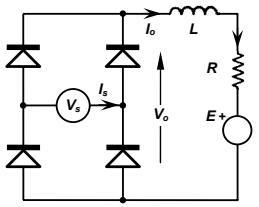
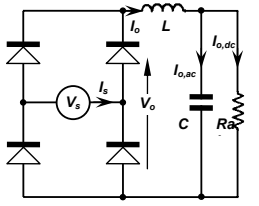
$$I_o = \frac{\sqrt{2}V_s}{Z}\left[\sin(\theta-\beta) + \frac{\sin\alpha}{\cos\theta}\right]$$

The conduction period β is found by iteratively solving

$$\sin(\beta-\alpha) = \left[1-e^{\frac{\alpha-\beta}{\tan\theta}}\right]\times\frac{\sin\alpha}{\cos\theta} - e^{\frac{\alpha-\beta}{\tan\theta}}\times\sin(\theta-\alpha) \quad (11.79)$$

Table 11.1: Single-phase full-wave uncontrolled rectifier circuits – continuous inductor current

Full-wave rectifier circuit: Load	circuit	2nd harmonic current $I_{o,2}$ (A)	average output current $\bar{I}_o$ (A)	output power P_R+P_E (W)
(a) ***R-L*** *see section 11.1.7i and 12.2.3 α = 0*		$\frac{V_{o,2}}{\sqrt{R^2+(2\omega L)^2}}$	$\frac{\bar{V}_o}{R}$ $= \frac{2\sqrt{2}V}{\pi R}$	$I_{o,rms}^2 R$

(b) **R-L-E** see section 11.7ii and 12.2.4 $\alpha = 0$		$\dfrac{V_{o,2}}{\sqrt{R^2+(2\omega L)^2}}$	$\dfrac{\overline{V}_o - E}{R}$ $= \dfrac{1}{R}\left(\dfrac{2\sqrt{2}V}{\pi} - E\right)$	$I_{o,rms}^2 R + \overline{I}_o E$
(c) **L-R//C** see section 11.1.7i		$\dfrac{V_{o,2}}{2\omega L}$	$\dfrac{\overline{V}_o}{R}$ $= \dfrac{2\sqrt{2}V}{\pi R}$	$I_{o,rms}^2 R = \overline{I}_o^2 R$

Example 11.5: ***Full-wave diode rectifier with an L-C filter and continuous load current***

A single-phase, full-wave, diode rectifier is supplied from a 230V ac, 50Hz voltage source and uses an *L-C* output filter with a resistor load, as shown in the last circuit in Table 11.1. The average inductor current is 10A with a 4A rms ripple current dominated by the 100Hz component. Ignoring diode voltage drops and initially assuming the output voltage is ripple free, determine

i. the dc output voltage, hence load resistance and power
ii. the dc filter inductance and its average voltage, whilst neglecting any capacitor voltage ripple
iii. the dc filter capacitance if its peak-to-peak ripple voltage is 5% the average voltage
iv. diode average, rms, and peak current
v. the supply power factor

Solution

Since $\sqrt{2}\, I_{o\,rms,2} < \overline{I}_o \left(\sqrt{2}\times 4\text{A} < 10\text{A}\right)$, the output current is continuous.

i. The dc output voltage is $\overline{V}_o$ = 0.9×230V = 207V. Assuming the 207V is ripple free, that is, $V_{rms} = V_{dc}$, then the load resistance and power dissipated are

$$R = \frac{\overline{V}_o}{\overline{I}_o} = \frac{207\text{V}}{10\text{A}} = 20.7\Omega$$

$$P_R = \overline{V}_o \times \overline{I}_o = 207\text{V}\times 10\text{A} = 2070\text{W}$$

ii. The 100Hz voltage component in the output voltage is given by equation (11.64), that is

$$V_{o,2} = \frac{2\sqrt{2}V}{\pi}\times\frac{2}{n^2-1}\cos n\omega t$$
$$= \frac{2\sqrt{2}V}{\pi}\times\frac{2}{3}\cos 2\omega t$$
$$= 0.60\times 230\text{V}\times\cos 2\omega t = 138\times\cos 2\omega t$$

which has an rms value of 138/√2 = 97.6V. The 100Hz rms current $I_{o,2}/\sqrt{2}$ produced by this voltage is 4A thus

from $\dfrac{I_{o,2}}{\sqrt{2}} = \dfrac{V_{o,2}}{2\omega L}$

$$L = \frac{V_{o,2}}{2\omega I_{o,2}} = \frac{97.6\text{V}}{2\times 2\pi 50\text{Hz}\times 4\text{A}} = 38.8\text{mH}$$

The average inductor voltage is zero.

iii. From part i, the dc output voltage is 207V. The peak-to-peak ripple voltage is 5% of 207V, that is 10.35V. This gives an rms value of 10.35V /2√2 = 3.66V. From

$$V_{o,2}/\sqrt{2} = I_{o,2}/\sqrt{2}\times X_{c,100\text{Hz}} = \frac{I_{o,2}/\sqrt{2}}{2\omega C}$$
$$\Rightarrow C = \frac{I_{o,2}}{2\omega\times V_{o,2}} = \frac{4\text{A}}{2\times 2\pi 50\text{Hz}\times 3.66\text{V}} = 1.7\text{mF}$$

iv. The diode currents are

$$I_{D,rms} = I_{o,rms}/\sqrt{2} = \sqrt{\overline{I}_o^2 + \tfrac{1}{2}I_{o,2}^2}/\sqrt{2} = \sqrt{10\text{A}^2 + 4\text{A}^2}/\sqrt{2} = 10.8\text{A}/\sqrt{2} = 7.64\text{A}$$

$$\overline{I}_D = \overline{I}_o/2 = 10\text{A}/2 = 5\text{A}$$

$$\hat{I}_D = \overline{I}_o + I_{o,2} = 10\text{A} + \sqrt{2}\times 4\text{A} = 15.7\text{A}$$

v. The input and output rms current is

$$I_s = I_{o,rms} = \sqrt{\overline{I}_o^2 + \tfrac{1}{2}I_{o,2}^2} = \sqrt{10\text{A}^2 + 4\text{A}^2} = 10.8\text{A}$$

Assuming the input power equals the output power, then from part i, $P_o = P_i$ = 2070W. The supply power factor is

$$pf = \frac{P_i}{S} = \frac{P_i}{V_s I_s} = \frac{2070\text{W}}{230\text{V}\times 10.8\text{A}} = 0.83$$

♣

11.1.7iii - Single-phase full-wave bridge rectifier with highly inductive load– constant load current

With a highly inductive load, which is the usual practical case, virtually constant load current flows, as shown dashed in figure 11.7c. The bridge diode currents are then square wave 180º blocks of current of magnitude $\overline{I}_o$. The diode current ratings can now be specified and depend on the pulse number *p*. For this full-wave single-phase application each input cycle comprises two 180º output current pulses, hence *p* = 2.

The mean current in each diode is

$$\overline{I}_D = \tfrac{1}{p}\overline{I}_o = \tfrac{1}{2}\overline{I}_o \quad \text{(A)} \qquad (11.80)$$

and the rms current in each diode is

$$I_D = \tfrac{1}{\sqrt{p}}\overline{I}_o = \overline{I}_o/\sqrt{2} \quad \text{(A)} \qquad (11.81)$$

whence the diode current form factor is

$$RF_{ID} = I_D/\overline{I}_D = \sqrt{p} = \sqrt{2} \qquad (11.82)$$

Since the load current is approximately constant, power delivered to the load is

$$P_o \approx \overline{V}_o\overline{I}_o = \tfrac{8}{\pi^2}\times V^2/R \quad \text{(W)} \qquad (11.83)$$

The supply power factor is $pf = V_o/V = 2\sqrt{2}/\pi = 0.90$, since $\overline{I}_o = I_{rms}$.

11.1.7iv - Single-phase full-wave bridge rectifier circuit with a C-filter and resistive load

The capacitor smoothed single-phase full-wave diode rectifier circuit shown in figure 11.8a is a common power rectifier circuit used to obtain unregulated dc voltages. The circuit is simple and cheap but the input current has high peak and rms values, high harmonics, and a poor power factor. The full-wave rectified case is an extension of the half-wave case considered in section 11.1.4X

The capacitor reduces the ripple voltage, so large voltage-polarised capacitance is used to produce an almost constant dc output voltage. Isolation and voltage matching (step-up or step down) are obtained by using a transformer before the diode rectification stage as shown in figures 11.7a and b. The resistor *R* across the filter capacitor represents a resistive dissipative load.

As the ac supply voltage rises to its extremes each half cycle, as shown in figure 11.8b, a pair of rectifier diodes D1-D2 or D3-D4, alternately become forward biased at time $\omega t = \alpha$. The ac supply provides load resistor current and simultaneously charges the capacitor, its voltage having drooped whilst providing the load current during the previous diode non-conduction period. The capacitor charging current period θ_c around the ac supply extremes is short, giving a high peak to rms ratio of diode and supply current. When all the rectifier diodes are reverse biased at $\omega t = \beta$ because the capacitor voltage is greater than the instantaneous supply ac voltage, the capacitor supplies the load current and its voltage decreases with an *R-C* time constant until $\omega t = \pi + \alpha$. The output voltage and diode voltages, plus load current v_o/R, and capacitor current $C\, dv_o/dt$ are defined in Table 11.2.

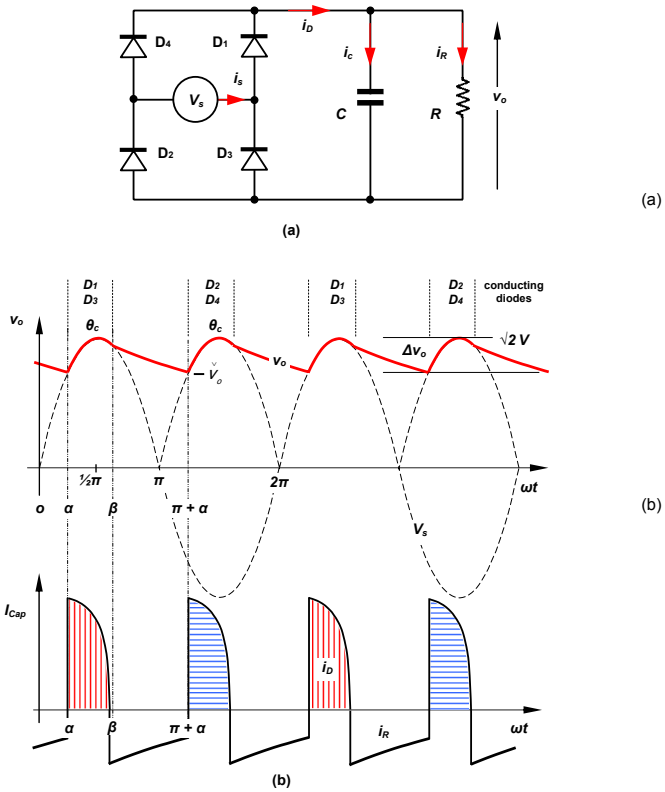


Figure 11.8. *Single-phase full-wave rectifier bridge: (a) circuit with C-filter capacitor and (b) circuit waveforms.*

Table 11.2: Single-phase, full-wave rectifier voltages and currents

$v_s(\omega t) = \sqrt{2}V\sin\omega t$		Diodes conducting	Diodes non-conducting
		$\alpha \le \omega t \le \beta$	$\beta \le \omega t \le \pi + \alpha$
Output voltage	$v_o(\omega t)$	$\sqrt{2}V\lvert\sin\omega t\rvert$	$\sqrt{2}V\sin\beta\times e^{-(\omega t-\beta)/\tan\beta}$
Diode voltage	$v_D(\omega t)$	0 and $-\sqrt{2}V\sin\omega t$	$-\sqrt{2}V\sin\beta\times e^{-(\omega t-\beta)/\tan\beta}+\sqrt{2}V\sin\omega t$
Capacitor current	$i_c(\omega t)$	$\frac{\sqrt{2}V}{X}\cos\omega t$	$\frac{\sqrt{2}V}{Z}\times e^{-(\omega t-\beta)/\tan\beta}$
Resistor current	$i_R(\omega t)$	$\frac{\sqrt{2}V}{R}\sin\omega t$	$\frac{\sqrt{2}V}{Z}\times e^{-(\omega t-\beta)/\tan\beta}=-i_c(\omega t)$
Diode bridge current	$i_D(\omega t)=$ $i_c(\omega t)+i_R(\omega t)$	$\frac{\sqrt{2}V}{R\cos\phi}\times\sin(\omega t+\phi)$	0

The start of diode conduction, α, the diode current extinction angle, β, hence diode conduction period, θ_c, are specified by the following equations.
From $i_c + i_R = 0$ at $\omega t = \beta$:

$$\frac{\sqrt{2}V}{X}\cos\beta+\frac{\sqrt{2}V}{R}\sin\beta=0 \qquad (11.84)$$

$$\beta=\tan^{-1}(-\omega RC)=\pi-\tan^{-1}(\omega RC) \qquad ½\pi\le\beta\le\pi$$

By equating the two expression for output voltage at the boundary $\omega t = \pi+\alpha$ gives

$$\sqrt{2}V\lvert\sin(\pi+\alpha)\rvert=\sqrt{2}V\sin\beta\times e^{-(\pi+\alpha-\beta)/\tan\beta} \qquad (11.85)$$

and a transcendental expression for α results:

$$\sin\alpha-\sin\beta\times e^{-(\pi+\alpha-\beta)/\omega RC}=0 \qquad (11.86)$$

The diode current conduction period θ_c is given by

$$\theta_c=\beta-\alpha \qquad (11.87)$$

When the diodes conduct, R and C are in parallel and $\tan\phi=\omega CR$.

When the diodes are not conducting, the output circuit current flows in a series R-C circuit with a fundamental impedance of:

$$Z=\sqrt{R^2+X^2} \quad \text{and} \quad X=1/\omega C$$

The resistor average voltage and current are

$$\bar{V}_R=\frac{\sqrt{2}V}{\pi}\frac{(1-\cos\theta_c)}{-\cos\beta}=\bar{I}_R R \qquad (11.88)$$

The maximum output voltage occurs at ωt = ½π when $v_o = \hat{V}_o = \hat{V}_s = \sqrt{2}V$, while the minimum output voltage occurs at the end of the capacitor discharge period when $\omega t = \alpha$ *and* $v_o = \check{V}_o = \sqrt{2}V\sin\alpha$. The output peak-to-peak ripple voltage is therefore the difference:

$$\Delta V_o=\hat{V}_o-\check{V}_o=\sqrt{2}V-\sqrt{2}V\sin\alpha=\sqrt{2}V(1-\sin\alpha) \qquad (11.89)$$

By assuming $\alpha \approx$ ½π, $\beta \approx$ ½π, and a series expansion for the exponent

$$\Delta V_o\approx\frac{\sqrt{2}V\pi}{\omega RC}=\frac{V}{\sqrt{2}fRC} \qquad (11.90)$$

The ac source current is the sum of the diode currents, that is

$$i_s=i_{D1,2}-i_{D3,4}=i_R+i_c \qquad (11.91)$$

when $\alpha<\omega t<\beta$. Otherwise $i_s = 0$.

Since the capacitor voltage is in steady-state, the average capacitor current is zero, thus for full-wave rectification, the average diode current is half the average load current.
The peak capacitor current occurs at $\omega t = \alpha$, when the diodes first conduct. From the capacitor current equation in table 11.2:

$$\hat{I}_c=\sqrt{2}V\omega C\cos\alpha \qquad (11.92)$$

From table 11.2, the peak diode current occurs at the same time as the peak capacitor current, $\omega t = \alpha$:

$$\hat{I}_D=i_c(\pi+\alpha)+i_R(\pi+\alpha)$$
$$=\sqrt{2}V\omega C\cos\alpha+\frac{\sqrt{2}V}{R}\sin\alpha=\frac{\sqrt{2}V}{X}\cos\alpha+\frac{\sqrt{2}V}{R}\sin\alpha=\frac{\sqrt{2}V}{R}\frac{Z}{X}\sin(\alpha+\phi) \qquad (11.93)$$

Similar expressions can be derived for the half-wave rectifier case. For the non-conduction period, $\beta=2\pi+\alpha$. The output ripple voltage is about twice that given by equation (11.90) and the average resistor voltage in equation (11.88) (after modification), is reduced.

Example 11.6: ***Single-phase full-wave bridge rectifier circuit with C-filter and resistive load***

A single-phase, full-wave, diode rectifier is supplied from a 230V ac, 50Hz voltage source and uses a capacitor output filter, 1000μF, with a resistor 100Ω load, as shown in Figure 11.8a. Ignoring diode voltage drops, determine

i. expressions for the output voltage
ii. output voltage ripple Δv_o and the % error in using the approximation equation (11.90)
iii. expressions for the capacitor current
iv. diode peak current
v. the average load voltage and current

Assuming the output ripple voltage is triangular, estimate

vi. the average output voltage and rms output ripple voltage
vii. capacitance C for Δv_o = 2% of the maximum output voltage

Solution

The supply voltage is $v_s = \sqrt{2}\times 230 \sin 2\pi 50t$, which has a peak value of $\hat{V}_s = 325.3$V.

$$\omega RC = 2\pi 50\text{Hz}\times 100\Omega \times 1000\mu\text{F} = 31.416 \text{ rad}$$

Thus $X = 1/\omega C = 3.1831\Omega$ and $Z = 100.0507\Omega$.
(5 figure accuracy is used because of the sensitivity of the applicable equations around α =90°.)
From equation (11.84) the diode current extinction angle β is

$$\beta = \pi - \tan^{-1}(\omega RC) = \pi - \tan^{-1}(31.416\text{rad}) = 1.6026\,\text{rad} = 91.8°$$

The diode current turn-on angle α is solve iteratively from equation (11.86), that is

$$\sin\alpha - \sin\beta \times e^{-(\pi+\alpha-\beta)/\omega RC} = 0$$

$$\sin\alpha - \sin 1.603 \times e^{-(\pi+\alpha-1.603)/31.416} = 0$$

gives α = 1.16095 rad or 66.5°. The diode conduction period is $\theta_c = \beta - \alpha$ =1.6026 -1.16095 = 0.44167rad or 25.3°.

i. From table 11.2, the output voltage, which is the capacitor voltage, is given by

$$v_o(\omega t) = \left|\sqrt{2}V\sin\omega t\right| = \left|325.27\text{V}\times\sin\omega t\right|$$

$$66.5° \le \omega t \le 91.8°$$

$$v_o(\omega t) = \sqrt{2}\times 230\text{V}\times\sin 1.6026\text{rad}\times e^{-(\omega t - 1.6026\text{rad})/31.416\text{rad}} = 325.13\times e^{-(\omega t - 1.6026\text{rad})/31.416\text{rad}}$$

$$91.8° \le \omega t \le 246.5°$$

ii. The output voltage ripple Δv_o is given by equation (11.89), that is

$$\Delta V_o = \sqrt{2}\,V(1-\sin\alpha) = \sqrt{2}\,230\text{V}\times(1-\sin 1.16026) = 26.94\text{V p-p}$$

From equation (11.90)

$$\Delta V_o \approx \frac{V}{\sqrt{2}\,f\,RC} = \frac{230\text{V}}{\sqrt{2}\times 50\text{Hz}\times 100\Omega\times 1000\mu\text{F}} = 32.5\text{V}$$

The approximation predicts a higher ripple: a +21% over-estimate.

iii. From table 11.2, the capacitor current is

$$i_c(\omega t) = \sqrt{2}V\omega C\cos\omega t = \sqrt{2}\,230\text{V}\times 2\pi 50\text{Hz}\times 1000\mu\text{F}\times\cos\omega t = 102.2\times\cos\omega t$$

$$66.5° \le \omega t \le 91.8°$$

$$i_c(\omega t) = \frac{\sqrt{2}V\sin\beta}{R}\times e^{-(\omega t-\beta)/\omega RC} = \frac{\sqrt{2}230\text{V}\times\sin 1.16}{100\Omega}\times e^{-(\omega t - 1.16)/31.4} = 3.0\times e^{-(\omega t-1.16)/31.4}$$

$$91.8° \le \omega t \le 246.5°$$

iv. The peak diode current is given by equation (11.93):

$$\hat{I}_D = \sqrt{2}V\omega C\cos\alpha + \frac{\sqrt{2}V}{R}\sin\alpha$$

$$= \sqrt{2}\,230\text{V}\times 2\pi 50\text{Hz}\times 1000\mu\text{F}\times\cos 1.16026 + \frac{\sqrt{2}\,230\text{V}}{100\Omega}\times\sin 1.16026$$

$$= 40.7\text{A} + 3\text{A} = 43.7\text{A}$$

The peak diode current is dominated by the capacitor initial charging current of 40.7A

v. The average load voltage and current are given by equation (11.88)

$$\bar{V}_R = \frac{\sqrt{2}V}{\pi}\frac{(1-\cos\theta_c)}{-\cos\beta}$$

$$= \frac{\sqrt{2}\,230\text{V}}{\pi}\times\frac{(1-\cos 0.4417)}{-\cos 1.603} = 312.3\text{V}$$

$$\bar{I}_R = \frac{\bar{V}_R}{R} = \frac{312.3\text{V}}{100\Omega} = 3.12\text{A}$$

vi. If the ripple voltage is assumed triangular then

(a) The average output voltage is the peak output voltage minus half the ripple voltage, that is

$\hat{V}_s$ - ½Δv_o = √2×230V - ½×26.9V = 311.8V

which is less than that given by the accurate equation (11.89), 312.3V.

(b) If the 26.9V p-p ripple voltage is assumed triangular then its rms value is ½×26.9/√3 =7.8V rms.

vii. Re-arrangement of equation (11.90), which under-estimates the capacitance requirement for 2% ripple, gives

$$C = \frac{V}{\sqrt{2}\,f\,R\times\Delta V_o} = \frac{\hat{V}_s}{2f\,R\times 2\%\text{ of }\hat{V}_s} = \frac{1}{2f\,R\times 2\%}$$

$$= \frac{1}{2\times 50\text{Hz}\times 100\Omega\times 0.02} = 5{,}000\mu\text{F}$$

♣

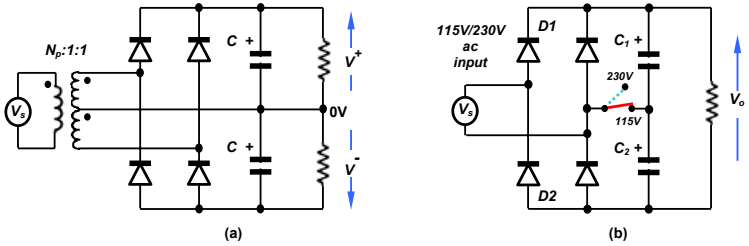


Figure 11.9. *Bridge rectifiers: (a) split rail dc supplies and (b) voltage doubler.*

11.1.7v - Other single-phase bridge rectifier circuit configurations

Figure 11.9a shows a transformer used to create a two-phase supply (each phase is 180° apart), which upon rectification produce equal split-rail dc output voltages, V^+ and V^-. The electrical characteristics can be analysed as in the case of the single-phase full-wave bridge rectifier circuit with a capacitive C-filter and resistive load, in section 11.1.7iv. In the split rail case, the rectifiers conduct every 180°, alternately feeding each output voltage rail capacitor. Thus the diode average and rms currents are increased by 2 and √2 respectively, above those of a conventional single phase rectifier.
The voltage doubler in figure 11.9b can be used in equipment that must be able to operate from both 115Vac and 230V ac voltage supplies, without the aid of a voltage-matching transformer. With the switch in the 115V position, the output is twice the peak of the input ac supply. The capacitor C_1 charges through diode D1, and when the supply reverses, capacitor C_2 charges through D2. Since C_1 and C_2 are in series, the output voltage is the sum $V_{C1}+V_{C2}$, where each capacitor is alternately charged (half-wave rectified) from the ac source V_s. The other, unused, two diodes remain reverse biased, and are only necessary if the dual input voltage function is required.
With the switch in the 230V ac position (open circuit), standard rectification occurs, with the two series capacitors charging simultaneously every half cycle. In dual frequency applications (110V ac, 60Hz and 230V ac, 50Hz), the capacitance requirements are based on the supply with the lower frequency, 50Hz.

11.2 Three-phase uncontrolled rectifier converter circuits

Single-phase supply circuits are adequate below a few kilowatts. At higher power levels, restrictions on unbalanced loading, line harmonics, current surge voltage dips, and filtering require the use of three-phase (or higher - polyphase) converter circuits. Generally it will be assumed that the output current is both continuous and smooth. This assumption is based on the dc load being highly inductive.

11.2.1 Three-phase half-wave rectifier circuit with an inductive R-L load

Figure 11.10 shows a half-wave, three-phase diode rectifier circuit along with various circuit voltage and current waveforms. A transformer having a star connected secondary is required for neutral access, N.
The diode with the highest potential with respect to the neutral conducts a rectangular current pulse. As the potential of another diode becomes the highest, load current is transferred to that device, and the previously conducting device is reverse-biased and naturally (line) commutated. Note that the load voltage, hence current never reaches zero, when the load is passive (no opposing back emf).
In general terms, for an *n*-phase *p*-pulse system, the mean output voltage is given by (see example 11.8)

$$V_o = \frac{1}{2\pi/p}\int_{-\pi/p}^{\pi/p}\sqrt{2}\,V\cos\omega t\,d\omega t \quad \text{(V)}$$

$$= \sqrt{2}\,V\frac{\sin(\pi/p)}{\pi/p} \quad \text{(V)} \qquad (11.94)$$

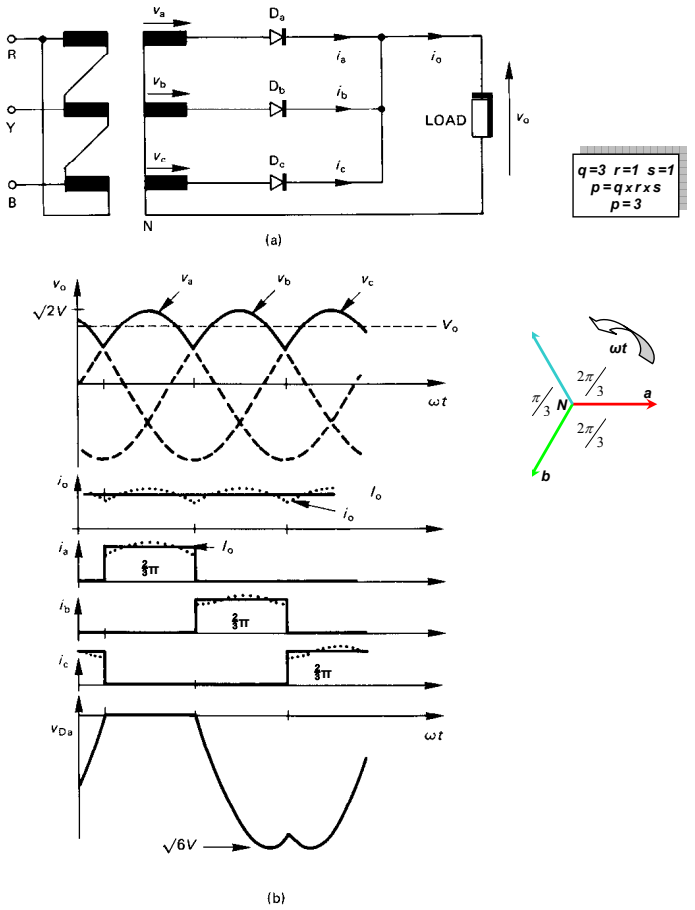


Figure 11.10. *Three-phase half-wave diode rectifier: (a) circuit diagram and (b) circuit voltage and current waveforms.*

For a three-phase, half-wave circuit (p = 3) the mean output voltage, (thence average current) is

$$V_o = \bar{I}_o R = \frac{1}{2\pi/3}\int_{\pi/6}^{5\pi/6} \sqrt{2}\,V \sin\omega t\, d\omega t$$
$$= \sqrt{2}V\frac{\frac{1}{2}\sqrt{3}}{\pi/3} \qquad = 1.17 \times V \quad \text{(V)} \tag{11.95}$$

The rms load voltage is

$$V_{rms} = \sqrt{\frac{1}{2\pi/3}\int_{\pi/6}^{5\pi/6} \left(\sqrt{2}\,V\right)^2 \sin^2\omega t\, d\omega t} = \sqrt{2}\,V\left[\frac{3}{2\pi}\left(\frac{\pi}{3}+\frac{\sqrt{3}}{4}\right)\right]^{\frac{1}{2}} = 1.19 \times V \tag{11.96}$$

The load voltage form factor is

$$FF_v = \frac{V_{rms}}{V} = 1.19V/1.17V = 1.01 \tag{11.97}$$

$$\text{RF}_v = \text{ripple factor} = \frac{\text{ac voltage across the load}}{\text{dc voltage across the load}} = \sqrt{\left(\frac{V_{rms}}{V}\right)^2 - 1} = 0.185 \tag{11.98}$$

The diode conduction angle is 2π/n, namely ⅔π. The peak diode reverse voltage is given by the maximum voltage between any two phases, √3√2 V = √6 V.

From equations (11.80), (11.81), and (11.82), for a constant output current, $\bar{I}_o = I_{o\,rms}$, the mean diode current is

$$\bar{I}_D = \tfrac{1}{n}\bar{I}_o = \tfrac{1}{3}\bar{I}_o \qquad \text{(A)} \tag{11.99}$$

and the rms diode current is

$$I_D = \tfrac{1}{\sqrt{n}} I_{o\,rms} \approx \tfrac{1}{\sqrt{n}}\bar{I}_o = \tfrac{1}{\sqrt{3}}\bar{I}_o \qquad \text{(A)} \tag{11.100}$$

The diode current form factor is

$$FF_{ID} = I_D / \bar{I}_D = \sqrt{3} \tag{11.101}$$

The input displacement factor cosΦ is unity and the input power factor (and displacement factor), assuming diode square currents, is

$$pf = \frac{V_o I_o}{3V_s I_{irms}} = \frac{\frac{3\sqrt{6}}{2\pi}V_s\, I_o}{3V_s \frac{I_o}{\sqrt{3}}} = \frac{3}{\sqrt{2}\,\pi} \tag{11.102}$$

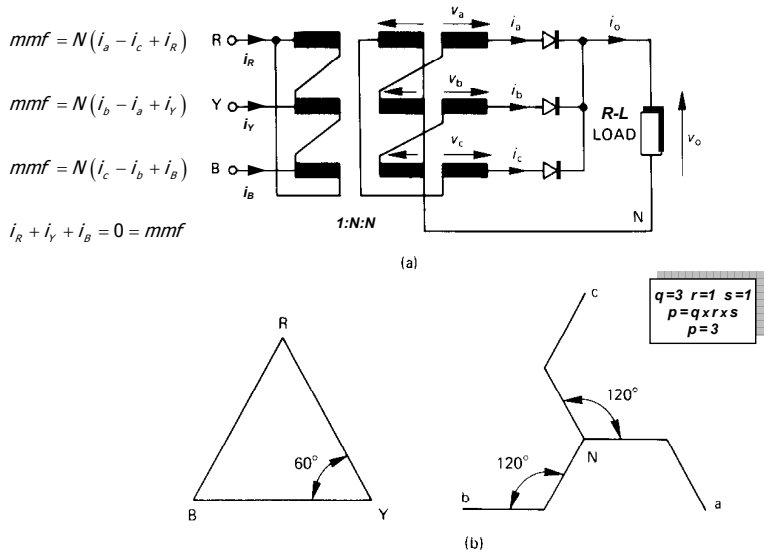


Figure 11.11. *Three-phase zig-zag interconnected star winding, with three windings per limb, 1:N:N: (a) transformer connection showing zero dc mmf in each limb (phase) and (b) phasor diagram of transformer primary and secondary voltages.*

If neutral is available, a transformer is not necessary. Then the full load current is returned via the neutral supply. This neutral current is generally not acceptable other than at low power levels. The simple delta-star connection of the supply in figure 11.10a is not appropriate since the unidirectional current in each phase is transferred from the supply to the transformer. This may result in increased magnetising current and iron losses if dc magnetisation occurs. As discussed in section 11.3.5, this problem is avoided in most cases by the special interconnected star winding, called zig-zag, shown in figure 11.11a and discussed in section 11.3.7. Each transformer limb has two equal voltage secondaries which are connected such that the magnetising forces balance. The resultant phasor diagram is shown in figure 11.11b. 15% more turns are needed than with a star connection. This transformer mmf problem resulting from half-wave rectification is considered in section 11.3.

As the number of phases increases, the windings become less utilised per cycle since the diode conduction angle decreases, from π for a single-phase circuit, to ⅔π for the three-phase case

11.2.2 Three-phase full-wave rectifier circuit with an inductive R-L load

Figure 11.12a shows a three-phase full-wave rectifier circuit where no neutral is necessary and it will be seen that two series diodes are always conducting. One diode (one of D_1, D_3, or D_5, at the highest potential) can be considered as being in the feed circuit, while the other (one of D_2, D_4, or D_6, at the lowest potential) is in the return circuit. As such, the line-to-line voltage is impressed across the load. Given no two series connected diodes conduct simultaneously, there are six possible diode pair combinations. The rectifier circuit waveforms in figure 11.12b show that the load ripple frequency is six times the supply. Each diode conducts for ⅔π and experiences a reverse voltage of the peak line voltage, $\sqrt{2}\,V_L$.

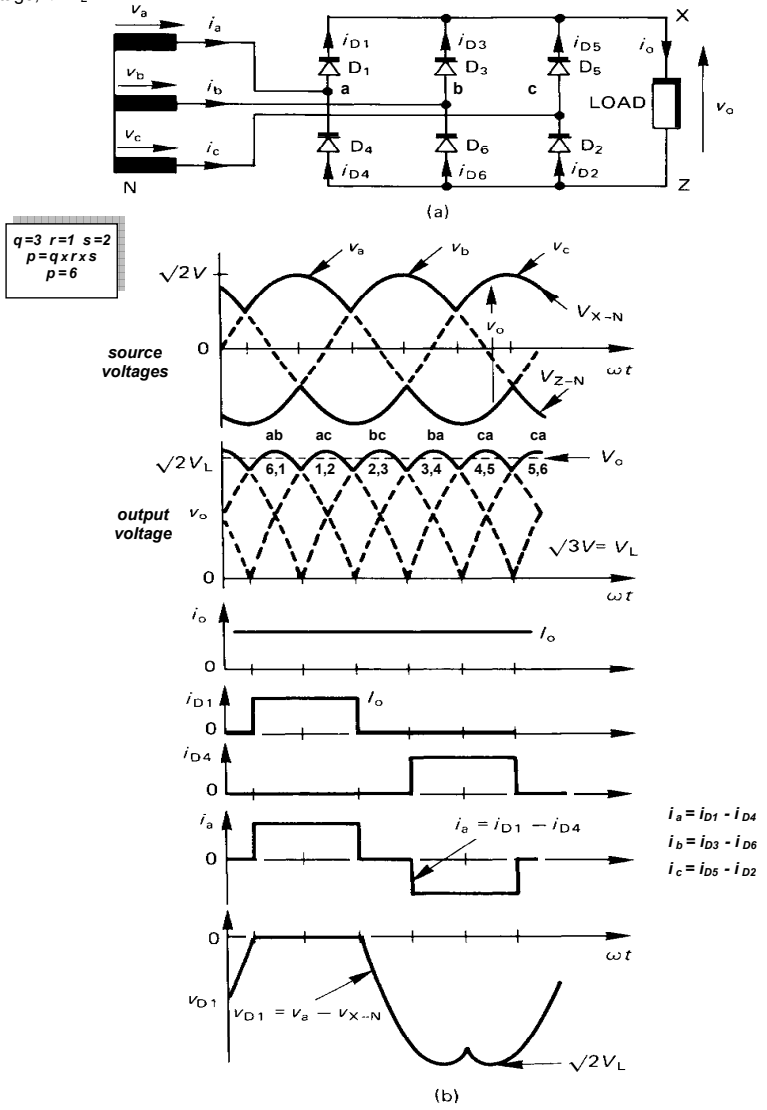


Figure 11.12. *Three-phase full-wave bridge rectifier: (a) circuit connection and (b) voltage and current waveforms.*

The mean load voltage is given by twice equation (11.95), that is

$$V_o = I_o R = \tfrac{1}{\pi/3}\int_{\pi/3}^{2\pi/3} \sqrt{2}\,V_L \sin\omega t\, d\omega t \qquad \text{(V)}$$
$$= \sqrt{2}V\frac{\sqrt{3}}{\pi/3} = \frac{3}{\pi}\sqrt{2}\,V_L = 1.35V_L = 2.34V \qquad (11.103)$$

where V_L is the line-to-line rms voltage ($V_L = \sqrt{3}V$).

Generally the peak-to-peak ripple voltage for *n*-phases is $\sqrt{2}V - \sqrt{2}\cos\pi/n$. (see table 11.4)

The critical load inductance (see figure 12.12) for continuous load current, is $L_{critical} = R/\tfrac{1}{2}\omega_\times\, p\left(p^2-1\right)$.

The output harmonics of a *p*-pulse voltage output are

$$V_{an} = -1^{p/2} \times \frac{\sqrt{2}V}{\pi/p}\sin\pi/p\,\frac{2}{\left[n^2-1\right]}$$
$$= -1^{p/2} \times V_o \times \frac{2}{\left[n^2-1\right]} \qquad (11.104)$$

where $n = mp$ and $m = 1,\ 2,\ 3,\ \ldots$ and V_o is the mean output voltage given by equation (11.94).

The output voltage harmonics for $p = 6$ are given by

$$V_{on} = \frac{6\hat{V}_L}{\pi\left(n^2-1\right)} \qquad (11.105)$$

for $n = 6, 12, 18, ..$

The rms output voltage is given by

$$V_{rms} = \left(\frac{1}{2\pi/6}\int_{\pi/3}^{2\pi/3} \sqrt{2}V_L \sin^2\omega t\, d\omega t\right)^{1/2}$$
$$= V_L\sqrt{1+\frac{3\sqrt{3}}{2\pi}} = 1.352V_L \qquad (11.106)$$

Generally, for a *p*-pulse rectifier output, the rms output voltage is

$$V_{rms} = V_L\sqrt{1+\frac{p}{2\pi}\sin 2\pi/p} \qquad (11.107)$$

The load voltage form factor = 1.352/1.35 = 1.001 and the ripple factor = √form factor -1 = 0.06.

11.2.2i Three-phase full-wave bridge rectifier circuit with continuous load current

If it is assumed that the load inductance is large, then (even with a load back emf), continuous load current flows and the dominate load current harmonic is due to the sixth harmonic current, that is let $I_{o,ac} = I_{o,6}$. By neglecting the higher order harmonics, the various circuit currents and voltages can be readily obtained as shown in table 11.3. From equations (11.103) and (11.105) the output voltage is given by

$$v_o(\omega t) = \overline{V}_o + V_{o,6}\cos 6\omega t$$
$$= \frac{3}{\pi}\sqrt{2}V_L + \frac{3}{\pi}\sqrt{2}V_L\frac{2}{\left(n^2-1\right)}\cos n\omega t \quad \text{for } n = 6$$
$$= \frac{3}{\pi}\sqrt{2}V_L + \frac{3}{\pi}\sqrt{2}V_L \times \frac{2}{35}\cos 2\omega t \qquad (11.108)$$
$$= 1.35V_L + 0.077V_L\cos 2\omega t$$

The fundamental voltage, hence current, V_o/R, is therefore much larger than the sixth harmonic current, $V_{o,6}/Z_6$, that is $\overline{I}_o > \hat{I}_{o,6}$. The load and supply ac currents are $I_{o,ac} = I_{s,ac} = I_{o,6}$. The output and supply rms currents are

$$I_{o,rms} = I_{s,rms} = \sqrt{\overline{I}_o^2 + I_{o,ac}^2} = \sqrt{\overline{I}_o^2 + I_{o,6}^2} \qquad (11.109)$$

and the power delivered to resistance *R* in the load is

$$P_R = I_{o,rms}^2 R \qquad (11.110)$$

11.2.2ii Three-phase full-wave bridge circuit with highly inductive load – constant load current

For a highly inductive load, that is a constant load current:

- the mean diode current is

$$\bar{I}_D = \tfrac{1}{3}\bar{I}_o = \tfrac{1}{3}\bar{I}_o \quad \text{(A)} \tag{11.111}$$

- and the rms diode current is

$$I_{D\,rms} = \tfrac{1}{\sqrt{3}} I_{o\,rms} \approx \tfrac{1}{\sqrt{3}}\bar{I}_o = \tfrac{1}{\sqrt{3}}\bar{I}_o \quad \text{(A)} \tag{11.112}$$

- and the power factor (distortion factor) for a constant load current is

$$pf = \frac{3}{\pi} = 0.955 \tag{11.113}$$

The rms input line currents are

$$I_{L\,rms} = \sqrt{\frac{2}{3}}\, I_{o\,rms} \tag{11.114}$$

The diode current form factor is

$$FF_{ID} = I_{D\,rms} / \bar{I}_D = \sqrt{3} \tag{11.115}$$

The diode current ripple factor is

$$RF_{ID} = \sqrt{FF_{ID}^2 - 1} = \sqrt{2} \tag{11.116}$$

A phase voltage and current are given by

$$v_a = \sqrt{2}V \sin\omega t \tag{11.117}$$

$$i_a = \frac{2\sqrt{3}}{\pi}\bar{I}_o\left[\sin\omega t + \frac{\sin(n-1)\omega t}{n-1} + \frac{\sin(n+1)\omega t}{n+1}\right] \qquad n = 6, 12, 18, .. \tag{11.118}$$

with phases b and c shifted by ⅔π. That is substitute ωt in equations (11.117) and (11.118) with $\omega t \pm \tfrac{2}{3}\pi$.

Each load current harmonic *n* produces harmonics *n+1* and *n-1* on the input current.

The total load instantaneous power is given by

$$p(\omega t) = 3 \times \sqrt{2}V\bar{I}_o \times \left(\tfrac{1}{2} - \frac{\cos n\omega t}{n^2 - 1}\right) \tag{11.119}$$

The supply apparent power is

$$S = \sqrt{3}V_L I_{s\,rms} \tag{11.120}$$

while the ac power, in terms of apparent ac resistance, is

$$P_{ac} = \frac{3 \times V^2}{R_{ac}} \tag{11.121}$$

Using the output voltage from equation (11.103), the output power is

$$P_{dc} = \left(\sqrt{2}V\frac{\sqrt{3}}{\pi/3}\right)^2 / R_{dc} \tag{11.122}$$

Since $P_{ac} = P_{dc}$, then $R_{dc} = 2\left(\frac{9}{\pi^2}\right)R_{ac} \approx 2R_{ac}$.

11.2.2iii Three-phase full-wave bridge circuit with highly inductive load with an EMF source

With continuous load current, the output voltage and input characteristics are unaffected by a load back emf, with the average and rms output voltages given by equations (11.103) and (11.106) respectively. The input power factor and distortion factor are 3/π, as per equation (11.113).

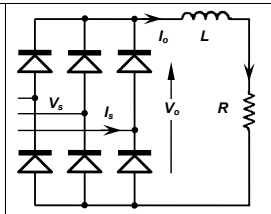
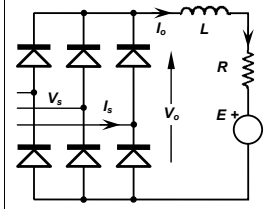
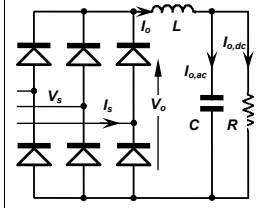
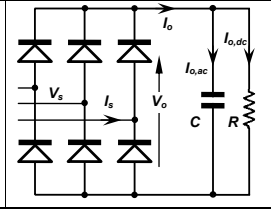
The output, that is, load current, is found from

$$L\frac{di_o}{dt} + Ri_o + E = \sqrt{2}V_s \sin\omega t \qquad \tfrac{1}{3}\pi \le \omega t \le \tfrac{2}{3}\pi$$

$$i_o(t) = I_o e^{-\frac{\omega t - \frac{1}{3}\pi}{\tan\theta}} + \sqrt{2}V_s\left[\sin(\omega t - \theta) - \frac{\sin\alpha}{\cos\theta}\right] \tag{11.123}$$

where $\tan\theta = \frac{\omega L}{R}$; $Z = \sqrt{R^2 + \omega^2 L^2}$; $\sin\alpha = \frac{E}{\sqrt{2}V_s}$; and $I_o = \frac{\sqrt{2}V_s}{Z}\frac{2\sin\theta}{1 - e^{\frac{-\frac{2}{3}\pi}{\tan\theta}}}$

Table 11.3: Three-phase full-wave uncontrolled rectifier circuits

load	Full-wave rectifier circuit	6th harmonic current $I_{o,6}$ (A)	average output current $\bar{I}_o$ (A)	output power P_R+P_E (W)
(a) *R-L* *see section 11.2.2i*		$\frac{V_{o,6}}{\sqrt{R^2+(6\omega L)^2}}$	$\frac{\bar{V}_o}{R}$	$I_{o,rms}^2 R$
(b) *11.2.2iii* *R-L-E*		$\frac{V_{o,6}}{\sqrt{R^2+(6\omega L)^2}}$	$\frac{\bar{V}_o - E}{R}$	$I_{o,rms}^2 R + \bar{I}_o E$
(c) *R-L-C*		$\frac{V_{o,6}}{6\omega L}$	$\frac{\bar{V}_o}{R}$	$I_{o,rms}^2 R = \bar{I}_o^2 R$
(d) *11.2.2iv* *R-C*			$\frac{\bar{V}_o}{R}$	$I_{o,rms}^2 R$

11.2.2iv Three-phase full-wave bridge circuit with capacitively filtered load resistance

Part d in Table 11.1 shows a three-phase full-wave rectifier circuit with an R-C load.

Interval $\alpha \le \omega t \le \beta$

In the interval $\alpha \le \omega t \le \beta$, two diodes are conducting connecting the supply voltage across the load. The input current provides both the resistive load and the output filter capacitor across the load.

$$i_s = i_o + i_C = \frac{v_o}{R} + C\frac{dv_o}{dt} \text{ where } v_o = V_s = \sqrt{2}V_s \sin\omega t$$

That is

$$\begin{aligned} i_s &= i_o + i_C \\ i_s(t) &= \frac{\sqrt{2}V_s \sin\omega t}{R} + \sqrt{2}V_s\,\omega C\cos\omega t \\ &= \frac{\sqrt{2}V_s}{R}\sqrt{1+\omega^2C^2R^2}\cos(\omega t - \phi) = \frac{\sqrt{2}V_s}{R\cos\phi}\cos(\omega t - \phi) \end{aligned} \tag{11.124}$$

where $\tan\phi = \dfrac{1}{\omega RC}$ and $\beta = \tfrac{1}{2}\pi + \phi$

Interval $\beta \le \omega t \le \alpha + \pi/3$

In the interval $\beta \le \omega t \le \alpha + \pi/3$, the bridge diodes are all reverse biased, isolating the source from the load (discontinuous input current), and the load current is provided from the output capacitor.

$$i_s = i_o + i_C = 0 = \frac{v_o}{R} + C\frac{dv_o}{dt}$$

In satisfying a boundary condition yields

$$\begin{aligned} v_o(t) = v_C(t) = v_R(t) = \\ = \sqrt{2}V_s \frac{\omega RC}{\sqrt{1+\omega^2R^2C^2}} e^{\frac{-(\omega t-\beta)}{\omega RC}} = \sqrt{2}V_s \frac{R\cos\phi}{X_C} \times e^{\frac{-(\omega t-\beta)}{\omega RC}} \\ = i_o R \end{aligned} \qquad (11.125)$$

Equating the two output voltage expressions, equations (11.124) and (11.125), at the boundary $\omega t = \alpha + \pi/3$ yields an equation for determining α iteratively.

$$\sin\alpha = \frac{\omega RC}{\sqrt{1+\omega^2R^2C^2}} e^{\frac{-\left(\alpha - \frac{1}{6}\pi - \tan^{-1}\frac{1}{\omega RC}\right)}{\omega RC}}$$

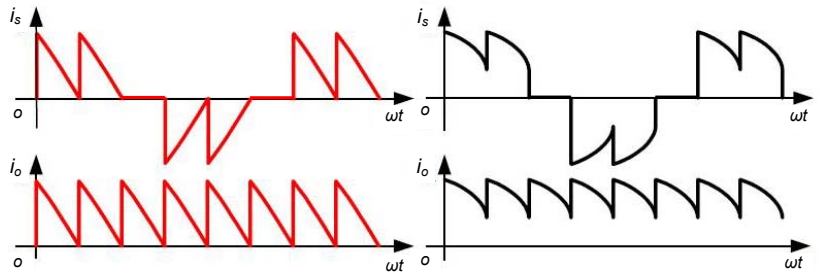


Figure 11.13. *Three-phase full-wave bridge rectifier a capacitive output filter: (a) verge of discontinuous conduction and (b) continuous current conduction.*

Example 11.7: ***Three-phase full-wave rectifier***

The full-wave three-phase dc rectifier in figure 11.12a has a three-phase 415V 50Hz source (240V phase), and a 10Ω, 50mH, series load. During the problem solution, verify that the only harmonic that need be considered is the sixth.

Determine

i. the average output voltage and current
ii. the rms load voltage and the ac output voltage
iii. the rms load current hence power dissipated and supply power factor
iv. the load power percentage error in assuming a constant load current
v. the diode average and rms current requirements

Solution

i. From equation (11.103) the average output voltage and current are

$$V_o = I_o R = 1.35V_L = 1.35 \times 415\text{V} = 560.45\text{V}$$

$$I_o = \frac{V_o}{R} = \frac{560.45V}{10\Omega} = 56.045\text{A}$$

ii. The rms load voltage is given by equation (11.106)

$$V_{rms} = 1.352V_L = 1.352 \times 415\text{V} = 560.94\text{V}$$

The ac component across the load is

$$V_{ac} = \sqrt{V_{rms}^2 - V_o^2} = \sqrt{560.94\text{V}^2 - 560.447\text{V}^2} = 23.52\text{V}$$

iii. The rms load current is calculated from the harmonic currents, which are calculated from the harmonic voltages given by equation (11.105).

harmonic n	$V_n = \dfrac{6\hat{V}_L}{\pi(n^2-1)}$	$Z_n = \sqrt{R^2+(n\omega L)^2}$	$I_n = \dfrac{V_n}{Z_n}$	$\tfrac{1}{2}I_n^2$
0	***(560.45)***	10.00	56.04	***(3141.01)***
6	32.03	94.78	0.34	0.06
12	7.84	188.76	0.04	0.00
Note the 12th harmonic current is not significant			$I_o^2 + \sum\tfrac{1}{2}I_n^2 =$	3141.07

The rms load current is

$$I_{rms} = \sqrt{I_o^2 + \sum\tfrac{1}{2}I_n^2} = \sqrt{3141.07} = 56.05\text{A}$$

The power absorbed by the 10Ω load resistor is

$$P_L = I_{rms}^2 R = 56.05\text{A}^2 \times 10\Omega = 31410.7\text{W}$$

The supply power factor is

$$pf = \frac{P_L}{V_{rms}I_{rms}} = \frac{P_L}{\sqrt{3}V_L I_L} = \frac{31410.7\text{W}}{\sqrt{3}\times 415\text{V} \times \sqrt{\frac{2}{3}} \times 56.05\text{A}} = 0.955$$

This power factor of 0.955 is as predicted by equation (11.113), $\frac{3}{\pi}$, for a constant current load.

iv. The percentage output power error in assuming the load current is constant is given by

$$1 - \frac{\tilde{P}_L}{P_L} = 1 - \frac{I_o^2 R}{I_{rms}^2 R} = 1 - \frac{56.045\text{A}^2 \times 10\Omega}{56.05\text{A}^2 \times 10\Omega} = 1 - \frac{31410.1\text{W}}{31410.7\text{W}} \simeq 0\%$$

v. The diode average and rms currents are given by equations (11.111) and (11.112)

$$\bar{I}_D = \tfrac{1}{3}\bar{I}_o = \tfrac{1}{3} \times 56.045 = 18.7\text{A}$$

$$I_{D\,rms} = \tfrac{1}{\sqrt{3}}I_{o\,rms} = \tfrac{1}{\sqrt{3}} \times 56.05 = 23.4\text{A}$$

Example 11.8: ***Rectifier average load voltage***

Derive a general expression for the average load voltage of an *p*-pulse rectifier.

Solution

Figure 11.14 defines the general output voltage waveform where *p* is the output pulse number per cycle of the ac supply. From the output voltage waveform

$$V_o = \frac{1}{2\pi/p}\int_{-\pi/n}^{\pi/n} \sqrt{2}V\cos\omega t\, d\omega t$$

$$= \frac{\sqrt{2}V}{2\pi/p}\left(\sin(\pi/p) - \sin(-\pi/p)\right) = \frac{\sqrt{2}V}{2\pi/p} 2\sin(\pi/p)$$

$$V_o = \frac{\sqrt{2}V}{\pi/p}\sin(\pi/p) \qquad \text{(V)}$$

where

for *p* = 2 for the single-phase (*n* = 1) full-wave rectifier in figure 11.7.
for *p* = 3 for the three-phase (*n* = 3) half-wave rectifier in figure 11.10.
for *p* = 6 for the three-phase (*n* = 3) full-wave rectifier in figure 11.12.

♣

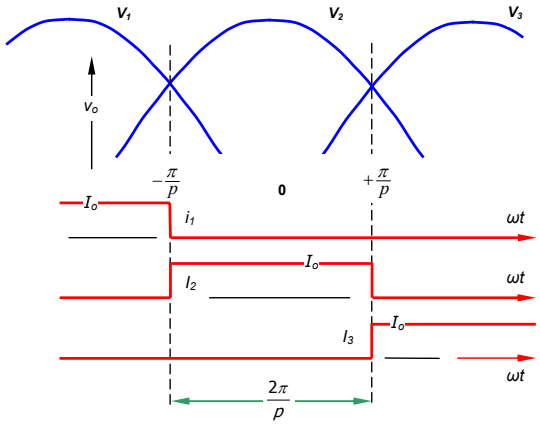


Figure 11.14. *A half-wave n-phase uncontrolled rectifier: output voltage and current waveforms.*

11.3 DC *MMF*s in converter transformers

Half-wave rectification – whether controlled, semi-controlled or uncontrolled, is notorious for producing a dc *mmf* in transformers and triplen harmonics in the ac supply neutral of three-phase circuits. Generally, a transformer based solution can minimise the problem. In order to simplify the underlying concepts, a constant dc load current I_o is assumed, that is, the load inductance is assumed infinite. The transformer is assumed linear, no load excitation is ignored, and the ac supply is assumed sinusoidal. Independent of the transformer and its winding connection, the average output voltage from a rectifier, when the rectifier bridge input rms voltage is V_B and there are q pulses in the output, is given by

$$V_o = \frac{\hat{V}_B}{2\pi/q}\int_{-\pi/q}^{\pi/q}\cos\omega t\, d\omega t = \hat{V}_B\frac{\sin\pi/q}{\pi/q} \qquad (11.126)$$

The rectifier bridge rms voltage output is dominated by the dc component and is given by

$$V_{o\,rms} = \frac{q}{2\pi}\int_{-\frac{\pi}{q}}^{+\frac{\pi}{q}} 2V_B^2\cos^2(\omega t)\, d\omega t = V_B\sqrt{1+\frac{q}{2\pi}\sin\frac{2\pi}{q}} \qquad (11.127)$$

The Fourier expression for the output voltage, which is also dominated by the dc component, is

$$v_o(\omega t) = V_o + V_o\sum_{k=1}^{\infty}\frac{2(-1)^{k+1}}{k^2n^2-1}\cos kn\omega t \qquad (11.128)$$

Table 11.4 summarizes the various rectifier characteristics that are independent of the transformer winding configuration.

Table 11.4: Rectifier characteristics with *q* phases (see section 11.6)

q phases	Parallel connected secondary windings		Series connected secondary windings
	Star, thus neutral always exists		Polygon, hence no neutral
	$v_1=\sqrt{2}V\sin[\omega t]$ $v_2=\sqrt{2}V\sin\left[\omega t-\frac{2\pi}{q}\right]$ . . $v_q=\sqrt{2}V\sin\left[\omega t-(q-1)\frac{2\pi}{q}\right]$		
	Half-wave	Full-wave	
V_o	$\frac{q}{\pi}\sqrt{2}V\sin\frac{\pi}{q}$	$2\frac{q}{\pi}\sqrt{2}V\sin\frac{\pi}{q}$	$\frac{q}{\pi}\sqrt{2}V$
Load harmonics	$n=q$	$n=q$ q even $n=2q$ q odd	$n=q$ q even $n=2q$ q odd
$\hat{V}_o$ $\check{V}_o$	$\sqrt{2}V$ $\sqrt{2}V\cos\frac{\pi}{q}$	$2\sqrt{2}V\cos\frac{\pi}{2q}$	
$\hat{V}_{DR}$	$\sqrt{2}V$ q even $\sqrt{2}V\cos\frac{\pi}{2q}$ q odd	$2\sqrt{2}V$ q even $2\sqrt{2}V\cos\frac{\pi}{2q}$ q odd	$\frac{\sqrt{2}V}{\sin\frac{\pi}{q}}$ q even $\frac{\sqrt{2}V}{2\sin\frac{\pi}{2q}}$ q odd
N° of diodes	q diodes	$2q$ diodes	$2q$ diodes
$\bar{I}_D$ $I_{D\,rms}$	$\bar{I}_D=\frac{I_o}{q}$ $I_{D\,rms}=\frac{I_o}{\sqrt{q}}$		
I_s	$I_s=I_o\sqrt{\frac{1}{q}}$	$I_s=I_o\sqrt{\frac{2}{q}}$	$I_s=½I_o$ q even $I_s=½I_o\frac{\sqrt{q^2-1}}{q}$ q odd
$P_o=V_oI_o$ $S=qV_sI_s$ $pf_{load}=\frac{P_o}{S}$	$\frac{\sqrt{2q}}{\pi}\sin\frac{\pi}{q}$	$\frac{2\sqrt{q}}{\pi}\sin\frac{\pi}{q}$	$\frac{2\sqrt{2}}{\pi}$ q even $\frac{2\sqrt{2}}{\pi}\frac{q}{\sqrt{q^2-1}}$ q odd

11.3.1 Effect of multiple coils on multiple limb transformers

The transformer for a single-phase two-pulse half-wave rectifier has three windings, a primary and two secondary windings as shown in figure 11.15. Two possible transformer core and winding configurations are shown, namely shell and core. In each case the winding turns ratios are identical, as is the load voltage and current, but the physical transformer limb arrangements are different. One transformer, figure 11.15a, has three limbs (made up from E and I laminations), while the second, figure 11.15b, is made from a circular core (shown as a square core). The reason for the two possibilities is related to the fact that the circular core can use a single strip of wound cold-rolled grain-orientated silicon steel as lamination material. Such steels offer better magnetic properties than the non-oriented steel that must be used for E core laminations. Single-phase toroidal core transformers are attractive because of the reduced size and weight but manufacturers do not highlight their inherent limitation and susceptibility to dc flux biasing, particularly in half-wave type applications. Although the solution is simple, the advantageous features of the toroidal transformer are lost, as will be shown.

i. The E-I three-limb transformer (shell)

The key feature of the three-limb shell is that the three windings are on the centre limb, as shown in figure 11.15a. The area of each outer limb is half that of the central limb. Assuming a constant load current I_o and equal secondary turns, N_s, excitation of only the central limb yields the following *mmf* equation

$$mmf = i_p N_p + i_{s1} N_s - i_{s2} N_s \tag{11.129}$$

Thus the primary current i_p is

$$i_p = \frac{N_s}{N_p}\left(i_{s2} - i_{s1}\right) + \frac{mmf}{N_p} \tag{11.130}$$

From the waveforms in figure 11.15a, since $i_{s2} - i_{s1}$ is alternating, an average primary current of zero in equation (11.130) can only be satisfied by $mmf = 0$.

The various transformer voltages and currents are

$$I_{s1} = I_{s2} = I_s = \frac{I_o}{\sqrt{2}}$$

$$I_p = I_o \frac{N_s}{N_p} \tag{11.131}$$

$$V_{s1} = V_{s2} = V_s = V_p \frac{N_s}{N_p} = \frac{\pi}{2\sqrt{2}} V_o$$

Therefore the transformer input, output and average VA ratings are

$$S_s = V_{s1} I_{s1} + V_{s2} I_{s2} = \sqrt{2}\,\frac{N_s}{N_p} V_p I_o \left(= \frac{\pi}{2} P_o = 1.57 P_o\right)$$

$$S_p = V_p I_p = \frac{N_s}{N_p} V_p I_o \left(= \frac{\pi}{2\sqrt{2}} P_o = 1.11 P_o\right) \tag{11.132}$$

$$\overline{S} = ½\left(S_s + S_p\right) = \frac{N_p}{N_s} V_p I_o \frac{1+\sqrt{2}}{2}$$

The average output voltage, hence output power, are

$$V_o = \frac{2\sqrt{2}}{\pi} V_s = \frac{2\sqrt{2}}{\pi} \frac{N_s}{N_p} V_p = 0.9 \frac{N_s}{N_p} V_p \tag{11.133}$$

$$P_o = I_o V_o$$

Thus

$$\overline{S} = \frac{1+\sqrt{2}}{4\sqrt{2}} \pi P_o = 1.34 P_o \tag{11.134}$$

Since the transformer primary current is the line current, the supply power factor is

$$pf = \frac{P_o}{\overline{S}} = \frac{V_o I_o}{V_p I_p} = \frac{\frac{2}{\pi} V_s I_o}{\frac{N_p}{N_s} V_s \frac{N_s}{N_p} I_o} = \frac{2\sqrt{2}}{\pi} = 0.9 \tag{11.135}$$

ii. The two-limb strip core transformer

Figure 11.15b shows the windings equally split on each transformer leg. In practice the windings can all be on one leg and the primary is one coil, but separation as shown allows visual *mmf* analysis. The load and diode currents and voltages are the same as for the E-I core arrangement, as seen in the waveforms in figure 11.15b. The *mmf* analysis necessary to assess the primary currents and core flux, is based on analysing each limb.

$$mmf_1 = -i_p ½N_p + i_{s1} N_s$$

$$mmf_2 = +i_p ½N_p + i_{s2} N_s \tag{11.136}$$

$$mmf_1 = mmf_2 = mmf$$

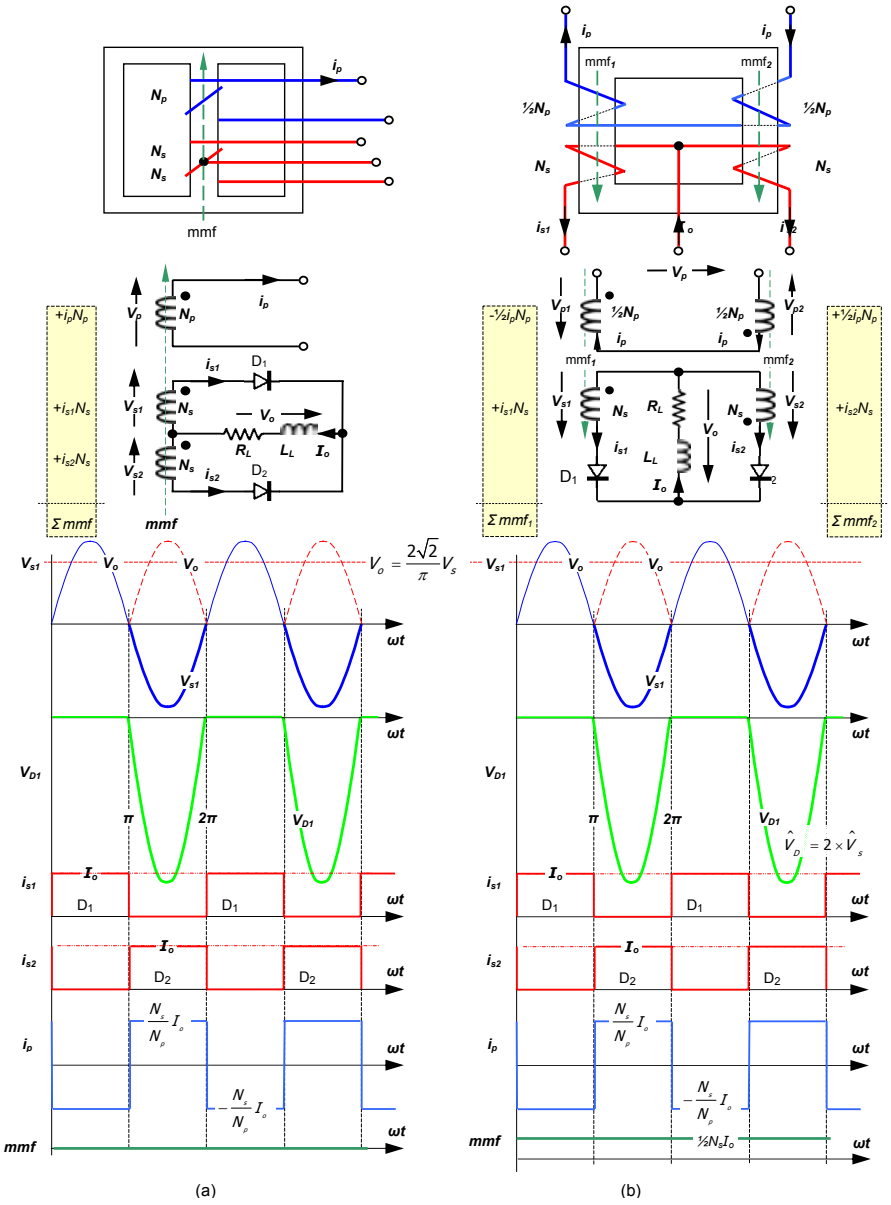


Figure 11.15. *Single-phase transformer core and winding arrangements:*
(a) E-I core with zero dc mmf bias and (b) square/circular core with dc mmf bias.

These equations yield

$$mmf = N_s \tfrac{1}{2}(i_{s1} + i_{s2}) = N_s \tfrac{1}{2} I_o$$
$$i_p = \frac{N_s}{N_p}(i_{s1} - i_{s2}) \qquad (11.137)$$

These two equations are used every ac half cycle to obtain the plots in figure 11.15b. It will be noticed that the core has a magnetic *mmf* bias of $\tfrac{1}{2}N_sI_o$ associated with the half-wave rectification process.

The various transformer ratings are

$$I_{s1} = I_{s2} = I_s = \frac{I_o}{\sqrt{2}} \qquad I_p = I_o \frac{N_s}{N_p}$$
$$V_{p1} = V_{p2} = \tfrac{1}{2}V_p \qquad V_{s1} = V_{s2} = V_s = \tfrac{1}{2}V_p \frac{N_s}{\tfrac{1}{2}N_p} = V_p \frac{N_s}{N_p} \qquad (11.138)$$
$$V_p = \frac{N_p}{N_s}\frac{\pi}{2\sqrt{2}} V_o$$

Therefore the transformer VA ratings are

$$S_s = V_{s1}I_{s1} + V_{s2}I_{s2} = \sqrt{2}\frac{N_s}{N_p}V_pI_o \quad \left(= \frac{\pi}{2}P_o = 1.57P_o\right)$$
$$S_p = V_{p1}I_p + V_{p2}I_p = \frac{N_s}{N_p}V_pI_o \quad \left(= \frac{\pi}{2\sqrt{2}}P_o = 1.11P_o\right) \qquad (11.139)$$
$$\overline{S} = \tfrac{1}{2}(S_s + S_p) = \frac{N_s}{N_p}V_pI_o\frac{1+\sqrt{2}}{2}$$

The average output voltage, hence output power, are

$$V_o = \frac{2\sqrt{2}}{\pi}V_s = \frac{2\sqrt{2}}{\pi}\frac{N_s}{\tfrac{1}{2}N_p}\tfrac{1}{2}V_p = \frac{2\sqrt{2}}{\pi}\frac{N_s}{N_p}V_p \qquad (11.140)$$
$$P_o = I_oV_o$$

Thus

$$\overline{S} = \pi\frac{1+\sqrt{2}}{4\sqrt{2}}P_o = 1.34P_o \qquad (11.141)$$

and the supply power factor is $pf = P_o / \overline{S} = 0.9$.

The interpretation for equations (11.139) and (11.141) (and equations (11.132) and (11.134)) is that the transformer has to be oversized by 11% on the primary side and 57% on the secondary. From equation (11.141), in terms of the average VA, the transformer needs to be 34% larger than that implied by the rated dc load power. Further, the secondary is rated higher than the primary because of a dc component in the secondary. This core saturation aspect requires special attention when dimensioning the core size. Additionally, a component of the over rating requirement is due to circulating harmonics that do not contribute to real power output. This component is particularly relevant in three-phase delta primary or secondary connections when cophasal triplens circulate. This discussion on apparent power aspects is relevant to all the transformer connections considered. Generally the higher the phase number the better the transformer core utilisation, but the poor the secondary winding and rectifying diode utilisation since the percentage current conduction decreases with increased pulse number.

The fundamental ripple in the output voltage, at twice the supply frequency, is $\tfrac{2}{3}V_o$.

The two cores give the same rated transformer apparent power and supply power factor, but importantly, undesirably, the toroidal core suffers an *mmf* magnetic bias.
In each core case each diode conducts for 180° and

$$\overline{I}_D = \tfrac{1}{2}I_o \qquad I_{Drms} = \frac{I_o}{\sqrt{2}} \qquad \hat{V}_D = 2\sqrt{2}\frac{N_s}{N_p}V_p \qquad (11.142)$$

With a purely resistive load, a full-wave rectifier with a centre-tapped primary gives

$$\overline{I}_D = \tfrac{1}{2}I_o \qquad I_{Drms} = \tfrac{1}{4}\pi I_o \qquad S_s = 1.75P_o \qquad S_p = 1.23P_o \qquad \overline{S} = 1.49P_o \qquad (11.143)$$

11.3.2 Single-phase toroidal core mmf imbalance cancellation – zig-zag winding

In figure 11.16, each limb of the core has an extra secondary winding, of the same number of turns, N_s. *MMF* analysis of each limb in figure 11.16 yields

$$\text{limb 1:} \quad mmf_o = -i_pN_p - i_{s2}N_s + i_{s1}N_s$$
$$\text{limb 2:} \quad mmf_o = i_pN_p + i_{s2}N_s - i_{s1}N_s \qquad (11.144)$$

Adding the two mmf equations gives $mmf_o = 0$ and the resulting alternating primary current is given by

$$i_p = \frac{N_s}{N_p}(i_{s1} - i_{s2}) \qquad (11.145)$$

The transformer apparent and real power are rated by the same equation as for the previous winding arrangements, namely

$$\overline{S} = \tfrac{1}{2}(S_p + S_s) = \tfrac{1}{2}\left(\frac{\pi}{2\sqrt{2}}P_o + \frac{\pi}{2}P_o\right) = 1.34P_o \qquad (11.146)$$
$$\text{where } P_o = V_oI_o \text{ and } V_o = \frac{N_s}{N_p}\frac{2\sqrt{2}}{\pi}V_p$$

Since the transformer primary current is the ac line current, the supply power factor is $pf = P_o / \overline{S} = 0.9$.
The general rule to avoid any core dc *mmf* is, each core leg must be effectively excited by a net alternating current.

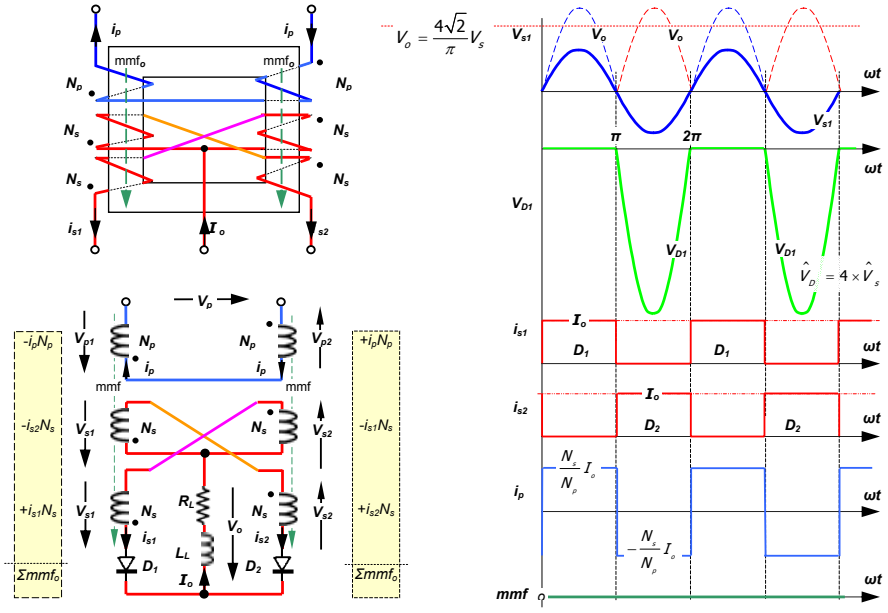


Figure 11.16. *Single-phase zig-zag transformer core and winding arrangement using square/circular core with zero dc mmf bias.*

11.3.3 Single-phase transformer connection, with full-wave rectification

The secondary current is ac with a zero average , thus no core mmf bias occurs.
The average output voltage and peak diode reverse voltage, in terms of the transformer secondary rms voltage, are

$$V_o = \frac{2\sqrt{2}}{\pi} V_s \qquad V_{Dr} = \sqrt{2}\, V_s \tag{11.147}$$

The rms output voltage is the bridge input rms voltage:

$$V_{o\,rms} = V_s \tag{11.148}$$

The various harmonic currents are

$$I_{s1} = \frac{2\sqrt{2}}{\pi} I_o = 0.9 I_o \qquad I_{sh} = \frac{I_{s1}}{h} \quad \text{for } h \text{ odd} \tag{11.149}$$

The power factor angle of the fundamental is unity, while the THD is 48.43%.

The transformer primary and secondary apparent powers are

$$S_p = S_s = \frac{\pi}{2\sqrt{2}} P_o = 1.11 P_o \tag{11.150}$$

The transformer average VAr rating is

$$\overline{S} = \frac{\pi}{2\sqrt{2}} P_o \tag{11.151}$$

Since the line current is the primary current, the supply power factor is

$$pf = \frac{P_o}{\overline{S}} = \frac{2\sqrt{2}}{\pi} \tag{11.152}$$

The fundamental ripple in the output voltage, at twice the supply frequency, is ⅔V_o.

With a purely resistive load, a full-wave bridge rectifier gives

$$I_o = \frac{V_o}{R} \qquad I_{o\,rms} = \frac{V_s}{R}$$
$$\overline{I}_D = ½ I_o \qquad I_{D\,rms} = ¼\pi I_o \qquad S_s = 1.23 P_o \qquad S_p = 1.23 P_o \qquad \overline{S} = 1.23 P_o \tag{11.153}$$

11.3.4 Three-phase transformer connections

Basic three-phase transformers can have a combination of star (wye) and delta, primary and secondary winding arrangements.

i. Y - y (WYE-wye) is avoided due to imbalance and third harmonic problems, but with an extra delta winding, triplen problems can be minimised. The arrangement is used to interconnect high voltage networks, 240kV/345kV or when two neutrals are needed for grounding.

ii. Y - δ (WYE-delta) is commonly used for step-down voltage applications.

iii. Δ – δ (DELTA-delta) is used in 11kV medium voltage applications where neither primary nor neutral connection is needed.

iv. Δ - y (DELTA-wye) is used as a step-up transformer at the point of generation, before transmission.

Independent of the three-phase connection of the primary and secondary, for a balance three-phase load, the apparent power, VA, from the supply to the load is

$$S = \sqrt{3} V_{line} I_{line} = 3 V_{phase} I_{phase} \tag{11.154}$$

Also the sum of the primary and secondary line voltages is zero, that is

$$V_{AB} + V_{BC} + V_{CA} = 0$$
$$V_{ab} + V_{bc} + V_{ca} = 0 \tag{11.155}$$

where upper case subscripts refer to the primary and lower case subscripts refer to the secondary.

Y-y (WYE-wye) connection

Electrically, the *Y-y* transformer connection shown in figure 11.17, can be summarized as follows.

$$\eta_{Y-y} = \frac{N_p}{N_s} = \frac{V_{AN}}{V_{an}} = \frac{I_a}{I_A} = \frac{I_a}{I_{L1}} = \frac{V_{BN}}{V_{bn}} = \frac{I_b}{I_B} = \frac{V_{CN}}{V_{cn}} = \frac{I_c}{I_C} \tag{11.156}$$

$$V_{AB} = V_{AN} + V_{NB} = V_{AN} - V_{BN} = \sqrt{3} V_{AN} e^{j30^\circ} = \sqrt{3} V_{AN} \angle 30^\circ$$
$$V_{BC} = V_{BN} + V_{NC} = V_{BN} - V_{CN} \qquad V_{CA} = V_{CN} + V_{NA} = V_{CN} - V_{AN}$$
$$V_{ab} = V_{an} - V_{bn} = \sqrt{3} V_{an} e^{j30^\circ} = \sqrt{3} V_{an} \angle 30^\circ \tag{11.157}$$
$$V_{bc} = V_{bn} - V_{cn} \qquad V_{ca} = V_{cn} - V_{an}$$

$$I_N = I_A + I_B + I_C \qquad I_n = I_a + I_b + I_c \tag{11.158}$$

The output current rating is

$$I_Y = \frac{|S|/3}{V/\sqrt{3}} = \frac{|S|}{\sqrt{3} V}$$

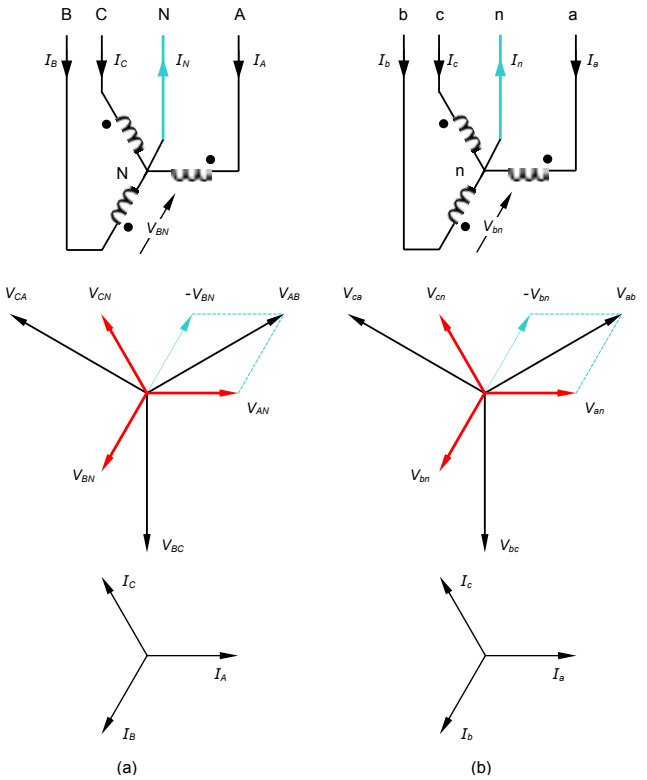


Figure 11.17. *Three-phase Y-y transformer: (a) winding arrangement and (b) phasor diagrams.*

Y- δ *(WYE-delta) connection*

The *Y*- δ transformer connection in figure 11.18 can be summarized as follows.

$$\eta_{Y-\delta} = \frac{N_p}{N_s} = \frac{V_{AN}}{V_{ab}} = \frac{I_{ba}}{I_A} = \frac{V_{BN}}{V_{bc}} = \frac{I_{cb}}{I_B} = \frac{V_{CN}}{V_{ca}} = \frac{I_{ac}}{I_C} \tag{11.159}$$

$$V_{AB} = V_{AN} - V_{BN} = V_{AN} - V_{AN}e^{-j120^\circ} = \sqrt{3}V_{AN}e^{j30^\circ}$$

$$I_a = I_{ba} - I_{ac} = I_{ba} - I_{ba}e^{-j240^\circ} = \sqrt{3}\, I_{ba}\, e^{-j30^\circ}$$

$$I_a + I_b + I_c = 0$$

The output current rating is

$$I_\Delta = \frac{|S|/3}{V} = \frac{|S|}{3V} \tag{11.160}$$

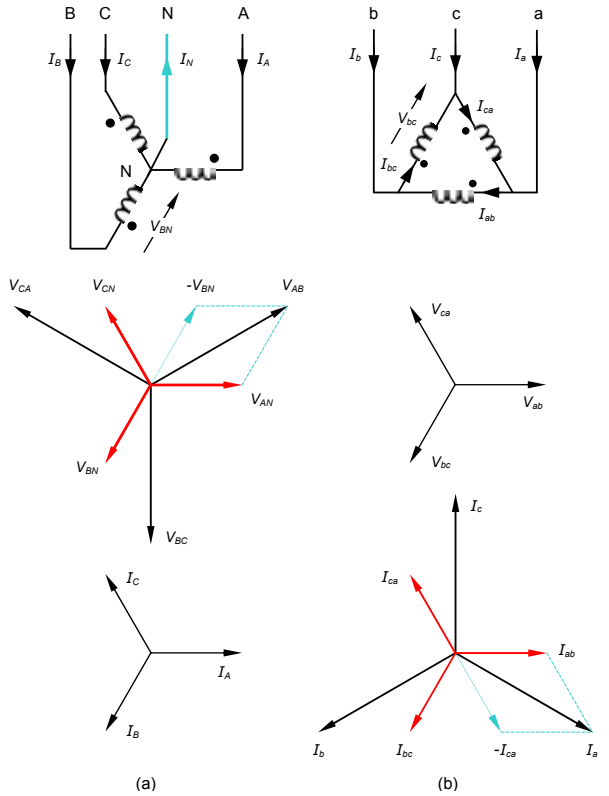


Figure 11.18. *Three-phase Y- δ transformer: (a) winding arrangement and (b) phasor diagrams.*

Δ-δ (DELTA-delta) connection

In figure 11.19, the Δ - δ transformer connection can be summarized as follows.

$$\eta_{\Delta-\delta} = \frac{N_p}{N_s} = \frac{V_{AB}}{V_{ab}} = \frac{I_a}{I_A} = \frac{I_{ba}}{I_{AB}} = \frac{V_{BC}}{V_{bc}} = \frac{I_b}{I_B} = \frac{I_{cb}}{I_{BC}} = \frac{V_{CA}}{V_{ca}} = \frac{I_c}{I_C} = \frac{I_{cb}}{I_{CA}} \tag{11.161}$$

$$I_A = I_{AB} - I_{CA} = \sqrt{3}\, I_{AB}e^{-j30^\circ} = \sqrt{3}I_{AB}\angle -30^\circ$$

$$I_B = I_{BC} - I_{AB} \qquad I_C = I_{CA} - I_{BC}$$

$$I_A + I_B + I_C = 0$$

$$I_a = I_{ab} - I_{ca} = \sqrt{3}\, I_{ab}e^{-j30^\circ} = \sqrt{3}I_{ab}\angle -30^\circ \tag{11.162}$$

$$I_b = I_{cb} - I_{ba} \qquad I_c = I_{ac} - I_{cb}$$

$$I_c + I_b + I_c = 0$$

The output current rating is

$$I_Y = \frac{|S|/3}{V/\sqrt{3}} = \frac{|S|}{\sqrt{3}V} \tag{11.163}$$

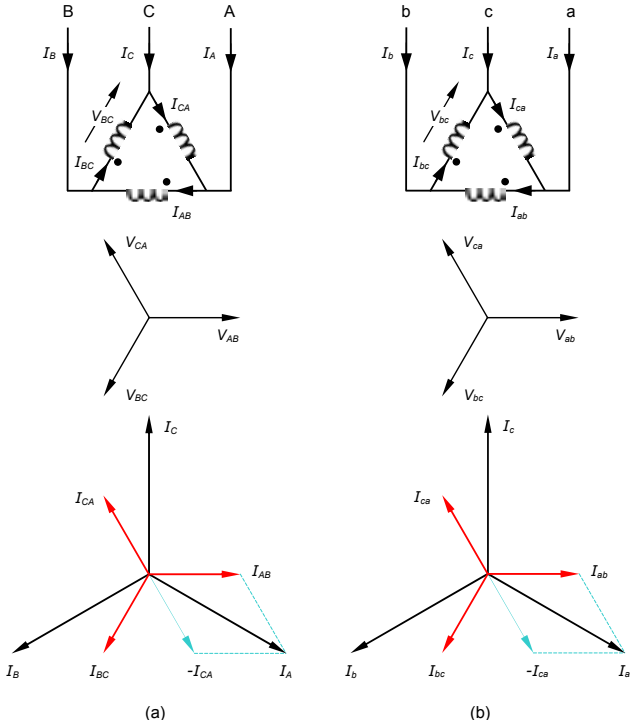


Figure 11.19. *Three-phase Δ-δ transformer: (a) winding arrangement and (b) phasor diagrams.*

Δ-y (DELTA-wye) connection

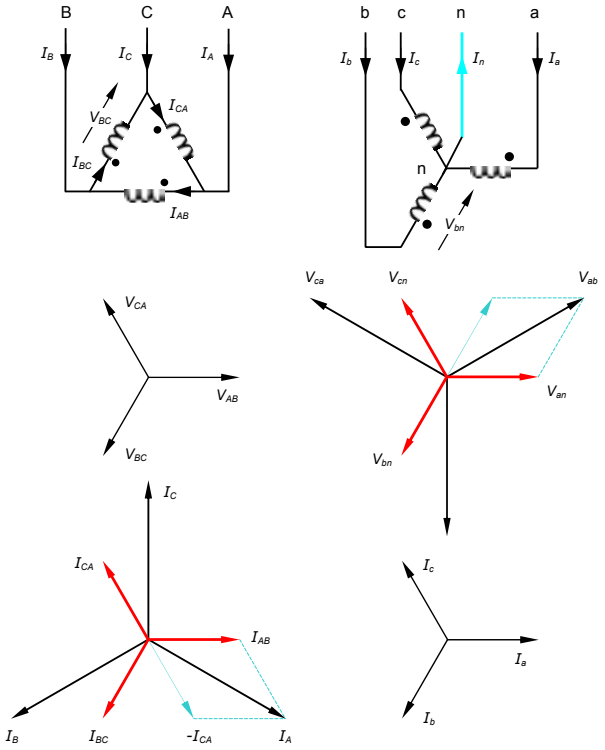
The Δ-*y* transformer connection in figure 11.20 can be summarized as follows.

$$\eta_{\Delta-y} = \frac{V_{AB}}{V_{ab}} = \frac{V_{AB}}{V_{ab}}\frac{e^{-j30^\circ}}{\sqrt{3}} = \frac{V_{AN}}{V_{an}} = \frac{I_a^*}{I_A^*} = \frac{I_a^*}{\left(\sqrt{3}\, I_{AB} e^{-j30^\circ}\right)^*} \tag{11.164}$$

$$I_A = I_{AB} - I_{CA} = I_{AB} - I_{AB}e^{-j240^\circ} = \sqrt{3}\, I_{an} e^{-j30^\circ}$$

The output current rating is

$$I_Y = \frac{|S|/3}{V/\sqrt{3}} = \frac{|S|}{\sqrt{3}V} \tag{11.165}$$



(a) (b)

Figure 11.20. *Three-phase* Δ-y *transformer: (a) winding arrangement and (b) phasor diagrams.*

11.3.5 Three-phase transformer, half-wave rectifiers - core mmf imbalance

Note that a delta secondary connection cannot be used for half-wave rectification as no physical neutral connection exists.

i. star connected primary Y-y (WYE-wye)

The three-phase half-wave rectifier with a star-star connected transformer in figure 11.21a is prone to magnetic *mmf* core bias. With a constant load current I_o, each diode conducts for 120º. Each leg is analysed on an *mmf* basis, and the current and *mmf* waveforms in figure 11.21a are derived as follows.

$$\begin{aligned} mmf_o &= N_s i_{s1} - N_p i_{p1} \\ mmf_o &= N_s i_{s2} - N_p i_{p2} \\ mmf_o &= N_s i_{s3} - N_p i_{p3} \end{aligned} \tag{11.166}$$

By symmetry and balance, the *mmf* in each leg must be equal.
If i_N is the neutral current then the equation for the currents is

$$i_{p1} + i_{p2} + i_{p3} = i_N \tag{11.167}$$

The same *mmf* equations are obtained if the load is purely resistive.
Any triplens in the primary will add algebraically, while any other harmonics will vectorially cancel to zero. Therefore the neutral may only conduct primary side triplen currents. Any input current harmonics are due to the rectifier and the rectifier harmonics of the order $h = cp \pm 1$ where $c = 0,1,2,\ldots$ and p is the pulse number, 3. No secondary-side third harmonics can exist hence $h \neq 3k$ for $k = 1, 2, 3, .$ Therefore no primary-side triplen harmonic currents exist to flow in the neutral, that is $i_N = 0$. In a balanced load condition, the neutral connection is redundant. The system equations resolve to

$$\begin{aligned} i_{p1} &= \frac{N_s}{N_p}\left(\frac{2}{3}i_{s1} - \frac{1}{3}i_{s2} - \frac{1}{3}i_{s3}\right) \\ i_{p2} &= \frac{N_s}{N_p}\left(-\frac{1}{3}i_{s1} + \frac{2}{3}i_{s2} - \frac{1}{3}i_{s3}\right) \\ i_{p3} &= \frac{N_s}{N_p}\left(-\frac{1}{3}i_{s1} - \frac{1}{3}i_{s2} + \frac{2}{3}i_{s3}\right) \\ mmf_o &= N_s\left(\frac{i_{s1} + i_{s2} + i_{s3}}{3}\right) = \frac{1}{3}N_s I_o \end{aligned} \tag{11.168}$$

Specifically, the core has an mmf dc bias of $\frac{1}{3}N_sI_o$.

Waveforms satisfying these equations are show plotted in figure 11.21a. The various transformer currents and voltages are

$$\begin{aligned} I_{s1} &= I_{s2} = I_{s3} = I_s = \frac{I_o}{\sqrt{3}} \\ I_{p1} &= I_{p2} = I_{p3} = I_p = \frac{\sqrt{2}}{3}\frac{N_s}{N_p} I_o \\ V_{p1} &= V_{p2} = V_{p3} = V_p = V_o \frac{N_p}{N_s}\frac{2\pi}{3\sqrt{6}} \\ V_{s1} &= V_{s2} = V_{s3} = V_s = V_o \frac{2\pi}{3\sqrt{6}} = \frac{V_o}{1.17} \end{aligned} \tag{11.169}$$

The fundamental ripple in the output voltage, at three times the supply frequency, is ¼V_o.
Therefore the various transformer VA ratings are

$$\begin{aligned} S_s &= V_{s1}I_{s1} + V_{s2}I_{s2} + V_{s3}I_{s3} = 3V_sI_s = \frac{\pi\sqrt{2}}{3}P_o = 1.48P_o \\ S_p &= V_{p1}I_p + V_{p2}I_p + V_{p3}I_p = 3V_pI_p = \frac{2}{3\sqrt{3}}P_o = 1.21P_o \\ \overline{S} &= ½\left(S_s + S_p\right) = P_o\frac{2+\pi\sqrt{6}}{3\sqrt{3}} = 1.34P_o \end{aligned} \tag{11.170}$$

The average output power is

$$P_o = I_oV_o \tag{11.171}$$

Since with a wye connected transformer primary, the transformer primary phase current is the line current, the supply power factor is

$$pf = \frac{P_o}{\overline{S}} = \frac{V_oI_o}{3V_pI_p} = \frac{\frac{3}{\pi}\sqrt{2}V_s\frac{\sqrt{3}}{2}I_o}{3\frac{N_p}{N_s}V_s\frac{\sqrt{3}}{2}\frac{N_s}{N_p}I_o} = 0.827 \tag{11.172}$$

Although the neutral connection is redundant for a constant load current, the situation is different if the load current has ripple at the three times the rectified ac frequency, as with a resistive load. Equations in (11.168) remain valid for the untapped neutral case. In such a case, when triplens exist in the load current, how they are reflected into the primary depends on whether or not the neutral is connected:

- No neutral connection – a triplen *mmf* is superimposed on the *mmf* dc bias of $⅓N_sI_o$.
- Neutral connected – a dc current (zero sequence) flows in the neutral and the associated zero sequence line currents in the primary, oppose the generation of any triplen *mmf* onto the dc *mmf* bias of $⅓N_sI_o$.

ii. delta connected primary Δ-y (DELTA-wye)

The three-phase half-wave rectifier with a delta-star connected transformer in figure 11.17b is prone to magnetic *mmf* core bias. With a constant load current I_o each diode conducts for 120º. Each leg is analysed on an *mmf* basis, and the current and *mmf* waveforms in figure 11.21b are derived as follows.

$$\begin{aligned} mmf_o &= N_s i_{s1} - N_p i_{p1} \\ mmf_o &= N_s i_{s2} - N_p i_{p2} \\ mmf_o &= N_s i_{s3} - N_p i_{p3} \end{aligned} \qquad (11.173)$$

$$i_{L1} = i_{p1} - i_{p3} \qquad i_{L2} = i_{p2} - i_{p1} \qquad i_{L3} = i_{p3} - i_{p2}$$

The line-side currents have average values of zero and if it is assumed that the core *mmf* has only a dc component, that is no alternating component, then based on these assumptions

$$mmf_o = N_s\left(\frac{i_{s1}+i_{s2}+i_{s3}}{3}\right) = \frac{1}{3}N_sI_o \qquad (11.174)$$

The primary currents are then

$$\begin{aligned} i_{p1} &= \left(i_{s1} - \frac{1}{3}I_o\right)\frac{N_s}{N_p} = \frac{N_s}{N_p}\left(\frac{2}{3}i_{s1} - \frac{1}{3}i_{s2} - \frac{1}{3}i_{s3}\right) \\ i_{p2} &= \left(i_{s2} - \frac{1}{3}I_o\right)\frac{N_s}{N_p} = \frac{N_s}{N_p}\left(-\frac{1}{3}i_{s1} + \frac{2}{3}i_{s2} - \frac{1}{3}i_{s3}\right) \\ i_{p3} &= \left(i_{s3} - \frac{1}{3}I_o\right)\frac{N_s}{N_p} = \frac{N_s}{N_p}\left(-\frac{1}{3}i_{s1} - \frac{1}{3}i_{s2} + \frac{2}{3}i_{s3}\right) \end{aligned} \qquad (11.175)$$

These line-side equations are the same as for the star connected primary, hence the same real and apparent power equations are also applicable to the delta connected primary transformer, viz. equations (11.170) and (11.171).
The line currents are

$$\begin{aligned} i_{L1} &= \frac{N_s}{N_p}\left(i_{p1} - i_{p3}\right) \\ i_{L2} &= \frac{N_s}{N_p}\left(i_{p2} - i_{p1}\right) \\ i_{L3} &= \frac{N_s}{N_p}\left(i_{p3} - i_{p2}\right) \end{aligned} \qquad (11.176)$$

The waveforms for these equations are shown plotted in figure 11.21b, where

$$I_p = \frac{N_s}{N_p}\frac{\sqrt{3}}{2}I_o \text{ and } I_L = \frac{N_s}{N_p}\sqrt{\frac{3}{2}}I_o \qquad (11.177)$$

that is $I_L = \sqrt{3}\,I_p$

$$\overline{S} = 1.34P_o \qquad (11.178)$$

The supply power factor is

$$pf = \frac{V_oI_o}{\sqrt{3}V_pI_L} = = 0.827 \qquad (11.179)$$

Although with a delta connected primary, the ac supply line currents are not the transformer primary currents, the supply power factor is the same as a star primary connection since the proportions of the input harmonics are the same.

The rms output voltage is

$$V_{o\,rms} = \sqrt{2}V_s\sqrt{\frac{3}{2\pi}\left(\frac{\pi}{3} + \frac{\sqrt{3}}{4}\right)} \qquad (11.180)$$

Each diode conducts for 120º and

$$\overline{I}_D = ⅓I_o \qquad I_{D\,rms} = \frac{I_o}{\sqrt{3}} \qquad \hat{V}_D = \sqrt{6}\frac{N_s}{N_p}V_p \qquad (11.181)$$

The primary connection, delta or wye, does not influence any ***dc*** mmf generated in the core, although the primary connection does influence if an ***ac*** mmf results.

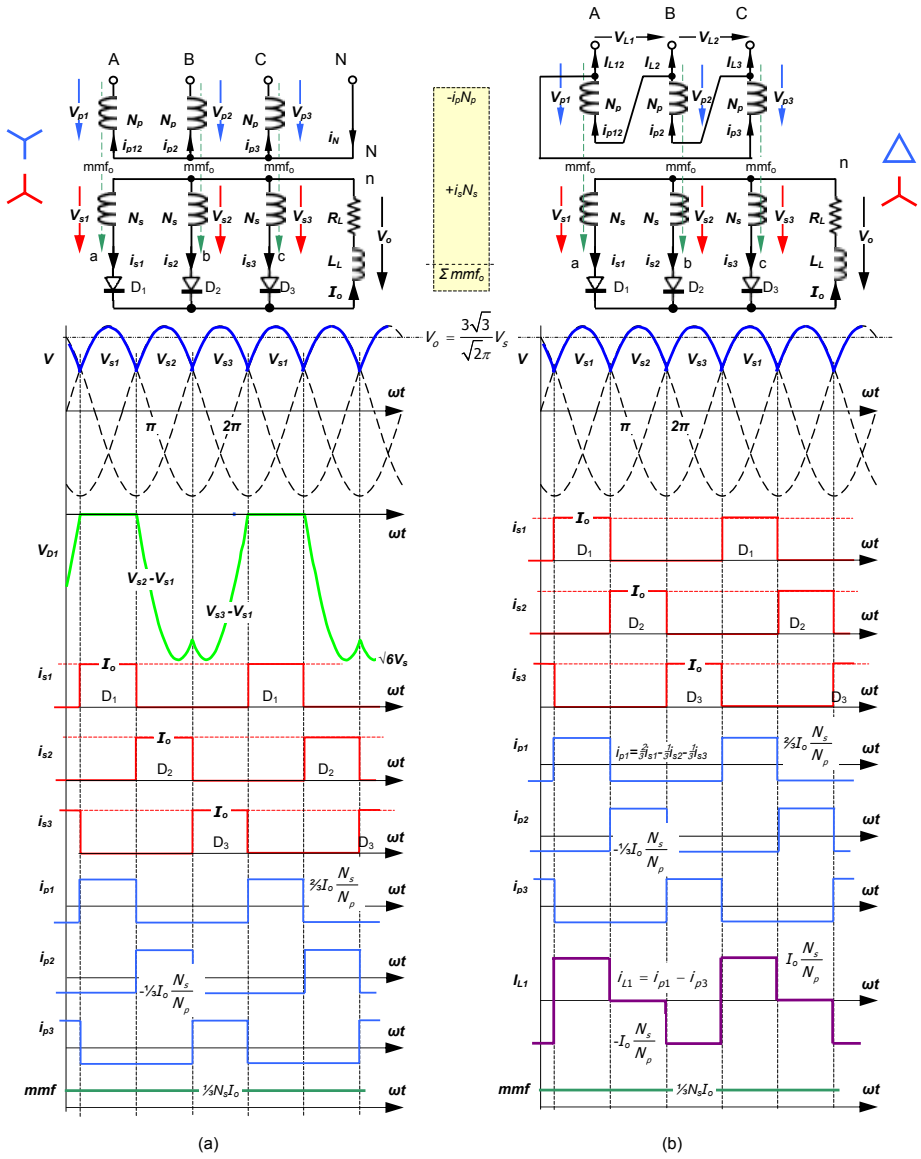


Figure 11.21. *Three-phase transformer winding arrangement with dc mmf bias: (a) star connected primary and (b) delta connected primary.*

11.3.6 Three-phase transformer with hexa-phase rectification, mmf imbalance

Figure 11.22 shown a tri-hexaphase half-wave rectifier, which can employ a wye or delta primary configuration, but only a star secondary connection is possible, since a neutral connection is required. The primary configuration can be shown to dictate core mmf bias conditions.

i. Y-y (WYE-wye) connection

The mmf balance for the wye primary connection in figure 11.22a is

$$\begin{aligned} N_s i_{s1} - N_s i_{s4} - N_p i_{p1} &= 0 \\ N_s i_{s3} - N_s i_{s6} - N_p i_{p2} &= 0 \\ N_s i_{s5} - N_s i_{s2} - N_p i_{p3} &= 0 \\ i_{p1} + i_{p2} + i_{p3} &= 0 \end{aligned} \qquad (11.182)$$

The primary currents expressed in terms of the secondary current are

$$\begin{aligned} i_{p1} &= \frac{N_s}{N_p}\left(\tfrac{2}{3} i_{s1} + \tfrac{1}{3} i_{s2} - \tfrac{1}{3} i_{s3} - \tfrac{2}{3} i_{s4} - \tfrac{1}{3} i_{s5} + \tfrac{1}{3} i_{s6}\right) \\ i_{p2} &= \frac{N_s}{N_p}\left(-\tfrac{1}{3} i_{s1} + \tfrac{1}{3} i_{s2} + \tfrac{2}{3} i_{s3} + \tfrac{1}{3} i_{s4} - \tfrac{1}{3} i_{s5} - \tfrac{2}{3} i_{s6}\right) \\ i_{p3} &= \frac{N_s}{N_p}\left(-\tfrac{1}{3} i_{s1} - \tfrac{2}{3} i_{s2} - \tfrac{1}{3} i_{s3} + \tfrac{1}{3} i_{s4} - \tfrac{2}{3} i_{s5} - \tfrac{1}{3} i_{s6}\right) \\ mmf &= N_s \frac{1}{3}\left(i_{s1} - i_{s2} + i_{s3} - i_{s4} + i_{s5} - i_{s6}\right) \end{aligned} \qquad (11.183)$$

These line side equations are plotted in figure 11.22a. Notice that an alternating mmf exists in the core related to the pulse frequency, $n = 2q = 6$.

The transformer primary currents and the line currents are

$$\begin{aligned} i_p &= \frac{\sqrt{2}}{3} I_o \\ i_L &= \frac{\sqrt{2}}{3} I_o \end{aligned} \qquad (11.184)$$

Note that because of the zero sequence current, triplens, in the delta primary that

$$i_L = \sqrt{2}\, i_p$$

not (11.185)

$$i_L = \sqrt{3}\, i_p$$

The transformer power ratings are

$$\begin{aligned} S_s &= 6\left(\frac{\pi}{3\sqrt{2}} V_o\right)\left(\frac{1}{\sqrt{6}} I_o\right) = \frac{\pi}{\sqrt{3}} P_o \\ S_p &= 3\left(\frac{\pi}{3\sqrt{2}} V_o\right)\left(\frac{\sqrt{2}}{3} I_o\right) = \frac{\pi}{3} P_o \\ \overline{S} &= ½\left(\frac{\pi}{\sqrt{3}} P_o + \frac{\pi}{3} P_o\right) = \frac{\pi}{6}\left(\sqrt{3}+1\right) P_o = 1.43 P_o \end{aligned} \qquad (11.186)$$

ii. Δ-y (DELTA-wye) connection

When the primary is delta connected, as shown in figure 11.22b, the mmf equations are the same as with a wye primary, namely

$$\begin{aligned} N_s i_{s1} - N_s i_{s4} - N_p i_{p1} &= 0 \\ N_s i_{s3} - N_s i_{s6} - N_p i_{p2} &= 0 \\ N_s i_{s5} - N_s i_{s2} - N_p i_{p3} &= 0 \end{aligned} \qquad (11.187)$$

but Kirchhoff's electrical current equation becomes of the following form for each phase:

$$mmf = \frac{1}{2\pi} \int_0^{2\pi} N_s \left(i_{s1} - i_{s4}\right) d\omega t = 0 \qquad (11.188)$$

Thus since each limb experiences an alternating current, similar to i_{s1}-i_{s4} for each limb, with an average value of zero, the line currents can be calculated from

$$i_{p1} = \frac{N_s}{N_p}\left(i_{s1} - i_{s4}\right) \qquad i_{p2} = \frac{N_s}{N_p}\left(i_{s3} - i_{s6}\right) \qquad i_{p3} = \frac{N_s}{N_p}\left(i_{s5} - i_{s2}\right) \qquad (11.189)$$

The line currents are

$$\begin{aligned} i_{L1} &= i_{p1} - i_{p3} = \frac{N_s}{N_p}\left(i_{s1} + i_{s2} - i_{s4} - i_{s5}\right) \\ i_{L2} &= i_{p2} - i_{p1} = \frac{N_s}{N_p}\left(-i_{s1} + i_{s3} + i_{s4} - i_{s6}\right) \\ i_{L3} &= i_{p3} - i_{p2} = \frac{N_s}{N_p}\left(-i_{s2} - i_{s3} + i_{s5} + i_{s6}\right) \end{aligned} \qquad (11.190)$$

The transformer primary currents and the line currents are

$$\begin{aligned} i_p &= \frac{1}{\sqrt{3}} I_o \\ i_L &= \sqrt{\frac{2}{3}} I_o \end{aligned} \qquad (11.191)$$

The transformer power ratings (which are relatively poor) are

$$\begin{aligned} S_s &= 6\left(\frac{\pi}{3\sqrt{2}} V_o\right)\frac{I_o}{\sqrt{6}} = \frac{\pi}{\sqrt{3}} P_o = 1.81 P_o \\ S_p &= 3\left(\frac{\pi}{3\sqrt{2}} V_o\right)\frac{I_o}{\sqrt{3}} = \frac{\pi}{\sqrt{6}} P_o = 1.28 P_o \\ \overline{S} &= ½\left(\frac{\pi}{\sqrt{3}} P_o + \frac{\pi}{\sqrt{6}} P_o\right) = ½\frac{\pi}{\sqrt{3}}\left(1 + \frac{1}{\sqrt{2}}\right) P_o = 1.55 P_o \end{aligned} \qquad (11.192)$$

The same primary and secondary apparent powers result for a purely resistive load.

The supply power factor is $pf = 3/\pi = 0.955$.

Independent of the primary connection, the average output voltage is

$$V_o = \frac{3\sqrt{2}}{\pi} V_s \qquad (11.193)$$

and the rms output voltage is

$$V_{o\,rms} = \sqrt{2} V_s \sqrt{\frac{6}{2\pi}\left(\frac{\pi}{6} + \frac{\sqrt{3}}{4}\right)} \qquad (11.194)$$

The diode average and rms currents are

$$I_D = \frac{I_o}{6} \qquad I_{D\,rms} = \frac{I_o}{\sqrt{6}} \qquad (11.195)$$

The maximum diode reverse voltage is

$$V_{Dr} = 2\sqrt{2} V_s \qquad (11.196)$$

The line currents are added to the waveforms in figure 11.22a and are also shown in figure 11.10b. The core mmf bias is zero, without any ac component associated with the 6-pulse rectification process. Zero sequence, triplen currents, can flow in the delta primary connection. A star connected primary is therefore not advisable.
If a single-phase inter-wye transformer is used between the neutrals of the two star rectifier group, the transformer apparent power factors improve significantly, to

$$S_s = 1.48 P_o \quad S_p = 1.05 P_o \text{ giving } \overline{S} = 1.26 P_o \qquad (11.197)$$

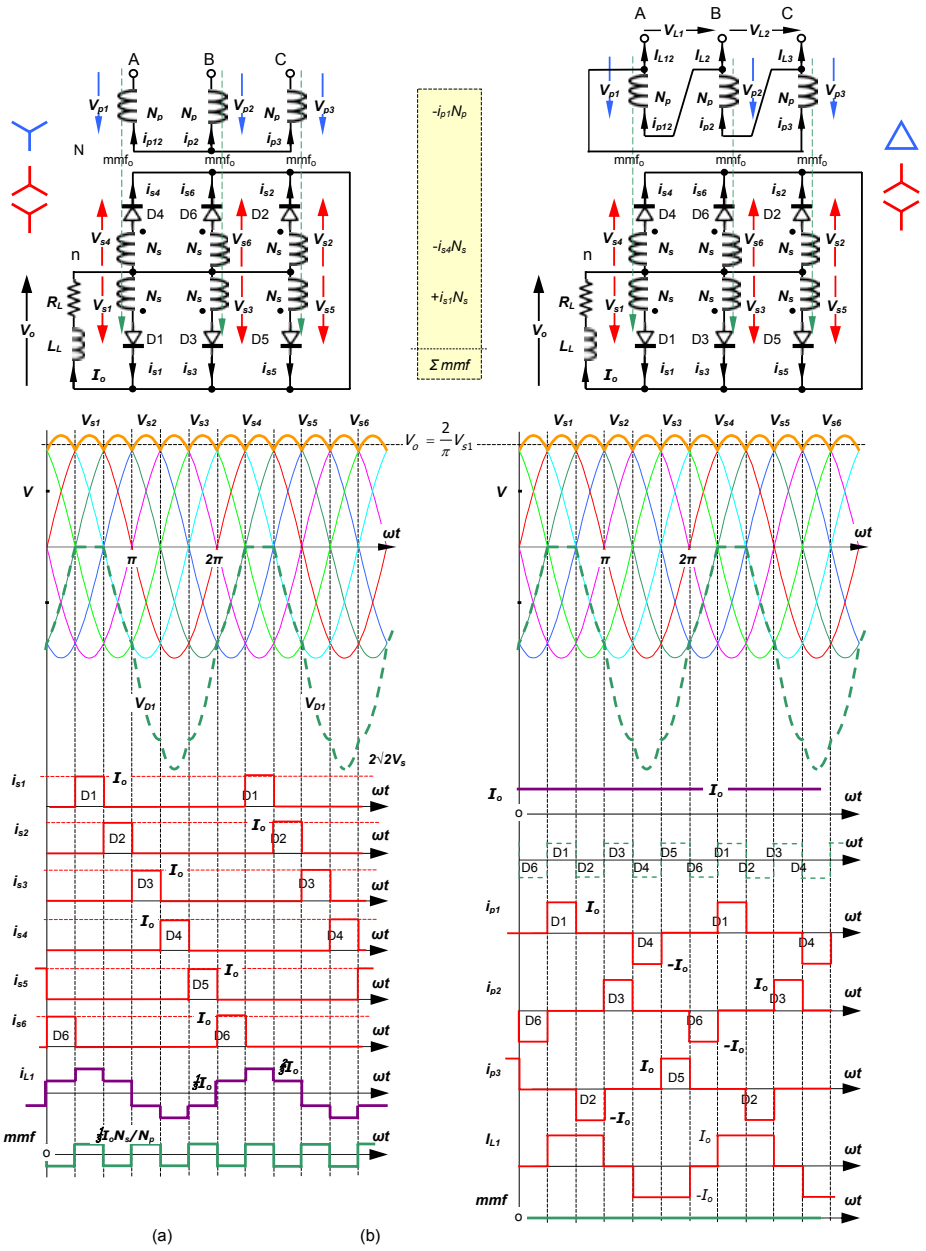


Figure 11.22. *Three-phase transformer winding arrangement with hexa-phase rectification: (a) star connected primary with dc mmf bias and (b) delta connected primary. (the transformer secondary and diode currents are the same in each case)*

11.3.7 Three-phase transformer mmf imbalance cancellation – zig-zag winding

In figures 11.23a and 11.24a, for balanced input currents and equal turns number N_s in the six windings

$$N_s\left(I_{aa'} + I_{nc'}\right) = N_s\left(I_{an} - I_{cn}\right)$$

whence (11.198)

$$N_s\left(I_{an} - I_{cn}\right) = \sqrt{3}\,N_s I_{an} \angle -30°$$

If the same windings were connected in series in a Y configuration the mmf would be $2NI_{an}$. Therefore 1.15 times more turns ($2/\sqrt{3}$) are needed with the zig-zag arrangement in order to produce the same mmf.

Similarly for the output voltage, when compared to the same windings used in series in a Y secondary configuration:

$$\begin{aligned} V_{na} &= V_{na'} + V_{a'a} \\ &= -V_{a'n} + V_{a'a} \\ &= \sqrt{3}V_{a'a}\angle 30° \end{aligned} \qquad (11.199)$$

That is, for a given line to neutral voltage, 1.15 times as many turns are needed as when Y connected.

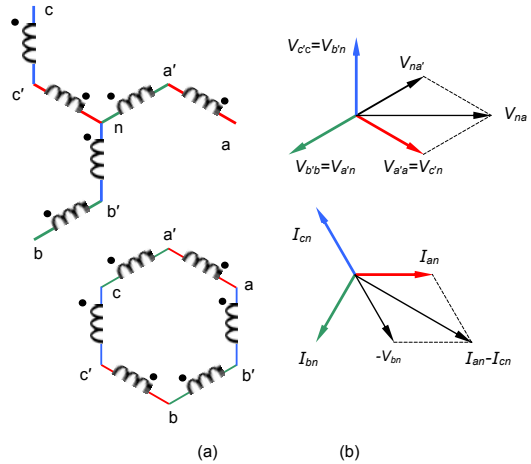


Figure 11.23. *Three-phase transformer secondary zig-zag winding arrangement: (a) secondary windings and (b) current and voltage phasors for the fork case.*

i. star connected primary Y-z (WYE-zigzag)

In figure 11.24, each limb of the core has an extra secondary winding, of the same number of secondary turns, N_s.
MMF analysis of each of the three limbs yields

$$\begin{aligned} &\text{limb 1:-} \quad mmf_o = -i_{p1}N_p + i_{s1}N_s - i_{s3}N_s \\ &\text{limb 2:-} \quad mmf_o = -i_{p2}N_p + i_{s2}N_s - i_{s1}N_s \\ &\text{limb 3:-} \quad mmf_o = -i_{p2}N_p + i_{s3}N_s - i_{s2}N_s \\ &i_{p1} + i_{p2} + i_{p3} = 0 \end{aligned} \qquad (11.200)$$

Adding the three *mmf* equations gives $mmf_o = 0$ and the alternating primary (and line) currents are

$$i_{p1} = \frac{N_s}{N_p}\left(i_{s1} - i_{s3}\right) \qquad i_{p2} = \frac{N_s}{N_p}\left(i_{s2} - i_{s1}\right) \qquad i_{p3} = \frac{N_s}{N_p}\left(i_{s3} - i_{s2}\right) \qquad (11.201)$$

These equations are plotted in figure 11.24a.

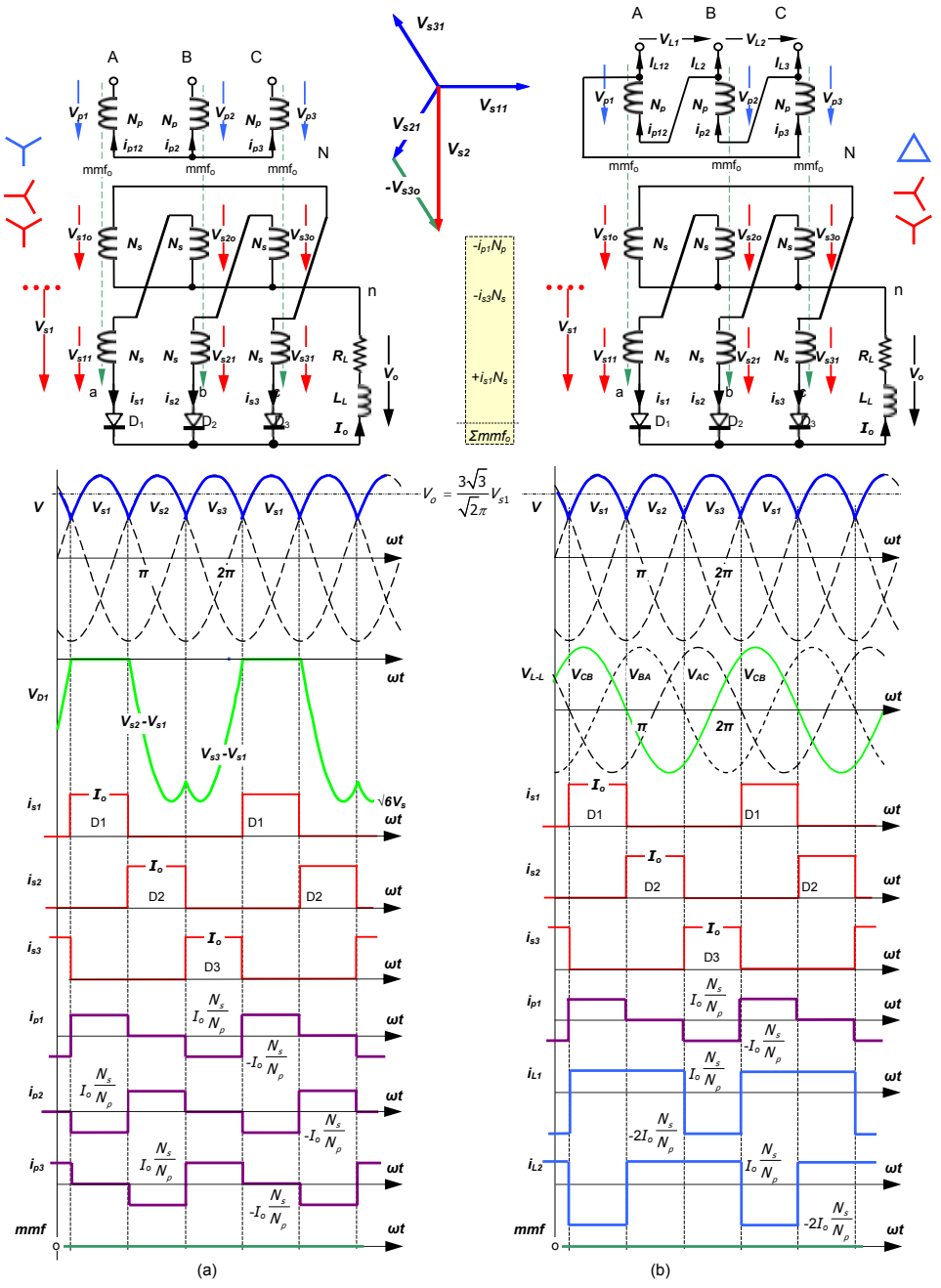


Figure 11.24. *Three-phase transformer winding zig-zag arrangement with no dc mmf bias: (a) star connected primary and (b) delta connected primary.*

If a 1:1:1 turns ratio is assumed, the power ratings of the transformer (which is independent of the turns ratio) involves the vectorial addition of the winding voltages.

$$\vec{V}_{s1} = \vec{V}_{s11} - \vec{V}_{s2o}$$
$$\vec{V}_{s2} = \vec{V}_{s21} - \vec{V}_{s3o} \qquad (11.202)$$
$$\vec{V}_{s3} = \vec{V}_{s31} - \vec{V}_{s1o}$$

The various transformer ratings are

$$S_s = 3I_sV_{s0} + 3I_sV_{s1} = 6I_sV_{so} = \frac{2\sqrt{2}\pi}{3\sqrt{3}}P_o \qquad S_p = 3I_pV_p = \frac{2\pi}{3\sqrt{3}}P_o$$
$$\bar{S} = \tfrac{1}{2}(S_s + S_p) = P_o\frac{\pi}{3\sqrt{3}}(\sqrt{2}+1) = 1.46P_o \qquad (11.203)$$

ii. delta connected primary Δ-z (DELTA-zigzag)

Carrying out an *mmf* balancing exercise, assuming no alternating *mmf* component, and the mean line current is zero, yields

$$i_{L1} = i_{p1} - i_{p3} = \frac{N_s}{N_p}(i_{s1} + i_{s2} - 2i_{s3})$$
$$i_{L2} = i_{p2} - i_{p1} = \frac{N_s}{N_p}(i_{s2} + i_{s3} - 2i_{s1}) \qquad (11.204)$$
$$i_{L3} = i_{p3} - i_{p2} = \frac{N_s}{N_p}(i_{s3} + i_{s1} - 2i_{s2})$$

The primary and secondary currents are the same whether for a delta or star connected primary, therefore

$$\bar{S} = \tfrac{1}{2}(S_s + S_p) = 1.46P_o \qquad (11.205)$$

If a 1:1:1 turns ratio is assumed, the line, primary and load current are related according to

$$I_L = \sqrt{\frac{2}{3}I_o^2 + \frac{4}{3}I_o^2} = \sqrt{2}I_o$$
$$I_p = \sqrt{\frac{2}{3}}I_o \qquad I_L = \sqrt{3}I_p \qquad (11.206)$$

A zig-zag secondary can be a Y-type fork for a possible neutral connection or alternatively, a Δ-type polygon when the neutral is not required.

Each diode conducts for 120° and

$$\bar{I}_D = \tfrac{1}{3}I_o \qquad I_{Drms} = \frac{I_o}{\sqrt{3}} \qquad \hat{V}_D = \sqrt{6}\frac{N_s}{N_p}V_p \qquad (11.207)$$

11.3.8 Three-phase transformer full-wave rectifiers – zero core mmf

Full-wave rectification is common in single and three phase applications, since, unlike half-wave rectification, the core *mmf* bias tends to be zero. In three-phase, it is advisable that either the primary or secondary be a delta connection. Any non-linearity in the core characteristics, namely hysteresis, causes triplen fluxes. If a delta connection is used, triplen currents can circulate in the winding, thereby suppressing the creation of triplen core fluxes. If a Y-y connection is used, a third winding set, delta connected, is usually added to the transformer in high power applications. The extra winding can be used for auxiliary type supply applications, and in the limit only one turn per phase need be employed if the sole function of the tertiary delta winding is to suppress core flux triplens.
The primary current harmonic content is the same for a given output winding configuration, independent of whether the primary is star or delta connected.

i. star connected primary Y-y (Wye-wye)

The Y-y connection shown in figure 11.25a (with primary and secondary neutral nodes *N*, *n* respectively) is the simplest to analyse since each phase primary current is equal to a corresponding phase secondary current.

$$\begin{aligned} mmf_o &= N_s i_{s1} - N_p i_{p1} \\ mmf_o &= N_s i_{s2} - N_p i_{p2} \\ mmf_o &= N_s i_{s2} - N_p i_{p2} \end{aligned} \qquad (11.208)$$

Adding the three *mmf* equations gives

$$3 \times mmf_o = N_p \sum_{i=1}^{3} i_{pi} - N_s \sum_{i=1}^{3} i_{si} \qquad (11.209)$$

but

$$i_{p1} + i_{p2} + i_{p3} = 0 \qquad (11.210)$$

and the secondary currents always sum to zero, then $mmf_o = 0$.
Additionally

$$i_{L1} = i_{p1} = \frac{N_s}{N_p} i_{s1} \qquad i_{L2} = i_{p2} = \frac{N_s}{N_p} i_{s2} \qquad i_{L3} = i_{p3} = \frac{N_s}{N_p} i_{s3} \qquad (11.211)$$

Generally

$$I_p = \frac{N_s}{N_p} I_s = \frac{N_s}{N_p} \sqrt{\frac{2}{3}}\, I_o \qquad (11.212)$$

whence

$$\begin{aligned} S_p &= \frac{2\pi}{3\sqrt{3}} P_o = 1.21 P_o \qquad S_s = \frac{\sqrt{2}\pi}{3} P_o = 1.48 P_o \\ \overline{S} &= ½\left(\frac{2\pi}{3\sqrt{3}} P_o + \frac{\sqrt{2}\pi}{3} P_o \right) = 1.35 P_o \end{aligned} \qquad (11.213)$$

The secondary harmonic currents are given by

$$I_{sh} = \frac{1}{h} I_{s1} = \frac{1}{h} \frac{\sqrt{6}}{\pi} I_o \quad \text{for } h = 6n \pm 1 \quad \forall \quad n > 0 \qquad (11.214)$$

The full-wave, three-phase rectified average output voltage (assuming the appropriate turns ratio, 1:1, to give the same output voltage for a given input line voltage) is

$$V_o = \frac{3\sqrt{3}}{\pi} V_p = \frac{3}{\pi} V_L \qquad (11.215)$$

The fundamental ripple in the output voltage, at six times the supply frequency, is $0.057V_o$.

Since with a star primary the line currents are the primary currents, the supply power factor is

$$pf = \frac{P_o}{S} = \frac{3}{\pi} = 0.955 \qquad (11.216)$$

ii. delta connected primary Δ-y (Delta-wye)

The secondary phase currents in figure 11.25b are the same as for the Y-y connection, but the line currents are composed as follows

$$i_{L1} = i_{p1} - i_{p3} \qquad i_{L2} = i_{p2} - i_{p1} \qquad i_{L3} = i_{p3} - i_{p2} \qquad (11.217)$$

Such that

$$\begin{aligned} I_p &= \frac{N_s}{N_p} I_s = \frac{N_s}{N_p} \sqrt{\frac{2}{3}}\, I_o \\ I_L &= \sqrt{3}\, I_p = \frac{N_s}{N_p} \sqrt{3}\, I_s = \frac{N_s}{N_p} \sqrt{2}\, I_o \end{aligned} \qquad (11.218)$$

The secondary harmonic currents are given by

$$I_{sh} = \frac{1}{h} I_{s1} = \frac{1}{h} \frac{\sqrt{6}}{\pi} I_o \quad \text{for } h = 6n \pm 1 \quad \forall \quad n > 0 \qquad (11.219)$$

The full-wave, three-phase rectified average output voltage (assuming the appropriate turns ratio, √3:1, to give the same output voltage for a given input line voltage) is

$$V_o = \frac{3\sqrt{3}}{\pi} V_p = \frac{3}{\pi} V_L \qquad (11.220)$$

The transformer apparent power components are

$$S_s = 1.05 P_o \quad S_p = 1.05 P_o \quad \text{hence} \quad \overline{S} = 1.05 P_o \qquad (11.221)$$

The fundamental ripple in the output voltage, at six times the supply frequency, is $0.057V_o$.

The supply power factor is

$$pf = \frac{3}{\pi} = 0.955 \qquad (11.222)$$

iii. star connected primary Y-δ (Wye-delta)

In the Y-δ configuration in figure 11.26a, there are no zero sequence currents hence no *mmf* bias arises, $mmf_o = 0$, and both transformer sides have positive and negative sequence currents.

$$\begin{aligned} &i_{p1} = \frac{N_s}{N_p} i_{s1} \qquad i_{p2} = \frac{N_s}{N_p} i_{s2} \qquad i_{p3} = \frac{N_s}{N_p} i_{s3} \\ &\text{and } i_{s1} + i_{s2} + i_{s3} = 0 \end{aligned} \qquad (11.223)$$

where

$$\begin{aligned} i_{L1} &= \frac{N_s}{N_p} \left(i_{s1} - i_{s2} \right) = i_{p1} - i_{p2} \\ i_{L2} &= \frac{N_s}{N_p} \left(i_{s2} - i_{s3} \right) = i_{p2} - i_{p3} \\ i_{L2} &= \frac{N_s}{N_p} \left(i_{s3} - i_{s1} \right) = i_{p3} - i_{p1} \end{aligned} \qquad (11.224)$$

Thus the transformer currents are related to the supply line currents by

$$\begin{aligned} i_{p1} &= \frac{N_s}{N_p} i_{s1} = \frac{2}{3} i_{L1} - \frac{2}{3} i_{L2} \\ i_{p2} &= \frac{N_s}{N_p} i_{s2} = \frac{2}{3} i_{L2} - \frac{2}{3} i_{L3} \\ i_{p3} &= \frac{N_s}{N_p} i_{s3} = \frac{2}{3} i_{L3} - \frac{2}{3} i_{L1} \end{aligned} \qquad (11.225)$$

where

$$i_{L1} + i_{L2} + i_{L3} = 0 \qquad (11.226)$$

Generally

$$I_p = \frac{N_s}{N_p} I_s = \frac{N_s}{N_p} \frac{2\sqrt{2}}{3} ½ I_o \qquad (11.227)$$

The full-wave, three-phase rectified average output voltage (assuming the appropriate turns ratio, √3:1, to give the same output voltage for a given input line voltage) is

$$V_o = \frac{3\sqrt{3}}{\pi} V_p = \frac{3}{\pi} V_L \qquad (11.228)$$

The fundamental ripple in the output voltage, at six times the supply frequency, is $2/5 \times 7 = 0.057 V_o$.

The supply power factor is

$$pf = \frac{3}{\pi} \qquad (11.229)$$

for an output power, $P_o = V_o I_o$,

iv. delta connected primary Δ-δ (Delta-delta)

The phase primary and secondary voltages are in phase.

As shown in figure 11.26b the line currents are composed as follows

$$i_{L1} = i_{p1} - i_{p3} \qquad i_{L2} = i_{p2} - i_{p1} \qquad i_{L3} = i_{p3} - i_{p2} \tag{11.230}$$

The transformer primary and secondary currents are

$$i_{p1} = \frac{N_s}{N_p} i_{s1} \qquad i_{p2} = \frac{N_s}{N_p} i_{s2} \qquad i_{p3} = \frac{N_s}{N_p} i_{s3} \tag{11.231}$$

and

$$\begin{aligned} i_{p1} + i_{p2} + i_{p3} &= 0 \\ i_{s1} + i_{s2} + i_{s3} &= 0 \\ i_{L1} + i_{L2} + i_{L3} &= 0 \end{aligned} \tag{11.232}$$

Generally

$$I_p = \frac{N_s}{N_p} I_s \tag{11.233}$$

The full-wave, three-phase rectified average output voltage (assuming the appropriate turns ratio, 1:1, to give the same output voltage for a given input line voltage) is

$$V_o = \frac{3\sqrt{3}}{\pi} V_p = \frac{3}{\pi} V_L \tag{11.234}$$

The rms output voltage is

$$V_{o\,rms} = \sqrt{2} V_s \sqrt{\frac{3}{2} + \frac{9\sqrt{3}}{4\pi}} \tag{11.235}$$

The fundamental ripple in the output voltage, at six times the supply frequency, is $0.057V_o$.

The primary and secondary apparent powers are

$$S_p = S_s = \frac{\pi}{3} P_o = 1.05 P_o \tag{11.236}$$

Thus the supply power factor is

$$pf = \frac{3}{\pi} \tag{11.237}$$

for an output power, $P_o = V_o I_o$,

In summary, when the primary and secondary winding configurations are the same (Δ-δ or Y-y) the input and output line voltages are in phase, otherwise (Δ-y or Y-δ) the input and output line voltages are shifted by 30° relative to one another.
Independent of the transformer primary and secondary connection, for a specified input and output voltage, the following electrical equations hold.

$$V_o = \frac{3\sqrt{3}}{\pi} V_p = \frac{3}{\pi} \frac{N_s}{N_p} V_L \qquad pf = \frac{3}{\pi}$$

$$\bar{I}_D = \frac{1}{3} I_o \qquad I_{D\,rms} = \frac{1}{\sqrt{3}} I_o \qquad \hat{V}_{DR} = \sqrt{3}\sqrt{2} V_s$$

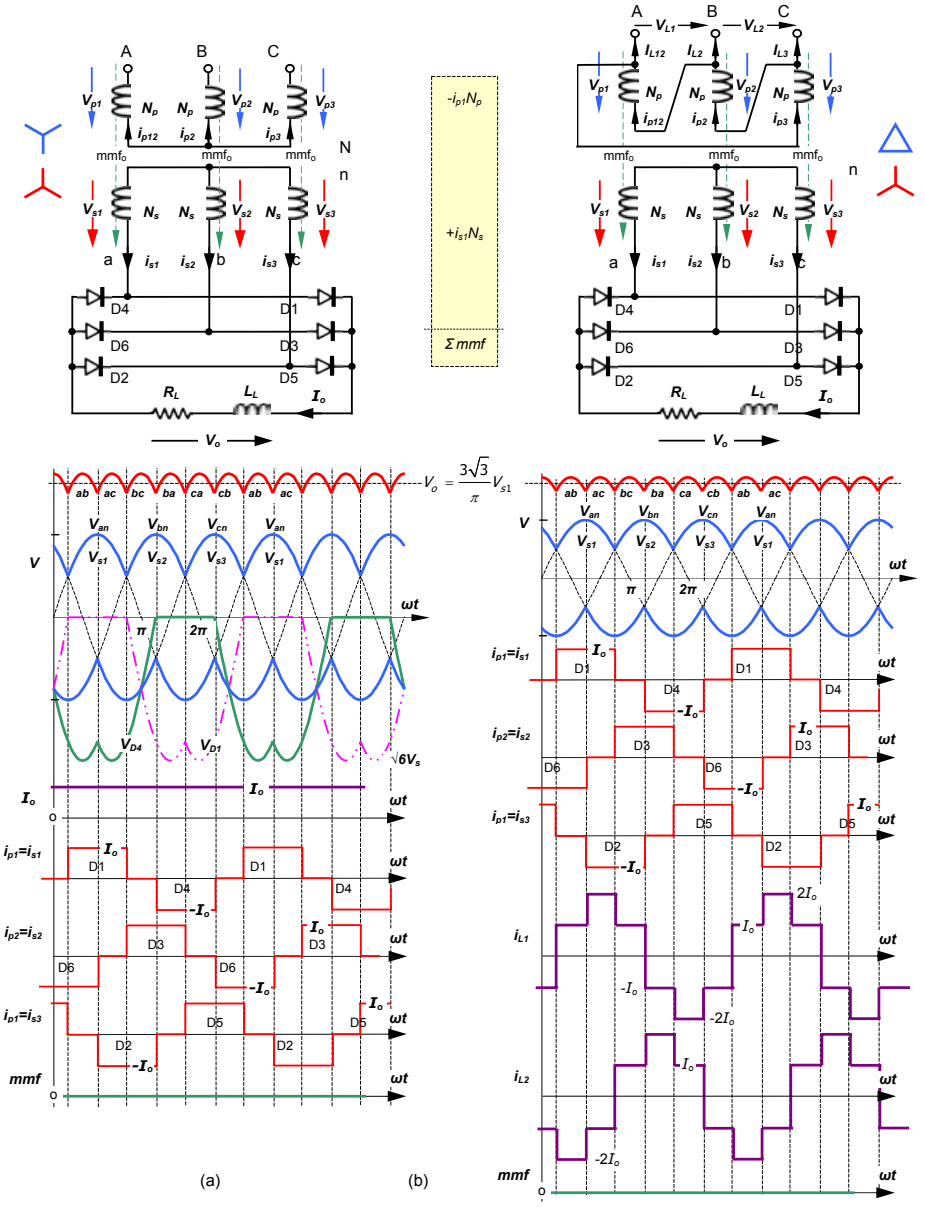


Figure 11.25. *Three-phase transformer wye connected secondary winding with full-wave rectification and no resultant dc mmf bias: (a) star connected primary Y-y and (b) delta connected primary Δ-y.*

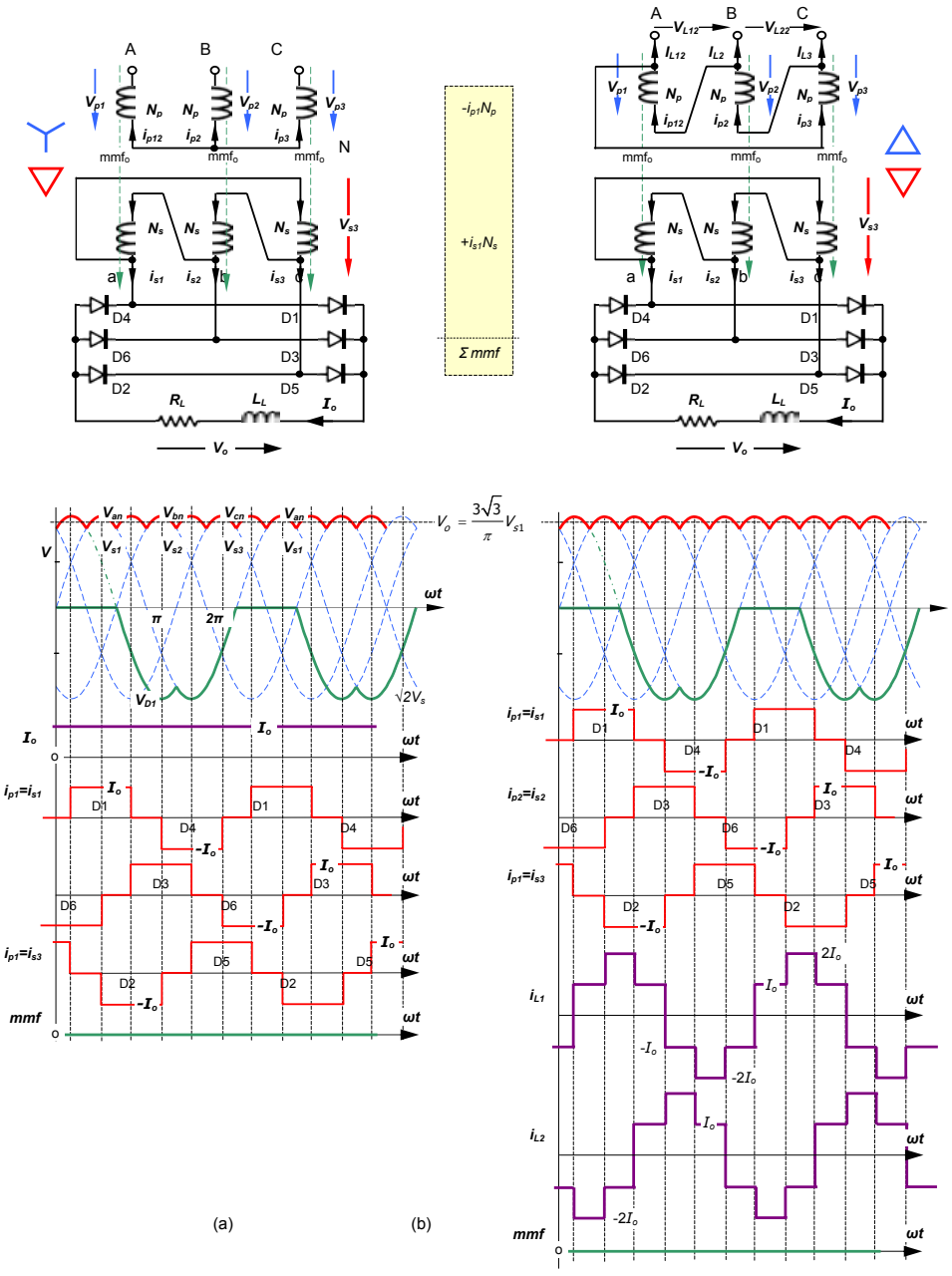


Figure 11.26. *Three-phase transformer with delta connected secondary winding with full-wave rectification and no resultant dc mmf bias: (a) star connected primary Y- δ and (b) delta connected primary Δ- δ.*

11.4 Voltage multipliers

Voltage multipliers are ac to dc power conversion circuits, comprised of diodes and capacitors that are interconnected so as to produce a high potential dc voltage from a lower voltage ac source. As in figure 11.27, multipliers are made up of cascaded stages each comprised of a diode and a capacitor.

Voltage multipliers are a simple way to generate high voltages at relatively low currents. By using only capacitors and diodes, the voltage multipliers can step up relatively low voltages to extremely high values, while at the same time being far lighter and cheaper than transformers. The advantage of the circuit is that the voltage across each cascaded stage is only equal to twice the peak input voltage, so it requires relatively low cost components and is easy to insulate. One can also tap the output from any stage, like a multi-tapped transformer.

The voltage multiplier has poor voltage regulation, that is, the voltage drops rapidly as a function the output current. The output *I-V* characteristic is approximately hyperbolic, so it is suitable for charging capacitor banks to high voltages at near constant charging power. Furthermore, the ripple on the output, particularly at high loads, is high. The output voltage is not isolated from the input voltage source, although transformer coupling provides general isolation.

The most commonly used multiplier circuit is the half-wave series multiplier. Other multiplier circuits can be derived from its operating principles.

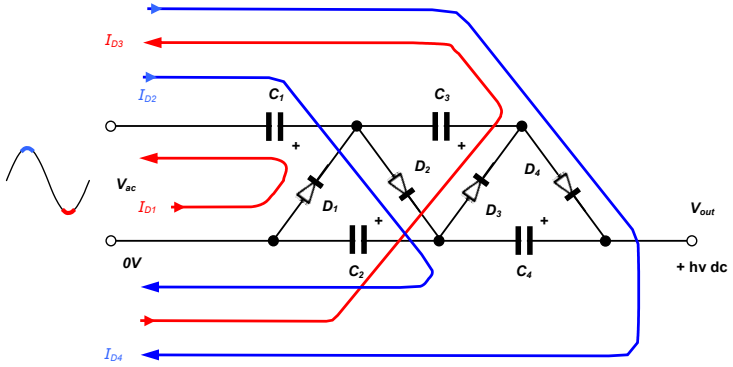


Figure 11.27. *Charging sequence of a half-wave series positive output voltage multiplier.*

The following description for a two-stage series voltage multiplier assumes no losses and represents sequential reversals of polarity of the source transformer T_s in the figure 11.27. The number of stages is equal to the number of smoothing capacitors between ground and V_{out}, which in this case is two, capacitors C_2 and C_4.

- V_{ac} = Negative Peak: C_1 charges through D_1 to V_{pk} by current $\boldsymbol{I_{D1}}$
- V_{ac} = Positive Peak: V_{pk} of Ts adds arithmetically to existing potential C_1, thus C_2 charges to 2 V_{pk} thru D_2 by current $\boldsymbol{I_{D2}}$
- V_{ac} = Negative Peak: C_3 is charged to $2V_{pk}$ through D_3 by current $\boldsymbol{I_{D3}}$
- V_{ac} = Positive Peak: C_4 is charged to $2V_{pk}$ by current $\boldsymbol{I_{D4}}$ through D_4 then V_{pk}.

For *N* stages (series capacitors) the output voltage is $N \times V_{pk}$.

11.4.1 Half-wave series multipliers

The capacitors are in series, so effectively capacitance is as for series connected capacitors, C/*N*, but voltage rating is the cumulative sum of the series capacitors between the output terminals. This multiplier is the most common, and is versatile, being used in high-voltage, low-current applications. The basic charging sequence in figure 11.28 is as for the circuit shown in figure 11.27, where the diodes conduct in the order D_1 to D_4, for both output polarity versions.

Half-wave series voltage multiplier features include:

- A wide range of multiplication stages
- Low cost
- Uniform stress per stage on diodes and capacitors, $2V_{pk}$ and V_{pk}

Any one capacitor can be eliminated from the capacitor filter bank if the load is capacitive. Whether full wave or half-wave, the series diodes prevent the output voltage from swinging negative. At high discharges, part of the output current is also drawn via a diode, hampering rapid high current discharge.

Dual polarity output voltage is produced by connecting positive and negative multipliers as shown in the four stage circuit is shown in figure 11.28c, where an unlimited stage number can be cascaded. Since regulation is proportional to N^3, a large number of stages eventually becomes ineffective. A centre tapped capacitor string connection reduces the maximum voltage potential with respect to ground. An odd number of stages can be produced as well as an even number of stages. The output voltage may be tapped at any point on the capacitor series filter bank.

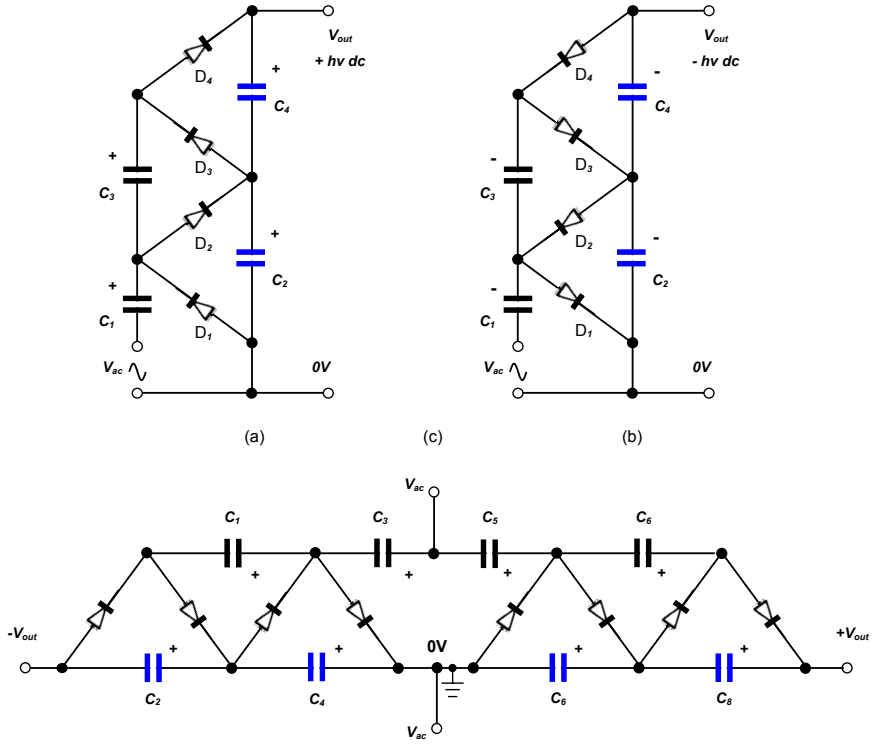


Figure 11.28. *Series half-wave voltage multipliers: (a) two stage positive hv output voltage; (b) two stage negative hv output voltage; and (c) four stage multiplier configured with ± output hv voltage.*

Once a load is connected at the output, the output voltage decreases due to the voltage regulation. Also, any small fluctuation of load impedance causes a large fluctuation in the multiplier output voltage due to the number of stages involved. For this reason, voltage multipliers are used only in special applications where the load is constant and has a high impedance or where voltage stability is not critical.

Half-wave Output Voltage

The open-circuit output voltage $V_{o/c}$ of each stage is nominally twice the peak input voltage V_{pk}. Assuming the ac input voltage and frequency are constant, for N cascaded stages, the output voltage is

$$V_{o/c} = 2N \times V_{pk} \qquad (11.238)$$

In practice, several cycles are required to reach full output voltage. The output voltage follows an RC network exponential curve, where R is the output impedance of the ac source, whilst C is the effective dynamic capacitance of the voltage multiplier, N×C. This charging occurs only upon switch-on of the voltage multiplier from a discharged state, and does not repeat itself unless the output is short circuited. The most common input ac waveforms are sine waves and square waves.

Output Voltage Regulation

DC output voltage drops as the dc output current increases. Regulation is the drop in dc output voltage from the ideal at a specified dc output current (assuming the ac input voltage and ac input frequency are constant). The voltage drop under load is mostly reactive and is calculated as:

$$V_{reg} = I_o \times \frac{4N^3 + 3N^2 - N}{6f \times C} = I_o \times \frac{4N^2 + 3N^2 - 1}{6f \times C/N} \qquad (11.239)$$

where:

I_o is the load or output dc current (A)
C is the stage capacitance (F)
f is the ac frequency (Hz)
N is the number of stages
C/N is the effective output capacitance.

Regulation voltage droop is not a power losses in a multiplier. Power losses are primarily diode forward conduction and rarely result in excessive multiplier temperatures at the low current loadings.
Substituting V_{reg} from equation (11.239):

$$V_{out} = V_{o/c} - V_{reg} = 2NV_{pk} - I_o \times \frac{4N^3 + 3N^2 - N}{6f \times C} \qquad (11.240)$$

Output Voltage Ripple

Ripple voltage is the magnitude of fluctuation in dc output voltage at a specific output current. This assumes the ac input voltage and frequency are maintained constant. The ripple voltage in the case where all stage capacitances, C_1 through C_{2N}, are equal, is:

$$V_{ripple} = I_o \times \frac{N^2 + N}{2f \times C} \qquad (11.241)$$

The ripple grows rapidly as the number of stages increases, with N squared. A common modification to the design is to make the stage capacitances larger at the input, with $C_1 = C_2 = N$×C, $C_3 = C_4 = (N\text{-}1)$×C, and so forth. Then the ripple is:

$$V_{ripple} = \frac{I_o}{f \times C} \qquad (11.242)$$

For a large number of stages, $N \geq 5$, the N^3 term in the voltage drop equation dominates. Differentiating the V_{out} equation without the negligible terms, with respect to the number of stages and equating to zero, gives an equation for the optimum (integer) number of stages N_{opt} for the equal valued capacitor design:

$$\frac{dV_{out}}{dN} = \frac{d}{dN}\left(2NV_{pk} - \frac{I_o}{6f \times C} \times 4N^3\right) = 0$$

:

$$N_{opt} = \text{int}\left[\left(\frac{V_{pk} f \times C}{I_o}\right)^{1/2}\right] \qquad (11.243)$$

Increasing the frequency can dramatically reduce the ripple, and the voltage drop under load, which accounts for the popularity driving a multiplier stack with a switching power supply.
If the driving voltage V_{pk} and the required output voltage $V_{o/c}$ are known, the optimum number of cascaded stages is:

$$N_{opt} = \text{int}\left[\frac{3V_{out}}{4V_{pk}}\right] \qquad (11.244)$$

11.4.2 Half-wave parallel multipliers

Opposite polarity half-wave parallel voltage multipliers are shown in figure 11.29. The output capacitors share a common connection but must have a high voltage rating. The output is usually low voltage but with high currents. The basic charging sequence in figure 11.29 is the same as shown in figure 11.27, where the diodes conduct in the order D_1 to D_4, for both output polarity versions.
Parallel multipliers offer the following features:

- Uniform stress on diodes
- compact
- Voltage stress on capacitors increases with successive stages by V_{pk}
- Highly efficient

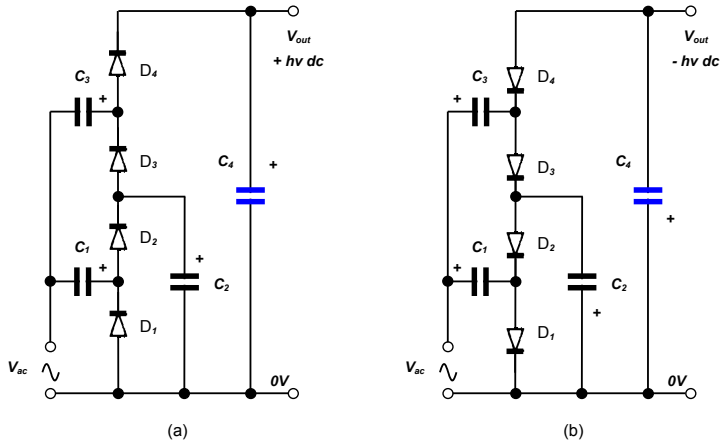


Figure 11.29. *Parallel half-wave voltage multipliers: (a) two stage positive hv output voltage and (b) two stage negative hv output voltage.*

11.4.3 Full-wave series multipliers

Increasing the frequency can dramatically reduce the ripple, and the voltage drop under load, which can be achieved by driving a multiplier stack with a switched mode power supply.
Figure 11.30 shows a typical full-wave two-stage series voltage multiplier. It is comprised of two anti-phase ac input half-wave multipliers sharing a common series output capacitor string. This effectively doubles the number of charging cycles per second, and thus reduces the voltage drop and ripple factor. The input is usually fed from a centre-tapped ac transformer or MOSFET H-bridge circuit.

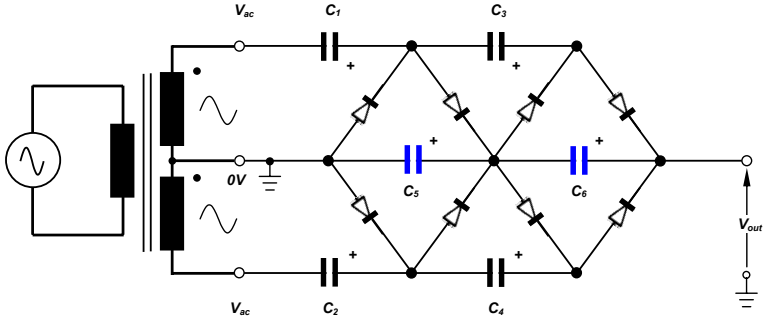


Figure 11.30. *Two-stage series full-wave voltage multiplier.*

The full-wave series voltage multiplier has the following general features:

- Uniform stress on components
- Highly efficient
- High Voltage
- High power capability
- Easy to produce
- Increased voltage stress on capacitors with successive stages
- Wide range of multiplication stages

Full-wave Output Voltage
As with the half-wave voltage multiplier, the full wave voltage multiplier output voltage is given by:

$$V_{o/c} = 2NV_{pk} \tag{11.245}$$

Output Voltage Regulation
DC output voltage decreases as dc output current increases. Regulation is the drop in dc output voltage from the ideal at a specified dc output current, assuming constant ac input voltage and frequency. The voltage drop under load is mostly reactive and is:

$$V_{reg} = I_o \times \frac{N^3 + 2N}{6f \times C} = I_o \times \frac{N^2 + 2}{6f \times C/N} \tag{11.246}$$

where:
I_o is the load or output dc current (A)
C is the stage capacitance (F)
f is the ac frequency (Hz)
N is the number of stages
C/N is the effective output capacitance.

Regulation voltage droop is not a power losses in a multiplier. Power losses are primarily diode forward conduction and rarely result in excessive multiplier temperatures at the low current loadings. Substituting equation (11.246) for V_{reg}:

$$V_{out} = V_{o/c} - V_{reg} = 2N \times V_{pk} - I_o \times \frac{N^3 + 2N}{6f \times C} \tag{11.247}$$

Output Voltage Ripple
The ripple voltage, in the case where all stage capacitances are equal, is given by:

$$V_{ripple} = I_o \times \frac{N}{2f \times C} \tag{11.248}$$

If the driving voltage V_{pk} and the required output voltage $V_{o/c}$ are known, the optimum number of cascaded stages is:

$$N_{opt} = \text{int}\left[\frac{0.521V_{out}}{V_{pk}}\right] \tag{11.249}$$

Example 11.9: ***Half-wave voltage multiplier***

A three-stage half-wave series voltage multiplier, is driven by a 50kHz peak voltage of 10kV, with 1nF capacitances, and a load current of 10mA.

i. Calculate the open circuit output voltage, regulated output voltage, ripple voltage, and optimal number of stages for the required voltage transfer function.
ii. What is the capacitance and voltage rating of each stage of a parallel connected multiplier?
iii. What is the output ripple if progressively smaller capacitance is used?.

Solution

i. In a three-stage voltage multiplier, the no load voltage $V_{o/c} = 2 \times N \times V_{pk} = 2 \times 3 \times 10\text{kV} = 60\text{kV}$

$$V_{reg} = I_o \frac{4N^3 + 3N^2 - N}{6f \times C} = 10\text{mA} \times \frac{3^3 + 3 \times 3^2 - 3}{6 \times 50\text{kHz} \times 1\text{nF}} = 1.7\text{kV}$$

$V_{out} = 60\text{kV} - 1.7\text{kV} = 58.3\text{kV}$
So the output voltage will swing between 6kV and 58.3kV, depending on the load current.
The output ripple voltage is

$$V_{ripple} = I_o \frac{N^2 + N}{2f \times C} = 10\text{mA} \frac{3^2 + 3}{2 \times 50\text{kHz} \times 1\text{nF}} = 3\text{kV}$$

The optimal number of stages, from equation (11.249), is

$$N_{opt} = \text{int}\left[\frac{0.521 \times V_{out}}{V_{pk}}\right] = \text{int}\left[\frac{0.521 \times 58.3\text{kV}}{10\text{kV}}\right] = 3$$

ii. An equivalent parallel multiplier would require each capacitor stage to equal the total series capacitance of the AC capacitor bank.

In this case, the three capacitors in the dc bank would equal 1000pF/3 or 330pF. The parallel equivalent would require 330pF capacitors in each stage. However, each successive stage, from the input, would require a higher voltage capacitor, 20kV, 40kV and 60kV, respectively.

iii. When $C_1 = C_2 = N{\times}C = 3nF$, $C_3 = C_4 = (N\text{-}1){\times}C = 2nF$, $C_5 = C_6 = (N\text{-}2){\times}C = 1nF$.

$$V_{ripple} = \frac{I_o}{f \times C} = \frac{10mA}{50kHz \times 1nF} = 200V$$

This modification reduces the ripple voltage from 3kV to just 200V.

♣

Example 11.10: ***Full-wave voltage multiplier***

A three-stage full-wave parallel voltage multiplier, is driven by a 50kHz peak voltage of 10kV, with 1nF capacitances, and a load current of 10mA. Calculate the output voltage and ripple voltage.

Solution

In a three-stage voltage multiplier, the no load voltage $V_{o/c} = 2{\times}N{\times}V_{pk} = 2{\times}3{\times}10kV = 60kV$.

$$V_{reg} = I_o \frac{N^3 + 2N}{6f \times C} = 10mA \frac{3^3 + 2 \times 3}{6 \times 50kHz \times 1nF} = 1.1kV$$

Full-wave rectification reduces the regulation voltage drop from 1.7kV in example 11.21, to 1.1kV. The output voltage is increased by 600V, from 58.3kV in example 11.21, to $V_{out} = 60kV - 1.1kV = 58.9kV$.
The ripple voltage reduces from 3kV for half-wave multiplication in example 11.21, to

$$V_{ripple} = I_o \frac{N}{2f \times C} = 10mA \frac{3}{2 \times 50kHz \times 1nF} = 200V$$

♣

11.4.4 Three-phase voltage multipliers

The full-wave multiplier in figure 11.31 is a special case of a poly-phase (0° and 180°) multiplier where more than one multiplier share a common series stack of load capacitors. In figure 11.31, the phase angle between phases is 0°, 120°, and 240°, respectively. The peak voltage supplied by each secondary winding is V_{pk}.
The three-phase circuit in figure 11.31b can be modified by disconnecting the centre point of the Y configuration from ground and omitting the first capacitor in each charging stack, as shown in figure 11.31c. As a result, the open-circuit dc voltage per stage is reduced from $2{\times}V_{pk}$ to $\sqrt{3}{\times}V_{pk}$. The output impedance, however, decreases dramatically, so the output voltage under load may be even higher, depending on the load current. Therefore, this variant is preferred if the multiplier has to supply higher currents.

11.4.5 Series versus parallel voltage multipliers

The theory of operation is the same for both series and parallel connected voltage multipliers. Parallel multipliers require less capacitance per cascaded stage than their series counterparts, however parallel multipliers require higher capacitor voltage ratings on successive cascaded stages. The parallel multiplier output is easier to RC filter in applications requiring low output ripple voltage.

11.5 Marx voltage generator

The Marx generator shown in figure 11.32, charges the energy storage capacitor of each stage in parallel with a relatively low voltage (1kV to 6kV), and then discharges them by means of active switches in series, into the load. The output voltage is then equal to the charging voltage multiplied by the number of stages. The series inductance of this type of generators is low, as a result the rise time and fall time of the output pulses can be less than 1µs. The pulse repetition rate can be more than 20kHz for short pulses, and the pulse length can be several ms.

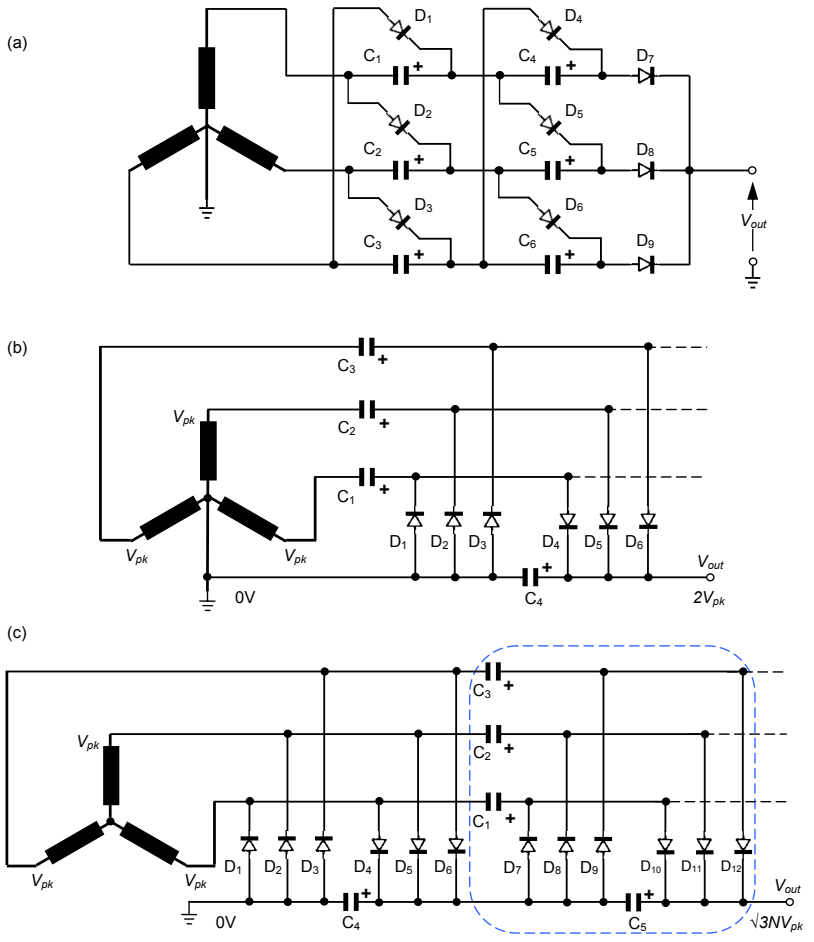


Figure 11.31. *Three-phase Y configuration voltage multipliers: (a) series diode output stage; (b) grounded centre point; and (b) floating centre point.*

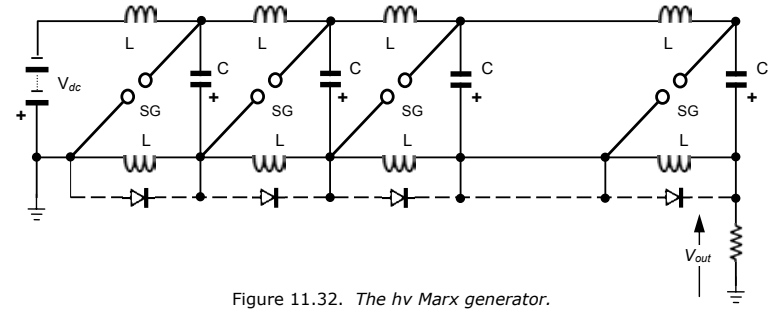


Figure 11.32. *The hv Marx generator.*

11.6 Definitions

The average (or mean or dc) rms (or effective) values, respectively, of a waveform, are defined by

$$V_o = \frac{1}{T}\int_o^T v_o(t)\,dt$$

and

$$V_{rms} = \sqrt{\frac{1}{T}\int_o^T v_o^2(t)\,dt}$$

V_o	average output voltage	$\bar{I}_o$	average output current
V_{rms}	rms output voltage	I_{rms}	rms output current
$\hat{V}$	peak output voltage	$\hat{I}$	peak output current

Load voltage form factor $= FF_v = V_{rms}/V_o$ Load voltage crest factor $= CF_v = \hat{V}/V_{rms}$

Load current form factor $= FF_i = I_{rms}/\bar{I}_o$ Load current crest factor $= CF_i = \hat{I}/I_{rms}$

Waveform smoothness = Ripple factor $= RF_v = \dfrac{\text{effective values of ac } V \text{ (or } I\text{)}}{\text{average value of } V \text{ (or } I\text{)}} = \dfrac{V_{Ri}}{V_o}$

$$= \sqrt{\frac{V_{rms}^2 - V_o^2}{V_o^2}} = \sqrt{FF_v^2 - 1}$$

where $V_{Ri} = \left[\sum_{n=1}^{\infty} \tfrac{1}{2}\left(V_{an}^2 + V_{bn}^2\right)\right]^{1/2}$

similarly the current ripple factor is $RF_i = \dfrac{I_{Ri}}{\bar{I}_o} = \sqrt{FF_i^2 - 1}$

$RF_i = RF_v$ for a resistive load

$$\text{Rectification efficiency} = \eta = \frac{\text{dc load power}}{\text{ac load power + rectifier losses}} = \frac{V_o I_o}{V_{rms} I_{rms} + \text{Loss}_{\text{rectifier}}}$$

Waveform fundamental and harmonic rms components are define by

$$V_1 = \sqrt{1/2\left(V_{1a}^2 + V_{1b}^2\right)}$$

where

$$V_{1a} = \frac{2}{T}\int_o^T v(t)\cos 2\pi\, t/T\, dt$$

$$V_{1b} = \frac{2}{T}\int_o^T v(t)\sin 2\pi\, t/T\, dt$$

and for the k^{th} harmonic component

$$V_k = \sqrt{1/2\left(V_{ka}^2 + V_{kb}^2\right)}$$

where

$$V_{ka} = \frac{2}{T}\int_o^T v(t)\cos 2\pi\, kt/T\, dt$$

$$V_{kb} = \frac{2}{T}\int_o^T v(t)\sin 2\pi\, kt/T\, dt$$

Distortion factor is defined as

$$DF_v = \frac{V_1}{V_{rms}}$$

The total harmonic distortion is

$$THD_v = \sqrt{\sum_{\substack{k=0\\k\neq 1}}^{\infty}\left(\frac{V_k}{V_1}\right)^2} = \frac{\sqrt{1 - DF_v^2}}{DF_v^2}$$

11.7 Output pulse number

Output pulse number p is the number of pulses in the output voltage that occur during one ac input cycle, of frequency f_s. The pulse number p therefore specifies the output harmonics, which occur at p x f_s, and multiples of that frequency, $m \times p \times f_s$, for $m = 1, 2, 3, \ldots$

$$p = \frac{\text{period of input supply voltage}}{\text{period of minimum order harmonic in the output } V \text{ or } I \text{ waveform}}$$

The pulse number p is specified in terms of

q	the number of elements in the commutation group
r	the number of parallel connected commutation groups
s	the number of series connected (phase displaced) commutating groups

Parallel connected commutation groups, r, are usually associated with (and identified by) intergroup reactors (to reduce circulating current), with transformers where at least one secondary is effectively star connected while another is delta connected. The rectified output voltages associated with each transformer secondary, are connected in parallel.

Series connected commutation groups, s, are usually associated with (and identified by) transformers where at least one secondary is effectively star while another is delta connected, with the rectified output associated with each transformer secondary, connected in series.

$q=3\ \ r=2\ \ s=2$
$p = q \times r \times s$
$p = 12$

The mean rectifier output voltage V_o can be specified by

$$V_o = s\frac{q}{\pi}\sqrt{2}V_\phi \times \sin\frac{\pi}{q} \qquad (11.250)$$

For a full-wave, single-phase rectifier, $r = 1$, $q = 2$, and $s = 1$, whence $p = 2$

$$V_o = 1 \times \frac{2}{\pi}\sqrt{2}V_\phi \times \sin\frac{\pi}{2} = \frac{2\sqrt{2}V_\phi}{\pi}$$

For a full-wave, three-phase rectifier, $r = 1$, $q = 3$, and $s = 2$, whence $p = 6$

$$V_o = 2 \times \frac{3}{\pi}\sqrt{2}V_\phi \times \sin\frac{\pi}{3} = \frac{3\sqrt{2}V_\phi}{\pi}$$

11.8 AC-dc converter generalised equations

Alternating sinusoidal voltages

$$V_1 = \sqrt{2}V\sin\omega t$$
$$V_2 = \sqrt{2}V\sin\left(\omega t - \tfrac{2\pi}{q}\right)$$
$$\cdot$$
$$V_q = \sqrt{2}V\sin\left(\omega t - (q-1)\tfrac{2\pi}{q}\right)$$

where q is the number of phases (number of voltage sources)

On the secondary or converter side of any transformer, if the load current is assumed constant I_o then the power factor is determined by the load voltage harmonics.

Voltage form factor

$$FF_v = \frac{V_{rms}}{V_o}$$

whence the voltage ripple factor is

$$RF_v = \frac{1}{V_o}\left[V_{rms}^2 - V_o^2\right]^{1/2} = \left[FF_v^2 - 1\right]^{1/2}$$

The power factor on the secondary side of any transformer is related to the voltage ripple factor by

$$pf = \frac{P_d}{S} = \frac{V_o I_o}{qVI_{rms}} = \frac{1}{\sqrt{RF_v^2 + 1}}$$

On the primary side of a transformer the power factor is related to the secondary power factor, but since the supply is assumed sinusoidal, the power factor is related to the primary current harmonics.
Relationship between current ripple factor and power factor

$$RF_i = \frac{1}{I_1}\sqrt{\sum_{h=3}^{\infty} I_h^2} = \frac{1}{I_1}\sqrt{I_{rms}^2 - I_1^2}$$

$$pf = \frac{I_1}{I_{rms}} = \frac{1}{\sqrt{1+RF_i^2}}$$

The supply power factor is related to the primary power factor and is dependent of the supply connection, star or delta, etc.

Half-wave diode rectifiers [see figures 11.2, 11.10]

Pulse number p=q. Pulse number is the number of sine crests in the output voltage during one input voltage cycle. There are q phases and q diodes and each diode conducts for 2π/q, with q crest (pulses) in the output voltage

Mean voltage

$$V_o = \frac{q}{2\pi}\int_{-\pi/q}^{+\pi/q} \sqrt{2}V\sin\omega t\, d\omega t$$

$$= \frac{q}{\pi}\sqrt{2}V\sin\frac{\pi}{q}$$

RMS voltage

$$V_{rms} = \left[\frac{q}{2\pi}\int_{-\pi/q}^{+\pi/q}\left(\sqrt{2}V\sin\omega t\right)^2 d\omega t\right]^{1/2}$$

$$= \sqrt{2}V\left[1/2 + \frac{q}{4\pi}\sin\frac{2\pi}{q}\right]^{1/2}$$

Normalised peak to peak ripple voltage

$$v_{p-p} = \sqrt{2}V - \sqrt{2}V\cos\frac{\pi}{q}$$

$$V_{n_{p-p}} = \frac{v_{p-p}}{V_o} = \frac{\sqrt{2}V - \sqrt{2}V\cos\frac{\pi}{q}}{\frac{q}{\pi}\sqrt{2}V\sin\frac{\pi}{q}} = \frac{\pi}{q}\,\frac{1-\cos\frac{\pi}{q}}{\sin\frac{\pi}{q}}$$

Voltage form factor

$$FF_v = \frac{V_{rms}}{V_o} = \frac{\left[1/2 + \frac{q}{4\pi}\sin\frac{2\pi}{q}\right]^{1/2}}{\frac{q}{\pi}\sin\frac{\pi}{q}}$$

whence the voltage ripple factor is

$$RF_v = \frac{1}{V_o}\left[V_{rms}^2 - V_o^2\right]^{1/2} = \left[FF_v^2 - 1\right]^{1/2}$$

Diode reverse voltage

$$\hat{V}_{DR} = 2\sqrt{2}V \quad \text{if } q \text{ is even}$$

$$\hat{V}_{DR} = 2\sqrt{2}V\cos\frac{\pi}{2q} \quad \text{if } q \text{ is odd}$$

For a constant load current I_o, diode currents are

$$\hat{I}_D = I_o \qquad \bar{I}_D = \frac{I_o}{q} \qquad I_{D\,rms} = \frac{I_o}{\sqrt{q}}$$

For a constant load current I_o the output power is

$$P_d = V_o I_o$$

The apparent power is

$$S = qVI_{rms}$$

The power factor on the secondary side of any transformer is

$$pf = \frac{P_d}{S} = \frac{V_o I_o}{qVI_{rms}} = \frac{1}{\sqrt{RF_v^2+1}}$$

$$= \frac{\frac{q}{\pi}\sqrt{2}V\sin\frac{\pi}{q}\times I_o}{qV\times I_o\sqrt{\frac{1}{q}}} = \frac{\sqrt{2q}}{\pi}\sin\frac{\pi}{q}$$

The primary side power factor is supply connection and transformer construction dependant.

For two-phase half-wave p=q=2

$$pf_{1\phi,1/2} = \frac{V_o I_o}{V I_o} = \frac{2\sqrt{2}}{\pi} = 0.90$$

For three-phase half wave p=q=3

$$pf_{3\phi,1/2} = \frac{V_o I_o}{3V I_o} = \frac{3\sqrt{3}}{2\pi} = 0.827$$

For six-phase half-wave p=q=6

$$pf_{6\phi,1/2} = \frac{V_o I_o}{3V I_o} == \frac{3}{\pi} = 0.995 \quad \text{(Y conection)}$$

The short circuit ratio (ratio actual s/c current to theoretical s/c current) is

$$K_{s/c} = \frac{q\sqrt{2}V \Big/ \omega L_c}{2\sqrt{2}V \Big/ \omega L_c \sin\frac{\pi}{q}} = \frac{q}{2\sin\frac{\pi}{q}}$$

Commutation overlap angle

$$1 - \cos\mu = \frac{\omega L_c I_o}{\sqrt{2}V\sin\frac{\pi}{q}}$$

The commutation voltage drop

$$v_{com} = \frac{q}{2\pi}\omega L_c I_o \text{ where } 2L_c = L_{s/c}$$

p=q=	$I_{sec\,rms}$	RF_v	$\bar{V}_o$	$\hat{V}_D$	%$V_{p\text{-}p}$	$K_{s/c}$	pf_{sec}	pf_{prim}
2	$I_o/\sqrt{2}$		0.90V	2√2 V	0.157	1	0.636	0.90
3	$I_o/\sqrt{3}$	0.68	1.17V	√6 V	0.604	1.73	0.675	0.827
6	$I_o/\sqrt{6}$	0.31	1.35V	2√2 V	0.140	6	0.55	0.995

For three-phase resistive load, with transformer turns ratio 1:N

$$I_o = \frac{\sqrt{2}V}{R}\frac{3\sqrt{3}}{2\pi} \qquad I_{o\,rms} = \frac{V}{R}\left[\tfrac{1}{3} + \frac{\sqrt{3}}{4\pi}\right]^{1/2}$$

$$FF_{i\,output} = \left[\tfrac{2\pi^2}{27} + \tfrac{\pi}{6\sqrt{3}}\right]^{1/2}$$

$$I_{p\Delta} = \frac{N}{1}\times\frac{V}{R}\left[\tfrac{1}{3} + \frac{\sqrt{3}}{4\pi} - \tfrac{3}{2\pi^2}\right]^{1/2} \qquad I_{L\Delta} = \frac{N}{1}\times\frac{V}{R}\left[\tfrac{2}{3} + \frac{\sqrt{3}}{2\pi}\right]^{1/2}$$

$$I_{LY} = \frac{N}{1}\times\frac{V}{R}\left[\tfrac{2}{9} + \frac{\sqrt{3}}{6\pi}\right]^{1/2}$$

Time domain half-wave single phase R-L-E load

$$i_o(\omega t) = -\frac{E}{R} + \frac{\sqrt{2}V}{Z}\left(\sin(\omega t - \phi) + \left[\frac{E}{R}\frac{Z}{\sqrt{2}V} - \sin(\omega t - \phi)\right]e^{\frac{\omega t-\alpha}{\tan\phi}}\right)$$

$$v_o(\omega t) = V_o\left[1 + \sum_{k=1}^{\infty}\frac{-2(-1)^k}{k^2q^2-1}\cos(kq\omega t)\right]$$

Full-wave diode bridge rectifiers - star [see figures 11.7, 11.12]

q phases and 2q diodes

Mean voltage

$$V_o = \frac{q}{\pi}\int_{\frac{1}{2}\pi-\frac{\pi}{q}}^{\frac{1}{2}\pi+\frac{\pi}{q}} \sqrt{2}V\sin\omega t\, d\omega t$$
$$= \frac{2q}{\pi}\sqrt{2}V\sin\frac{\pi}{q}$$

Pulse number

$p=q$ if q is even
$p=2q$ if q is odd

Diode reverse voltage

$$\hat{V}_{D_R} = 2\sqrt{2}V \quad \text{if } q \text{ is even}$$
$$\hat{V}_{D_R} = 2\sqrt{2}V\cos\frac{\pi}{2q} \quad \text{if } q \text{ is odd}$$

For a constant load current I_o, diode currents are

$$\hat{I}_D = I_o \qquad \bar{I}_D = \frac{I_o}{q} \qquad I_{D\,rms} = \frac{I_o}{\sqrt{q}}$$

The current and power factor are

$$I_{rms} = I_o\sqrt{\frac{2}{q}}$$

$$pf = \frac{P_d}{S} = \frac{V_o I_o}{qVI_{rms}} = \frac{\frac{2q}{\pi}\sqrt{2}V\sin\frac{\pi}{q}\times I_o}{qV\times I_o\sqrt{\frac{2}{q}}} = \frac{2\sqrt{q}}{\pi}\sin\frac{\pi}{q}$$

which is √2 larger than the half-wave case.

For single-phase, full-wave $p=q=2$

$$pf_{1\phi} = \frac{V_o I_o}{V\,I_o} = \frac{2\sqrt{2}}{\pi} = 0.90$$

For three-phase full-wave $p=2$ $q=6$

$$pf_{3\phi} = \frac{V_o I_o}{3VI_o} = \frac{\frac{6}{\pi}\sqrt{2}V\times ½I_o}{3V\times\frac{\sqrt{2}}{3}I_o} = \frac{3}{\pi} = 0.955$$

p, q	I_{sec}	RF_v	$\bar{V}_o$	$\hat{V}_D$	$\%V_{p\text{-}p}$	$K_{s/c}$	$pf_{Y\,prim}$	pf_{sec}
$p=q=2$	I_o	0.483	1.80V	2√2 V	0.157	2/π	0.90	0.90
$p=2$ $q=6$	$\sqrt{⅔}\,I_o$	0.31	2.34V	√6 V	0.140	6/π	0.995	0.995

The short circuit ratio (ratio actual s/c current to theoretical s/c current) is

$$K_{s/c} = \frac{q}{2\pi\sin\frac{\pi}{q}}$$

which is smaller by a factor π than the half-wave case.

$$K_{s/c} = \frac{q}{\pi} \quad \text{for } q = 2$$

Relationship between current ripple factor and supply side power factor on the primary

$$RF_I = \frac{1}{I_1}\sqrt{\sum_{h=3}^{\infty} I_h^2} = \frac{1}{I_1}\sqrt{I_{rms}^2 - I_1^2}$$
$$pf = \frac{I_1}{I_{rms}} = \frac{1}{\sqrt{1+RF_I^2}}$$

For single phase $p=2$

$$RF_I = \frac{1}{I_1}\sqrt{I_{rms}^2 - I_1^2}$$
$$= \frac{\sqrt{I_o^2 - \left(\frac{1}{\sqrt{2}}\frac{4}{\pi}I_o\right)^2}}{\frac{1}{\sqrt{2}}\frac{4}{\pi}I_o} = \sqrt{\frac{\pi^2-8}{8}} = 0.483$$

$$pf = \frac{1}{\sqrt{1+RF_I^2}} = \frac{1}{\sqrt{1+\frac{\pi^2-8}{8}}} = \frac{2\sqrt{2}}{\pi} = 0.90$$

The rms of the fundamental component is

$$I_1 = \frac{1}{\sqrt{2}}\frac{4}{\pi}I_o$$

The rms of the harmonic components are

$$I_h = \frac{I_1}{h} = \frac{I_1}{kp\pm1} \text{ for } k \geq 1,2,3...$$

For p-pulse

$$RF_v = \sqrt{\frac{\frac{\pi^2}{p^2}}{\sin^2\frac{\pi}{p}} - 1}$$

$$pf = \frac{1}{\sqrt{1+RF_v^2}} = \frac{p}{\pi}\sin\frac{\pi}{p}$$

Commutation overlap angle

$$1-\cos\mu = \frac{\omega L_c I_o}{\sqrt{2}V\sin\frac{\pi}{q}}$$

The commutation voltage drop

$$v_{com} = \frac{q}{\pi}\omega L_c I_o \text{ where } 2L_c = L_{s/c}$$

For $p=q=2$, only

$$1-\cos\mu = \frac{2\omega L_c I_o}{\sqrt{2}V}$$
$$v_{com} = \frac{4}{\pi}\omega L_c I_o$$

Load characteristics

$$\text{Current Form Factor} = FF_I = \frac{I_{o\,rms}}{I_o} = \frac{I_o\sqrt{\frac{2}{q}}}{I_o} = \sqrt{\frac{2}{q}}$$

Full-wave diode bridge rectifiers – delta

Same expression as for delta connected secondary, except supply voltages V are replaced by

$$\frac{V}{2\sin\frac{\pi}{q}}$$

For example in three-phase, V is replaced by $V/\sqrt{3}$, that is, $V_{L-L} = \sqrt{3}V_{L-N} = \sqrt{3}V_{phase}$

The mean output voltage is

$$V_o = \frac{2q}{\pi}\sqrt{2}V_\Delta\sin\frac{\pi}{q} = \frac{2q}{\pi}\sqrt{2}\,\frac{V}{2\sin\frac{\pi}{q}}\sin\frac{\pi}{q} = \frac{q}{\pi}\sqrt{2}V$$

Pulse number

$p=q$ if q is even
$p=2q$ if q is odd

diode reverse voltage and currents

$$\hat{V}_{D_R} = \frac{\sqrt{2}V}{\sin\frac{\pi}{q}} \quad \text{if } q \text{ is even}$$

$$\hat{V}_{D_R} = \frac{\sqrt{2}V}{2\sin\frac{\pi}{2q}} \quad \text{if } q \text{ is odd}$$

$$\hat{I}_D = I_o \qquad \bar{I}_D = I_o / q \qquad I_{D\,rms} = I_o / \sqrt{q}$$

rms current and power factor

$$I_{rms\,even} = \frac{I_o}{2} \qquad pf_{q\,even} = \frac{V_o I_o}{qVI_{rms}} = \frac{\frac{q}{\pi}\sqrt{2}V\,I_o}{qV\,½I_o} = \frac{2\sqrt{2}}{\pi}$$

$$I_{rms\,odd} = \frac{I_o}{2}\frac{\left[q^2-1\right]^{½}}{q} \qquad pf_{q\,odd} = \frac{V_o I_o}{qVI_{rms}} = \frac{2\sqrt{2}}{\pi}\frac{q}{\left[q^2-1\right]^{½}}$$

Commutation angle and voltage

$$1-\cos\mu = \frac{\omega L I_o}{\sqrt{2}V} \qquad v_{com} = \frac{q}{2\pi}\omega L_c I_o \qquad q \text{ even}$$

$$1-\cos\mu = \frac{\omega L I_o}{\sqrt{2}V}\left(1-\frac{1}{q}\right) \qquad v_{com} = \frac{q}{2\pi}\omega L_c I_o\left(1-\frac{1}{q}\right) \qquad q \text{ odd}$$

The short circuit ratio (ratio actual s/c current to theoretical s/c current) is

$$K_{s/c\,even} = \frac{q}{\pi}\sin\frac{\pi}{q} \qquad K_{s/c\,odd} = \frac{q-1}{\pi}\sin\frac{\pi}{q}$$

For single-phase resistive load, with transformer turns ratio 1:N

$$I_o = \frac{\sqrt{2}V}{R}\frac{4}{\pi} \qquad I_{o\,rms} = \frac{2V}{R}$$

$$FF_{i\,output} = \frac{\pi}{2\sqrt{2}} \qquad RF_v = \sqrt{FF^2-1} = \sqrt{\frac{\pi^2}{8}-1}$$

$$I_p = \frac{N}{1}I_{sec} = \frac{N}{1}\times\frac{2V}{R} \qquad pf = \frac{1}{\sqrt{RF^2+1}} = \frac{2\sqrt{2}}{\pi}$$

Reading list

Dewan, S. B. and Straughen, A., *Power Semiconductor Circuits,* John Wiley and Sons, New York, 1975.

Sen, P.C., *Power Electronics*, McGraw-Hill, 5th reprint, 1992.

Shepherd, W *et al*. *Power Electronics and motor control*, Cambridge University Press, 2nd Edition 1995.

http://www.ipes.ethz.ch/

http://www.celnav.de/hv/hv9.htm

http://www.physiqueindustrie.com/custom_converter.php

http://www.voltagemultipliers.com/html/multdesign.html

Problems

11.1. Derive equations (11.35) and (11.36) for the circuit in figure 11.5.

11.2. Assuming a constant load current, derive an expression for the mean and rms device current and the device form factor, for the circuits in figure 11.7.

11.3. The single-phase full-wave uncontrolled rectifier is operated from the 415 V line-to-line voltage, 50 Hz supply, with a series load of 10 Ω + 5 mH + 40 V battery. Derive the load voltage expression in terms of a Fourier series. Determine the rms value of the fundamental of the load current.

11.4. A single-phase uncontrolled rectifier has a 24Ω resistive load a 240V ac 50Hz supply. Determine the average, peak and rms current and peak reverse voltage across each rectifier diode for
- i. an isolating transformer with a 1:1 turns ratio
- ii. centre-tapped transformer with turns ratio 1:1:1.

11.5. A single-phase bridge rectifier has an *R-L* of R = 20Ω and L = 50mH and a 240V ac 50Hz source voltage. Determine:
- i. the average and rms currents of the diodes and load
- ii. rms and average 50Hz source currents
- iii. the power absorbed by the load
- iv. the supply power factor

11.6. A single-phase, full-wave uncontrolled rectifier has a back emf E_b in its load. If the supply is 240Vac 50Hz and the series load is R = 20Ω, L = 50mH, and E_b = 120V dc, determine:
- i. the power absorbed by the dc source in the load
- ii. the power absorbed by the load resistor
- iii. the power delivered from the ac source
- iv. the ac source power factor
- v. the peak-to-peak load current variation if only the first ac term of the Fourier series for the load current is considered.

11.7. A three-phase uncontrolled rectifier is supplied from a 50Hz 415V ac line-to-line voltage source. If the rectifier load is a 75 Ω resistor, determine
- i. the average load current
- ii. the rms load current
- iii. the rms source current
- iv. the supply power factor.

11.8. A three-phase uncontrolled rectifier is supplied from a 50Hz 415V ac line-to-line voltage source. If the rectifier load is a series *R-L* circuit where R = 10Ω and L = 100mH, determine:
- i. the average and rms load currents
- ii. the average and rms diode currents
- iii. the rms source and power current
- iv. the supply power factor.

12

Naturally Commutating AC to DC Converters

- Controlled Rectifiers

The converter circuits considered in this chapter have in common an ac voltage supply input and a dc load output. The function of the converter circuit is to convert the ac source energy into controllable dc load power, mainly for highly inductive loads. Turn-off of converter semiconductor devices is brought about by the ac supply voltage reversal, a process called *line commutation* or *natural commutation.*

Converter circuits employing only diodes are termed *uncontrolled* (or *rectifiers*) while the incorporation of only thyristors results in a (fully) *controlled converter.* The functional difference is that the diode conducts when forward-biased whereas the turn-on of the forward-biased thyristor can be controlled from its gate. An uncontrolled converter provides a fixed output voltage for a given ac supply and load.

Converters employing a combination of both diodes and thyristors are generally termed *half-controlled* (or *semi-controlled*). Both fully controlled and half-controlled converters allow an adjustable output voltage by controlling the phase angle at which the forward biased thyristors are turned on. The polarity of the output (load) voltage of a fully controlled converter can reverse (but the current flow direction is not reversible), allowing power flow into the supply, a process called *inversion.* Thus a fully controlled converter can be described as a *bidirectional converter* as it facilitates power flow in either direction.

The half-controlled converter, as well as the uncontrolled converter, contains diodes which prevent the output voltage from going negative. Such converters only allow power flow from the ac supply to the dc load, termed *rectification,* and can therefore be described as *unidirectional converters.*

Although all these converter types provide a dc output, they differ in characteristics such as output ripple and mean voltage as well as efficiency and ac supply harmonics.

An important converter characteristic is that of pulse number, which is defined as the repetition rate in the direct output voltage during one complete cycle of the input ac supply.

A useful way to judge the quality of the required dc output, is by the contribution of its superimposed ac harmonics. The harmonic or ripple factor *RF* is defined by

$$RF_v = \sqrt{\frac{V_{rms}^2 - V_{dc}^2}{V_{dc}^2}} = \sqrt{\frac{V_{rms}^2}{V_{dc}^2} - 1} = \sqrt{FF^2 - 1}$$

where *FF* is termed the form factor. RF_v is a measure of the voltage harmonics in the output voltage while if currents are used in the equation, RF_i gives a measure of the current harmonics in the output current. Both *FF* and *RF* are applicable to the input and output, and are fully defined in section 12.8.

The general analysis in this chapter is concerned with single and three phase ac supplies mainly feeding inductive dc loads. A load dc back emf is used in modelling the dc machine. Generally, uncontrolled rectifier equations can be derived from the corresponding controlled converter circuit equations by setting the controlled *delay angle* α to zero. Also purely resistive load equations generally can be derived by setting inductance *L* to zero in the *L*-*R* load equations and *R*-*L* load equations can be derived from *R*-*L*-*E* equations by setting *E*, the load back emf, to zero.

12.1 Single-phase full-wave half-controlled converter

12.1.1 Single-phase, full-wave half-controlled circuit with an R-L load

When a converter contains both diodes and thyristors, for example as shown in figure 12.1 parts a to d, the converter is termed half-controlled (or semi-controlled). These four circuits produce identical load and supply waveforms, neglecting any differences in the number and type of semiconductor voltage drops. The power to the load is varied by controlling the angle *α*, shown in figure 12.1e, at which the bridge thyristors are triggered (after first becoming forward biased). The circuit diodes prevent the load voltage from going negative, extend the conduction period, and reduce the output ac ripple.

The particular application will determine which one of the four circuits should be employed. For example, circuit figure 12.1a contains five devices of which four are thyristors, whereas the other circuits contain fewer devices, of which only two are thyristors. The circuit in figure 12.1b uses the fewest semiconductors, but requires a transformer which introduces extra cost, weight, and size. Also the thyristors experience twice the voltage of the thyristors in the other circuits, $2\sqrt{2}\ V$ rather than $\sqrt{2}\ V$. The transformer does provide isolation and voltage matching.

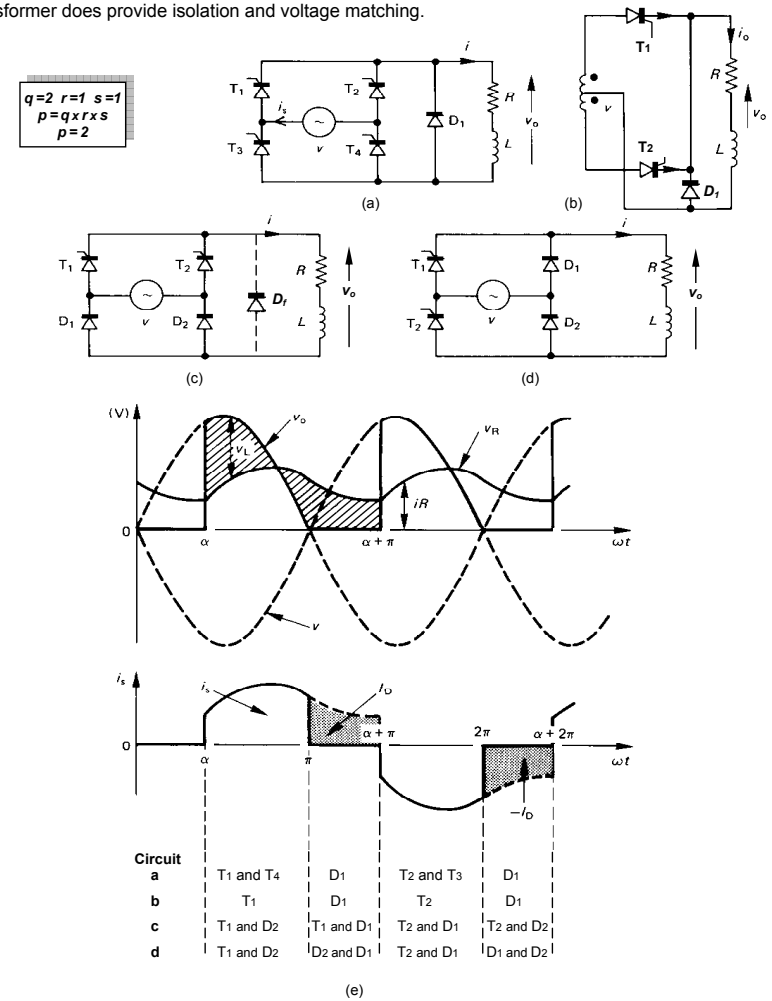


Figure 12.1. *Full-wave half-controlled converters with freewheel diodes: (a), (b), (c), and (d) different circuit configurations producing the same output; and (e) circuit voltage and current waveforms and device conduction table.*

The thyristor triggering requirements of the circuits in figures 12.1b and c are simple since both thyristors have a common cathode connection. Figure 12.1c may suffer from prolonged shut-down times with highly inductive loads. The diode in the freewheeling path will hold on the freewheeling thyristor, allowing conduction during that thyristors next positive cycle without any gate drive present. The extra diode D_f in figure 12.1c bypasses the bridge thyristors allowing them to drop out of conduction. This is achieved at the expense of an extra device, but the freewheel path conduction losses are decreased since that series circuit now involves only one semiconductor voltage drop. This continued conduction problem does not occur in circuits 12.1a and d since freewheeling does not occur through the circuit thyristors, hence they will drop out of conduction at converter shut-down. The table in figure 12.1e shows which semiconductors are active in each circuit during the various periods of the load cycle.

Circuit waveforms are shown in figure 12.1e. Since the load is a passive *L-R* circuit, independent of whether the load current is continuous or discontinuous, the mean output voltage and current (neglecting diode voltage drops) are

$$V_o = \bar{I}_o R = \frac{1}{\pi}\int_{\alpha}^{\pi} \sqrt{2}\, V \sin(\omega t)\; d\omega t = \frac{\sqrt{2}\, V}{\pi}(1+\cos\alpha) \qquad \text{(V)}$$
$$\bar{I}_o = {}^{V_o}\!/_{R} = \frac{\sqrt{2}\, V}{\pi R}(1+\cos\alpha) \qquad \text{(A)} \tag{12.1}$$

where α is the delay angle from the point at which the associated thyristor first becomes forward-biased and is therefore able to be turned on and conduct current. The maximum mean output voltage, $\hat{V}_o = 2\sqrt{2}V/\pi$ (also predicted by equation 11.54), occurs at $\alpha = 0$.
The normalised mean output voltage V_n is

$$V_n = V_o / \hat{V}_o = \tfrac{1}{2}(1+\cos\alpha) \tag{12.2}$$

The Fourier coefficients of the 2-pulse output voltage are given by equation (12.128). For the single-phase, full-wave, half-controlled case, $p = 2$, thus the output voltage harmonics occur at $n = 2, 4, 6, \ldots$

Equation (12.1) shows that the load voltage is independent of the passive load (because the diodes clamp the load to zero volts thereby preventing the load voltage from going negative), and is a function only of the phase delay angle for a given supply voltage.
The rms value of the load circuit voltage v_o is

$$V_{rms} = \sqrt{\frac{1}{\pi}\int_{\alpha}^{\pi} \left(\sqrt{2}V\sin\omega t\right)^2 d\omega t} = V\sqrt{\frac{\pi-\alpha+\tfrac{1}{2}\sin 2\alpha}{\pi}} \qquad \text{(V)} \tag{12.3}$$

From the load voltage definitions in section 12.7, the load voltage form factor is

$$FF_v = \frac{V_{rms}}{V_o} = \frac{\sqrt{\pi\left(\pi-\alpha+\tfrac{1}{2}\sin 2\alpha\right)}}{\sqrt{2}\left(1+\cos\alpha\right)} \tag{12.4}$$

The ripple voltage is

$$V_{Ri} \triangleq \sqrt{V_{rms}^2 - V_o^2} \tag{12.5}$$

hence the voltage ripple factor RF_v is

$$RF_v \triangleq V_{Ri}/V_o = \sqrt{FF_v^2 - 1} \tag{12.6}$$

The load and supply waveforms and equations, for continuous and discontinuous load current, are the same for all the circuits in figure 12.1. The circuits differ in the device conduction paths as shown in the table in figure 12.1e. After deriving the general load current equations, the current equations applicable to the different circuit devices can be decoded.

12.1.1i - Discontinuous load current, with $\alpha < \pi$ and $\beta - \alpha < \pi$, the load current (and supply current) is based on equation 11.14 namely

$$i(\omega t) = i_s(\omega t) = {}^{\sqrt{2}\,V}\!/_{Z}\left(\sin(\omega t-\phi) - \sin(\alpha-\phi)e^{-\omega t+\alpha/\tan\phi}\right) \qquad \text{(A)}$$
$$\alpha \le \omega t \le \pi \tag{12.7}$$

where $Z = \sqrt{R^2 + (\omega L)^2}$ and $\phi = \tan^{-1} {}^{\omega L}\!/_{R}$

After $\omega t = \pi$ the load current decreases exponentially to zero through the freewheel diode according to

$$i(\omega t) = i_{Df}(\omega t) = I_{01\pi}e^{-\omega t/\tan\phi} \qquad \text{(A)} \qquad 0 \le \omega t \le \alpha \tag{12.8}$$

where for $\omega t = \pi$ in equation (12.7)

$$I_{o1\pi} = {}^{\sqrt{2}\,V}\!/_{Z}\,\sin(\phi-\alpha)(1-e^{-\pi/\tan\phi})$$

The various semiconductor average current ratings can be determined from the average half-cycle freewheeling current, $\bar{I}_{½F}$, and the average half-cycle supply current, $\bar{I}_{½s}$. For discontinuous load current

$$\bar{I}_{½F} = \tfrac{1}{2}\frac{\sqrt{2}\,V}{\pi R}\sin\phi\left(\sin\phi - \sin(\alpha-\phi)e^{(\alpha-\pi)/\tan\phi}\right) \tag{12.9}$$

$$\bar{I}_{½s} = \tfrac{1}{2}\bar{I}_o - \bar{I}_{½F} = \tfrac{1}{2}\frac{\sqrt{2}\,V}{\pi R}\left(\cos^2\phi + \cos\alpha + \sin\phi\sin(\alpha-\phi)e^{(\alpha-\pi)/\tan\phi}\right) \tag{12.10}$$

12.1.1ii - Continuous load current, with $\alpha < \phi$ and $\beta - \alpha \ge \pi$, the load current is given by equations similar to equations 11.20 and 11.21, specifically

$$i(\omega t) = i_s(\omega t) = {}^{\sqrt{2}\,V}\!/_{Z}\left(\sin(\omega t-\phi) + \frac{\sin\phi\, e^{-\alpha/\tan\phi} - \sin(\alpha-\phi)}{1-e^{-\pi/\tan\phi}}e^{-\omega t+\alpha/\tan\phi}\right)$$
$$\alpha \le \omega t \le \pi \qquad \text{(A)} \tag{12.11}$$

while the load current when the freewheel diode conducts is

$$i(\omega t) = i_{Df}(\omega t) = I_{01\pi}e^{-\omega t/\tan\phi} \qquad \text{(A)}$$
$$0 \le \omega t \le \alpha \tag{12.12}$$

where, for $\omega t = \pi$ in equation (12.11)

$$I_{01\pi} = {}^{\sqrt{2}\,V}\!/_{Z}\;\frac{\sin\phi - \sin(\alpha-\phi)e^{-\pi+\alpha/\tan\phi}}{1-e^{-\pi/\tan\phi}} \qquad \text{(A)}$$

The various semiconductor average current ratings can be determined from the average half cycle freewheeling current, $\bar{I}_{½F}$, and the average half cycle supply current, $\bar{I}_{½s}$. For continuous load current

$$\bar{I}_{½F} = \tfrac{1}{2}\frac{\sqrt{2}\,V}{\pi R}\sin\phi\,\frac{\sin\phi - \sin(\alpha-\phi)e^{-\pi+\alpha/\tan\phi}}{1-e^{-\pi/\tan\phi}}\left(1-e^{-\alpha/\tan\phi}\right) \tag{12.13}$$

$$\bar{I}_{½s} = \tfrac{1}{2}\bar{I}_o - \bar{I}_{½F}$$
$$= \tfrac{1}{2}\frac{\sqrt{2}\,V}{\pi R}\cos\phi\left(\tan\phi\,\frac{1-e^{-(\pi+\alpha)/\tan\phi}}{1-e^{-\pi/\tan\phi}}\left(e^{-\alpha/\tan\phi}\sin\phi - \sin(\alpha-\phi)\right) + \cos\phi + \cos(\alpha-\phi)\right) \tag{12.14}$$

Table 12.1: Semiconductor average current ratings

Bridge circuit figure 12.1	Number of devices	Average device current	
		Thyristor	Diode
a	**4T+1D**	**1×** $\bar{I}_{½s}$	**2×** $\bar{I}_{½F}$
b	**2T+1D**	**1×** $\bar{I}_{½s}$	**2×** $\bar{I}_{½F}$
c	**2T+2D**	**½×** $\bar{I}_o$	**½×** $\bar{I}_o$
d	**2T+2D**	**1×** $\bar{I}_{½s}$	**1×** $\bar{I}_{½s}$ **+ 2×** $\bar{I}_{½F}$

The device conduction table in figure 12.1e can be used to specify average devices currents, for both continuous and discontinuous load current for each of the circuits in figure 12.1, parts a to d.
For a highly inductive load, constant load current, the supply power factor is $pf = {}^{2}\!/_{\pi}\sqrt{2}\cos\alpha$.

Critical load inductance
The critical load inductance, to prevent the current falling to zero, is given by

$$\frac{\omega L_{crit}}{R} = \theta - \alpha - \tfrac{1}{2}\pi + \frac{\alpha + \sin\alpha + \pi\cos\theta}{1+\cos\alpha} \tag{12.15}$$

for $\alpha \le \theta$ where

$$\theta = \sin^{-1}\frac{V_o}{\sqrt{2}V} = \sin^{-1}\frac{1+\cos\alpha}{\pi} \tag{12.16}$$

The minimum current occurs at the angle θ, where the mean output voltage V_o equals the instantaneous load voltage, v_o. When the phase delay angle α is greater than the critical angle θ, $\theta = \alpha$ in equation (12.16) yields (see figure 12.14)

$$\frac{\omega L_{crit}}{R} = -\tfrac{1}{2}\pi + \frac{\alpha + \sin\alpha + \pi\cos\alpha}{1+\cos\alpha} \tag{12.17}$$

It is important to note that converter circuits employing diodes cannot be used when inversion is required. Since the converter diodes prevent the output voltage from being negative, (and the current is unidirectional), regeneration from the load into the supply is not achievable.

Figure 12.1a is a fully controlled converter with an *R-L* load and freewheel diode. In single-phase circuits, this converter essentially behaves as a half-controlled converter.

12.1.2 Single-phase, full-wave, half-controlled circuit with R-L and emf load, E

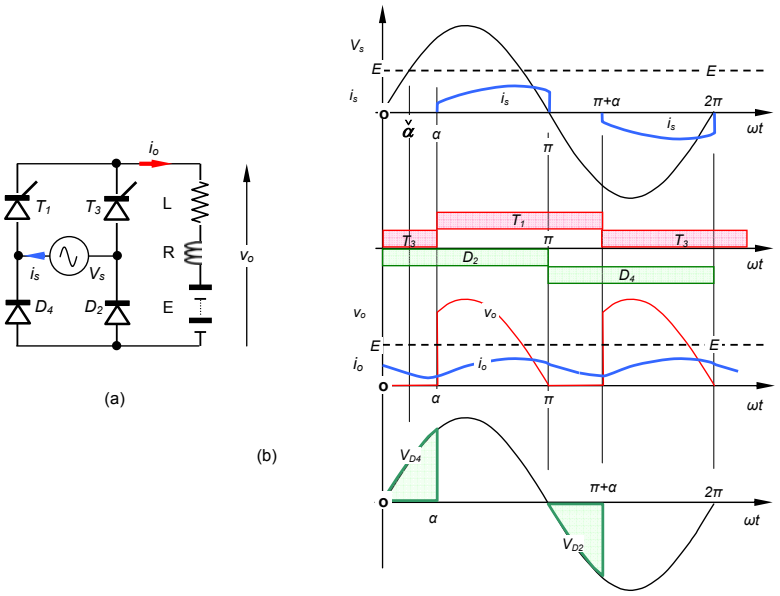


Figure 12.2. *Full-wave half-controlled converter with freewheel diodes and back emf: (a) circuit configurations and (b) circuit voltage and current waveforms and device conduction table.*

In figure 12.2a, with a load back emf, current begins to flow when the supply instantaneous voltage exceeds the back emf magnitude *E*, that is when

$$\check{\alpha} = \sin^{-1}\frac{E}{\sqrt{2}V} \tag{12.18}$$

The average load voltage and current are given by

$$V_o = \frac{1}{\pi}\int_{\alpha}^{\pi}\sqrt{2}V\sin(\omega t)\,d\omega t = \frac{\sqrt{2}V}{\pi}(1+\cos\alpha) \quad \text{(V)}$$
$$\bar{I}_o = \frac{V_o - E}{R} = \frac{\sqrt{2}V}{\pi R}(1+\cos\alpha - \pi\sin\check{\alpha}) \quad \text{(A)} \tag{12.19}$$

The time domain solution for the load current has two components.

In the period $\alpha \le \omega t \le \pi$
When current flows, Kirchhoff's voltage law gives

$$v_o(t) = \sqrt{2}V\sin\omega t = Ri + L\frac{di}{dt} + E \tag{12.20}$$

Assuming **continuous current conduction**, using $R = Z\cos\Phi$ and $E = \sqrt{2}V\sin\check{\alpha}$, which yields

$$i(\omega t) = I_1 e^{-\omega t+\alpha/\tan\phi} + \frac{\sqrt{2}V}{Z}\left(\sin(\omega t-\phi) - \frac{E}{\sqrt{2}V}\right)$$
$$= I_1 e^{-\omega t+\alpha/\tan\phi} + \frac{\sqrt{2}V}{Z}\left(\sin(\omega t-\phi) - \frac{\sin\check{\alpha}}{\cos\phi}\right) \tag{12.21}$$
$$\alpha \ge \check{\alpha} \text{ and } i(\omega t=\pi) > 0$$

where $Z = \sqrt{R^2 + (\omega L)^2}$ and $\phi = \tan^{-1}\omega L/R$

In the period $\pi \le \omega t \le \pi+\alpha$

$$v_o(t) = 0 = Ri + L\frac{di}{dt} + E \tag{12.22}$$

Solving gives

$$i(\omega t) = I_1 e^{-\omega t+\alpha/\tan\phi} + \frac{\sqrt{2}V}{Z}\left(e^{-\omega t+\pi/\tan\phi}\sin\phi - \frac{E}{\sqrt{2}V}\right)$$
$$= I_1 e^{-\omega t+\alpha/\tan\phi} + \frac{\sqrt{2}V}{Z}\left(e^{-\omega t+\pi/\tan\phi}\sin\phi - \frac{\sin\check{\alpha}}{\cos\phi}\right) \tag{12.23}$$
$$\alpha \ge \check{\alpha} \text{ and } i(\omega t=\pi) > 0$$

For continuous current, satisfying continuous and periodic boundary conditions

$$I_1 = \frac{\sqrt{2}V}{Z}\frac{\left(\sin(\phi-\alpha) - e^{-\alpha/\tan\phi}\sin\phi\right)}{1-e^{-\pi/\tan\phi}}$$

For $\alpha \le \omega t \le \pi$

$$i(\omega t) = \frac{\sqrt{2}V}{Z}\left\{\frac{e^{-\omega t+\alpha/\tan\phi}}{1-e^{-\pi/\tan\phi}}\left(\sin(\phi-\alpha) + e^{-\alpha/\tan\phi}\sin\phi\right) + \sin(\omega t-\phi) - \frac{\sin\check{\alpha}}{\cos\phi}\right\} \tag{12.24}$$

For $\pi \le \omega t \le \pi+\alpha$

$$i(\omega t) = \frac{\sqrt{2}V}{Z}\left\{\frac{e^{-\omega t+\alpha/\tan\phi}}{1-e^{-\pi/\tan\phi}}\left(\sin(\phi-\alpha) + e^{-\alpha/\tan\phi}\sin\phi\right) + e^{-\omega t+\pi/\tan\phi}\sin\phi - \frac{\sin\check{\alpha}}{\cos\phi}\right\} \tag{12.25}$$

Discontinuous conduction
At $\omega t \le \alpha$ equation (12.24) must be greater than zero for continuous conduction, that is

$$\frac{\sin(\phi-\alpha) + e^{-\alpha/\tan\phi}\sin\phi}{1-e^{-\pi/\tan\phi}} + \sin(\alpha-\phi) - \frac{\sin\check{\alpha}}{\cos\phi} \ge 0$$

That is

$$\frac{e^{-\pi/\tan\phi}\sin(\phi-\alpha) + e^{-\alpha/\tan\phi}\sin\phi}{1-e^{-\pi/\tan\phi}} - \frac{\sin\check{\alpha}}{\cos\phi} \ge 0 \tag{12.26}$$

Two discontinuous conduction conditions exist:

i. The current is forced to zero before load freewheeling, when *E* exceeds the instantaneous source voltage, that is $\pi - \check{\alpha} < \beta \le \pi$
ii. The current is forced to zero during load freewheeling, that is $\pi < \beta \le \pi+\alpha$

In both cases average output current is

$$\bar{I}_o = \frac{V_o - E}{R}$$

i. In the first case
The average output voltage is

$$V_o = \frac{1}{\pi}\left[\int_{\alpha}^{\beta}\sqrt{2}V\sin\omega t\,d\omega t + \int_{\pi}^{\pi+\alpha}\sqrt{2}V\sin\check{\alpha}\,d\omega t\right] \qquad \beta \le \pi$$
$$V_o = \frac{\sqrt{2}V}{\pi}\left[-\cos\beta + \cos\alpha + \alpha\sin\check{\alpha}\right] \tag{12.27}$$

Hence

$$\bar{I}_o = \frac{V_o - E}{R} = \frac{\sqrt{2}V}{\pi R}\left[-\cos\beta + \cos\alpha + (\alpha-\pi)\sin\check{\alpha}\right] \tag{12.28}$$

The rms output voltage is

$$V_{o\,rms} = \frac{1}{\pi}\left[\int_{\alpha}^{\beta}\left(\sqrt{2}V\sin\omega t\right)^2 d\,\omega t + \int_{\pi}^{\pi+\alpha}\left(\sqrt{2}V\sin\overset{\vee}{\alpha}\right)^2 d\,\omega t\right] \qquad (12.29)$$

$$= \frac{\sqrt{2}V}{\pi}\left[\tfrac{1}{2}(\beta-\alpha) + \tfrac{1}{4}\sin 2\alpha - \tfrac{1}{4}\sin 2\beta + \alpha\right]^{1/2}$$

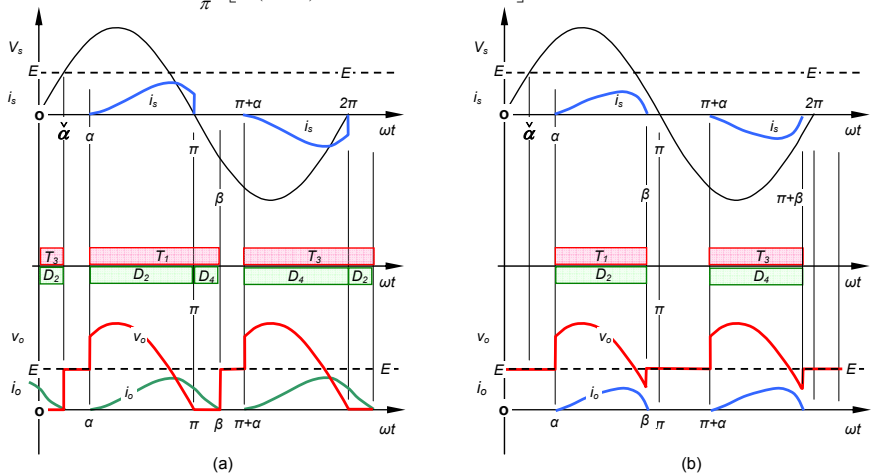


Figure 12.3. *Full-wave half-controlled converter with freewheel diodes and back emf during discontinuous conduction: (a) π-$\overset{\vee}{\alpha}$ < β ≤ π and (b) π < β ≤ π+α.*

In each case the rms output current can be derived from the time current equations, which are used to find the current extinction angle *β*.

For $\alpha \le \omega t \le \beta \le \pi$

$$v_o(t) = \sqrt{2}V\sin\omega t = Ri + L\frac{di}{dt} + E \qquad (12.30)$$

which has the general solution

$$i(\omega t) = I_o e^{-\omega t+\alpha/\tan\phi} + \frac{\sqrt{2}V}{Z}\left(\sin(\omega t-\phi) - \frac{\sin\overset{\vee}{\alpha}}{\cos\phi}\right)$$

where a zero initial current boundary condition gives

$$I_o = \frac{\sqrt{2}V}{Z}\left(\sin(\phi-\alpha) + \frac{\sin\overset{\vee}{\alpha}}{\cos\phi}\right)$$

That is

$$i(\omega t) = \frac{\sqrt{2}V}{Z}\left[\left(\sin(\phi-\alpha) + \frac{\sin\overset{\vee}{\alpha}}{\cos\phi}\right)e^{-\omega t+\alpha/\tan\phi} + \sin(\omega t-\phi) - \frac{\sin\overset{\vee}{\alpha}}{\cos\phi}\right] \qquad (12.31)$$

At $\omega t = \beta \le \pi$, $i(\beta) = 0$, that is *β* is found iteratively from

$$\left(\sin(\phi-\alpha) + \frac{\sin\overset{\vee}{\alpha}}{\cos\phi}\right)e^{-\beta+\alpha/\tan\phi} + \sin(\beta-\phi) = \frac{\sin\overset{\vee}{\alpha}}{\cos\phi} \qquad (12.32)$$

ii. In the second case:
The average output voltage is

$$V_o = \frac{1}{\pi}\left[\int_{\alpha}^{\pi}\sqrt{2}V\sin\omega t\,d\,\omega t + \int_{\beta}^{\pi+\alpha}\sqrt{2}V\sin\overset{\vee}{\alpha}\,d\,\omega t\right] \qquad \beta \ge \pi$$

$$V_o = \frac{\sqrt{2}V}{\pi}\left[1 + \cos\alpha + (\pi+\alpha-\beta)\sin\overset{\vee}{\alpha}\right] \qquad (12.33)$$

hence

$$\bar{I}_o = \frac{V_o - E}{R} = \frac{\sqrt{2}V}{\pi R}\left[1 + \cos\alpha + (\alpha-\beta)\sin\overset{\vee}{\alpha}\right] \qquad (12.34)$$

The rms output voltage is

$$V_{o\,rms} = \frac{1}{\pi}\left[\int_{\alpha}^{\pi}\left(\sqrt{2}V\sin\omega t\right)^2 d\,\omega t + \int_{\beta}^{\pi+\alpha}\left(\sqrt{2}V\sin\overset{\vee}{\alpha}\right)^2 d\,\omega t\right] \qquad (12.35)$$

$$= \frac{\sqrt{2}V}{\pi}\left[\tfrac{1}{2}(\pi-\alpha) + \tfrac{1}{4}\sin 2\alpha + (\pi+\alpha-\beta)\sin^2\overset{\vee}{\alpha}\right]^{1/2}$$

For $\alpha \le \omega t \le \pi$

$$v_o(t) = \sqrt{2}V\sin\omega t = Ri + L\frac{di}{dt} + E \qquad (12.36)$$

yielding

$$i(\omega t) = I_o e^{-\omega t+\alpha/\tan\phi} + \frac{\sqrt{2}V}{Z}\left(\sin(\omega t-\phi) - \frac{\sin\overset{\vee}{\alpha}}{\cos\phi}\right)$$

where a zero initial current boundary condition gives

$$I_o = \frac{\sqrt{2}V}{Z}\left(\sin(\phi-\alpha) + \frac{\sin\overset{\vee}{\alpha}}{\cos\phi}\right)$$

That is

$$i(\omega t) = \frac{\sqrt{2}V}{Z}\left[\left(\sin(\phi-\alpha) + \frac{\sin\overset{\vee}{\alpha}}{\cos\phi}\right)e^{-\omega t+\alpha/\tan\phi} + \sin(\omega t-\phi) - \frac{\sin\overset{\vee}{\alpha}}{\cos\phi}\right] \qquad (12.37)$$

For $\pi \le \omega t \le \beta$

$$v_o(t) = 0 = Ri + L\frac{di}{dt} + E \qquad (12.38)$$

which yields

$$i(\omega t) = I_1 e^{-\omega t+\pi/\tan\phi} - \frac{\sqrt{2}V}{Z}\frac{\sin\overset{\vee}{\alpha}}{\cos\phi}$$

where for continuous current at the boundary

$$I_1 = \frac{\sqrt{2}V}{Z}\left(\sin\phi + e^{\alpha-\pi/\tan\phi}\left(\sin(\phi-\alpha) + \frac{\sin\overset{\vee}{\alpha}}{\cos\phi}\right)\right) \qquad (12.39)$$

That is

$$i(\omega t) = \frac{\sqrt{2}V}{Z}\left[\left(\sin(\phi-\alpha) + \frac{\sin\overset{\vee}{\alpha}}{\cos\phi}\right)e^{-\omega t+\alpha/\tan\phi} + e^{-\omega t+\pi/\tan\phi}\sin\phi - \frac{\sin\overset{\vee}{\alpha}}{\cos\phi}\right] \qquad (12.40)$$

Equating to zero at the conduction extinction angle *β* gives

$$\left(\sin(\phi-\alpha) + \frac{\sin\overset{\vee}{\alpha}}{\cos\phi}\right)e^{-\beta+\alpha/\tan\phi} + e^{-\beta+\pi/\tan\phi}\sin\phi - \frac{\sin\overset{\vee}{\alpha}}{\cos\phi} = 0$$

$$\beta = \tan\phi \times \ell n\left\{\frac{\left(\sin(\phi-\alpha) + \frac{\sin\overset{\vee}{\alpha}}{\cos\phi}\right)e^{\alpha/\tan\phi} + e^{\pi/\tan\phi}\sin\phi}{\frac{\sin\overset{\vee}{\alpha}}{\cos\phi}}\right\} \qquad (12.41)$$

12.2 Single-phase controlled thyristor converter circuits

12.2.1 Single-phase, half-wave controlled circuit with an R-L load

The rectifying diode in the circuit of figure 11.1 can be replaced by a thyristor as shown in figure 12.4a to form a half-wave controlled rectifier circuit with an *R-L* load. The output voltage is now controlled by the thyristor trigger angle, *α*. The output voltage ripple is at the supply frequency. Circuit waveforms are shown in figure 12.4b, where the load inductor voltage equal areas are shaded.

The output current, hence output voltage, for the series circuit are given by

$$v_o(t) = L\frac{di}{dt} + Ri = \sqrt{2}V\sin\omega t \qquad \text{(V)}$$
$$\alpha \le \omega t \le \beta \qquad \text{(rad)} \qquad (12.42)$$

where phase delay angle *α* and current extinction angle *β* are shown in the waveform in figure 12.4b and are the zero load (and supply) current points.

Solving equation (12.42) yields the load and supply current

$$i(\omega t) = \frac{\sqrt{2}\,V}{Z}\{\sin(\omega t - \phi) - \sin(\alpha - \phi)e^{(\alpha - \omega t)/\tan\phi}\} \qquad \text{(A)}$$

$$\text{where } Z = \sqrt{R^2 + (\omega L)^2} \qquad \text{(ohms)} \qquad \alpha \le \omega t \le \beta \qquad (12.43)$$
$$\tan\phi = \omega L / R$$

and *i* is zero elsewhere.

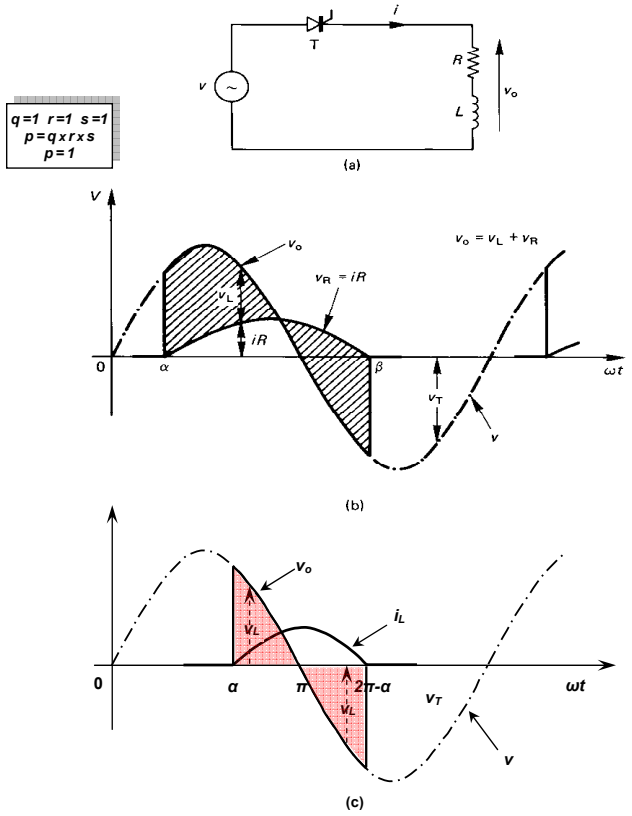


Figure 12.4. *Single-phase half-wave controlled converter: (a) circuit diagram; (b) circuit waveforms for an R-L load; and (c) purely inductive load.*

The current extinction angle *β* is dependent on the load impedance and thyristor trigger angle *α*, and can be determined by solving equation (12.43) with *ωt* = *β* when *i(β)* = 0, that is

$$\sin(\beta - \phi) = \sin(\alpha - \phi)\, e^{(\alpha - \beta)/\tan\phi} \qquad (12.44)$$

This is a transcendental equation. A family of curves of current conduction angle versus delay angle, that is *β* - *α* versus *α*, is shown in figure 12.5a. The straight line plot for $\phi = \frac{1}{2}\pi$ is for a purely inductive load, whereas *ø* = 0 is a straight line for a purely resistive load.

The mean load voltage, whence the mean load current, is given by

$$V_o = \frac{1}{2\pi}\int_\alpha^\beta \sqrt{2}\,V\sin\omega t\, d\omega t$$
$$V_o = \bar{I}_o R = \frac{\sqrt{2}\,V}{2\pi}(\cos\alpha - \cos\beta) \qquad \text{(V)} \qquad (12.45)$$

where the angle *β* can be extracted from figure 12.5a.

The rms load voltage is

$$V_{rms} = \left[\tfrac{1}{2\pi}\int_\alpha^\beta \left(\sqrt{2}\,V\right)^2 \sin^2\omega t\, d\omega t\right]^{\frac{1}{2}}$$
$$= V\left[\tfrac{1}{2\pi}\{(\beta - \alpha) - \tfrac{1}{2}(\sin 2\beta - \sin 2\alpha)\}\right]^{\frac{1}{2}} \qquad (12.46)$$

The rms current involves integration of equation (12.43), squared, giving

$$I_{rms} = \frac{V}{Z}\left[\frac{1}{2\pi}\left((\beta - \alpha) - \frac{\sin(\beta - \alpha)\cos(\alpha + \phi + \beta)}{\cos\phi}\right)\right]^{\frac{1}{2}} \qquad (12.47)$$

Iterative solutions to equation (12.44) are shown in figure 12.5a, where it is seen that two straight-line relationships exist that relate *α* and *β-α*. Exact solutions to equation (12.44) exist for these two cases. That is, exact tractable solutions exist for the purely resistive load, *Φ* = 0, and the purely inductive load, *Φ* = ½π.

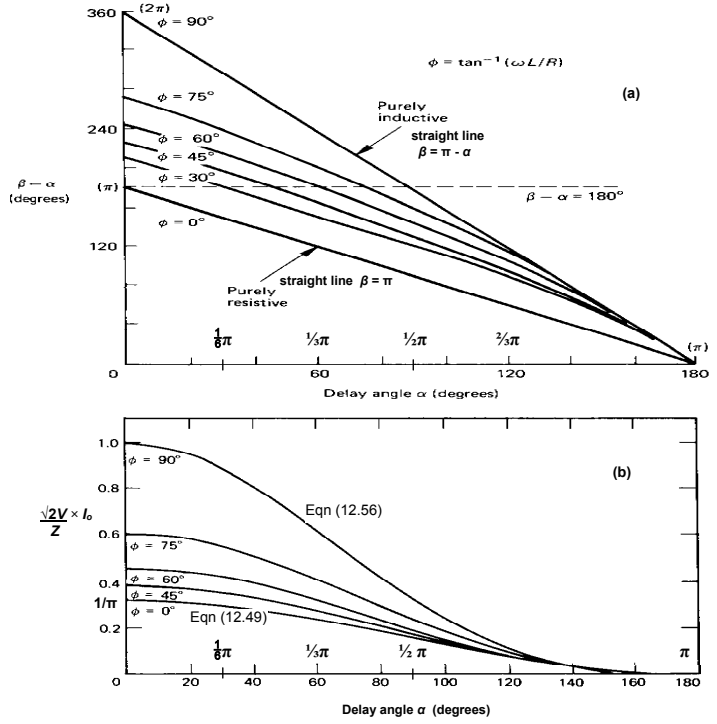


Figure 12.5. *Half-wave, controlled converter thyristor trigger delay angle α versus: (a) thyristor conduction angle, β-α, and (b) normalised mean load current.*

12.2.1i - Case 1: Purely resistive load. From equation (12.43), $Z = R$, $\phi = 0$, and the current is given by

$$i(\omega t) = \frac{\sqrt{2}\,V}{R}\sin(\omega t) \quad \text{(A)} \qquad \alpha \le \omega t \le \pi \text{ and } \beta = \pi \;\forall \alpha \tag{12.48}$$

The average load voltage, hence average load current, are

$$V_o = \frac{1}{2\pi}\int_{\alpha}^{\pi}\sqrt{2}V\sin\omega t\, d\omega t$$
$$V_o = \bar{I}_o R = \frac{\sqrt{2}V}{2\pi}(1+\cos\alpha) \quad \text{(V)} \tag{12.49}$$

where the maximum output voltage is 0.45V for zero delay angle.

The rms output voltage is

$$V_{rms} = \left[\tfrac{1}{2\pi}\int_{\alpha}^{\pi}\left(\sqrt{2}V\right)^2\sin^2\omega t\, d\omega t\right]^{\frac{1}{2}}$$
$$= V\left[\tfrac{1}{2\pi}\left\{(\pi-\alpha)+\tfrac{1}{2}\sin 2\alpha)\right\}\right]^{\frac{1}{2}} \tag{12.50}$$

Since the load is purely resistive, $I_{rms} = V_{rms}/R$ and the voltage and current factors (form and ripple) are equal. The power delivered to the load is $P_o = I_{rms}^2 R$.

The output voltage form factor is

$$FF_{vo} = \frac{V_{rms}}{V_o} = \frac{\sqrt{(\pi-\alpha)+\tfrac{1}{2}\sin 2\alpha}}{(1+\cos\alpha)} \tag{12.51}$$

The supply power factor, for a resistive load, is $P_{out}/V_{rms}I_{rms}$, that is

$$pf = \sqrt{\tfrac{1}{2} - \frac{\alpha}{2\pi} + \frac{\sin 2\alpha}{4\pi}} \tag{12.52}$$

12.2.1ii - Case 2: Purely inductive load. Circuit waveforms showing equal inductor voltage areas are shown in figure 12.4c. From equation (12.43), $Z = \omega L$, $\phi = \tfrac{1}{2}\pi$, and the output voltage and current ares given by

$$v_o(\omega t) = \begin{cases}\sqrt{2}V\sin\omega t & \alpha \le \omega t \le 2\pi - \alpha \\ 0 & \text{elsewhere}\end{cases} \tag{12.53}$$

$$i(\omega t) = \frac{\sqrt{2}\,V}{\omega L}\left(\sin(\omega t - \tfrac{1}{2}\pi) - \sin(\alpha - \tfrac{1}{2}\pi)\right) \quad \text{(A)}$$
$$= \frac{\sqrt{2}\,V}{\omega L}\left(\cos\alpha - \cos\omega t\right) \qquad \alpha \le \omega t \le \beta \text{ and } \beta = 2\pi - \alpha \tag{12.54}$$

The average load voltage, based on the equal area criterion, is zero

$$V_o = \frac{1}{2\pi}\int_{\alpha}^{2\pi-\alpha}\sqrt{2}V\sin\omega t\, d\omega t = 0 \tag{12.55}$$

The average output current is

$$\bar{I}_o = \tfrac{1}{2\pi}\int_{\alpha}^{2\pi-\alpha}\frac{\sqrt{2}\,V}{\omega L}\{\cos\alpha - \cos\omega t\}\, d\omega t$$
$$= \frac{\sqrt{2}\,V}{\pi\omega L}\left[(\pi-\alpha)\cos\alpha + \sin\alpha\right] \tag{12.56}$$

The rms output current is derived from

$$I_{rms} = \frac{\sqrt{2}\,V}{\omega L}\left[\tfrac{1}{2\pi}\int_{\alpha}^{2\pi-\alpha}(\cos\alpha - \cos\omega t)^2 d\omega t\right]^{\frac{1}{2}}$$
$$= \frac{V}{X}\left[\frac{1}{\pi}\left((\pi-\alpha)(2+\cos 2\alpha) + \frac{3}{2}\sin 2\alpha\right)\right]^{\frac{1}{2}} \tag{12.57}$$

The rms output voltage is

$$V_{rms} = \left[\tfrac{1}{2\pi}\int_{\alpha}^{2\pi-\alpha}\left(\sqrt{2}V\right)^2\sin^2\omega t\, d\omega t\right]^{\frac{1}{2}}$$
$$= V\left[\tfrac{1}{\pi}\left\{(\pi-\alpha)+\tfrac{1}{2}\sin 2\alpha)\right\}\right]^{\frac{1}{2}} \tag{12.58}$$

Since the load is purely inductive, $P_o = 0$ and the load voltage ripple factor is undefined since $V_o = 0$.

By setting $\alpha = 0$, the equations (12.48) to (12.58) are valid for the uncontrolled rectifier considered in section 11.1.3, for a purely resistive and purely inductive load, respectively.

12.2.1iii - Case 3: Back emf E and R-L load. With a load back emf, current begins to flow when the supply instantaneous voltage exceeds the back emf magnitude E, that is when

$$\check{\alpha} = \sin^{-1}\frac{E}{\sqrt{2}V} \tag{12.59}$$

When current flows, Kirchhoff's voltage law gives

$$v_o(t) = \sqrt{2}V\sin\omega t = Ri + L\frac{di}{dt} + E \tag{12.60}$$

Assuming **continuous current conduction**, using $R = Z\cos\Phi$ and $E = \sqrt{2}V_s\sin\check{\alpha}$, which yields

$$i(\omega t) = \frac{\sqrt{2}V}{Z}\sin(\omega t - \phi) - \frac{E}{R} - \left(\frac{\sqrt{2}V}{Z}\sin(\alpha - \phi) - \frac{E}{R}\right)e^{-\omega t + \alpha)/\tan\phi}$$
$$= \frac{\sqrt{2}V}{Z}\left[\sin(\omega t - \phi) - \frac{\sin\check{\alpha}}{\cos\phi} - \left(\sin(\alpha - \phi) - \frac{\sin\check{\alpha}}{\cos\phi}\right)e^{-\omega t + \alpha)/\tan\phi}\right] \tag{12.61}$$
$$\alpha \ge \check{\alpha} \text{ and } i(\omega t = 2\pi + \alpha) > 0$$

The load power is given by

$$P_L = I_{rms}^2 R + \bar{I}_o R \tag{12.62}$$

while the supply power factor is given by

$$pf = \frac{P_L}{V\,I_{rms}} = \frac{I_{rms}^2 R + \bar{I}_o R}{V\,I_{rms}} \tag{12.63}$$

The solution for the uncontrolled converter (a half-wave rectifier) is found by setting $\alpha = \check{\alpha}$, eqn (12.59).

The boundary for continuous current conduction is when $i > 0$ at the end of the conduction period when $\omega t = 2\pi + \alpha$, that is

$$\sin(\alpha - \phi) - \frac{\sin\check{\alpha}}{\cos\phi} - \left(\sin(\alpha - \phi) - \frac{\sin\check{\alpha}}{\cos\phi}\right)e^{-2\pi/\tan\phi} \ge 0$$

That is, continuous conduction occurs when

$$\sin(\alpha - \phi) \ge \frac{\sin\check{\alpha}}{\cos\phi} \tag{12.64}$$

With **discontinuous conduction,** the output current is still given by equation (12.61), until the current falls to zero at the extinction angle β. The extinction angle β is found from the boundary condition $i(\omega t) = i(\beta) = 0$, for $2\pi + \alpha > \beta > \tfrac{1}{2}\pi$, in equation(12.61). That is, β is found iteratively from:

$$\sin(\beta - \phi) - \frac{\sin\check{\alpha}}{\cos\phi}\left(1 - e^{\alpha - \beta)/\tan\phi}\right) - e^{\alpha - \beta)/\tan\phi}\sin(\alpha - \phi) = 0 \tag{12.65}$$

In the interval between $\beta \le \omega t \le 2\pi + \alpha$ no current flows and the output voltage is the load back emf, $E = \sqrt{2}V_s\sin\check{\alpha}$. The average output voltage, hence current are given by

$$V_o = \int_{\alpha}^{\beta}\sqrt{2}V\sin\omega t\, d\omega t + \int_{\beta}^{2\pi+\alpha} E\, d\omega t$$
$$= \int_{\alpha}^{\beta}\sqrt{2}V\sin\omega t\, d\omega t + \int_{\beta}^{2\pi+\alpha}\sqrt{2}V\sin\check{\alpha}\, d\omega t$$
$$V_o = \frac{\sqrt{2}V}{2\pi}\left[\cos\alpha - \cos\beta + (2\pi + \alpha - \beta)\times\sin\check{\alpha}\right] \tag{12.66}$$

Therefore

$$\bar{I}_o = \frac{V_o - E}{R} = \frac{V_o - \sqrt{2}V\sin\check{\alpha}}{R}$$
$$\bar{I}_o = \frac{\sqrt{2}V}{R}\frac{1}{2\pi}\left[\cos\alpha - \cos\beta + (\pi + \alpha - \beta)\times\sin\check{\alpha}\right] \tag{12.67}$$

Example 12.1: ***Single-phase, half-wave controlled rectifier***

The ac supply of the half-wave controlled single-phase converter in figure 12.4a is $v = \sqrt{2}\,240\sin\omega t$. For the following loads

Load #1: $R = 10\Omega$, $\omega L = 0\ \Omega$
Load #2: $R = 0\ \Omega$, $\omega L = 10\Omega$
Load #3: $R = 7.1\Omega$, $\omega L = 7.1\Omega$

Determine in each load case, for a firing delay angle $\alpha = \pi/6$

- the conduction angle $\gamma=\beta-\alpha$, hence the current extinction angle β
- the dc output voltage and the average output current
- the rms load current and voltage, load current and voltage ripple factor, and power dissipated in the load
- the supply power factor

Solution

Load #1: $Z = R = 10\Omega$, $\omega L = 0\ \Omega$

From equation (12.43), $Z = 10\Omega$ and $\phi = 0°$.
From equation (12.48), $\beta = \pi$ for all α, thus for $\alpha = \pi/6$, $\gamma = \beta - \alpha = 5\pi/6$.
From equation (12.49)

$$V_o = \bar{I}_o R = \frac{\sqrt{2}V}{2\pi}(1+\cos\alpha)$$

$$= \frac{\sqrt{2}V}{2\pi}(1+\cos\pi/6) = 100.9\text{V}$$

The average load current is

$$\bar{I}_o = V_o / R = \frac{\sqrt{2}V}{2\pi R}(1+\cos\alpha) = 100.9\text{V}/10\Omega = 10.1\text{A}.$$

The rms load voltage is given by equation (12.50), that is

$$V_{rms} = V\left[\tfrac{1}{2\pi}\left\{(\pi-\alpha)+\tfrac{1}{2}\sin 2\alpha\right\}\right]^{\frac{1}{2}}$$

$$= 240\text{V}\times\left[\tfrac{1}{2\pi}\left\{(\pi-\pi/6)+\tfrac{1}{2}\sin\pi/3\right\}\right]^{\frac{1}{2}} = 167.2\text{V}$$

Since the load is purely resistive, the power delivered to the load is

$$P_o = I_{rms}^2 R = V_{rms}^2 / R = 167.2\text{V}^2/10\Omega = 2797.0\text{W}$$

$$I_{rms} = V_{rms} / R = 167.9\text{V}/10\Omega = 16.8\text{A}$$

For a purely resistive load, the voltage and current factors are equal:

$$FF_i = FF_v = \frac{167.2\text{V}}{100.9\text{V}} = \frac{16.8\text{A}}{10.1\text{A}} = 1.68$$

$$RF_i = RF_v = \sqrt{FF^2 - 1} = 1.32$$

The power factor is

$$pf = \frac{2797\text{W}}{240\text{V}\times 16.7\text{A}} = 0.70$$

Alternatively, use of equation (12.52) gives

$$pf = \sqrt{\tfrac{1}{2} - \frac{\pi/6}{2\pi} + \frac{\sin\pi/6}{4\pi}} = 0.70$$

Load #2: $R = 0\ \Omega$, $Z = X = \omega L = 10\Omega$

From equation (12.43), $Z = X = 10\Omega$ and $\phi = \tfrac{1}{2}\pi$.
From equation (12.54), which is based on the equal area criterion, $\beta = 2\pi - \alpha$, thus for $\alpha = \pi/6$, $\beta = 11\pi/6$ whence the conduction period is $\gamma = \beta - \alpha = 5\pi/3$.
From equation (12.55) the average output voltage is

$$V_o = 0\text{V}$$

The average load current is

$$\bar{I}_o = \frac{\sqrt{2}\,V}{\pi\,\omega L}\left[(\pi-\alpha)\cos\alpha + \sin\alpha\right]$$

$$= \frac{\sqrt{2}\ 240}{\pi\times 10}\times\left[(5\pi/6)\cos\pi/6 + \sin\pi/6\right] = 14.9\text{A}$$

Using equations (12.57) and (12.58), the load rms voltage and current are

$$V_{rms} = 240\text{V}\left[\frac{1}{\pi}\left\{\pi - \pi/6 + \tfrac{1}{2}\sin\pi/3\right\}\right]^{\frac{1}{2}} = 236.5\text{V}$$

$$I_{rms} = \frac{240\text{V}}{10\Omega}\left[\frac{1}{\pi}\left\{(\pi-\pi/6)(2+\cos 2\alpha) + \tfrac{3}{2}\sin\pi/3\right\}\right]^{\frac{1}{2}} = 37.9\text{A}$$

Since the load is purely inductive, the power delivered to the load is zero, as is the power factor, and the output voltage ripple factor is undefined. The output current ripple factor is

$$FF_i = \frac{I_{rms}}{\bar{I}_o} = \frac{37.9\text{A}}{14.9\text{A}} = 2.54 \quad \text{whence} \quad RF_i = \sqrt{2.54^2 - 1} = 2.34$$

Load #3: $R = 7.1\Omega$, $\omega L = 7.1\Omega$

From equation (12.43), $Z = 10\Omega$ and $\phi = \tfrac{1}{4}\pi$.
From figure 12.5a, for $\phi = \tfrac{1}{4}\pi$ and $\alpha = \pi/6$, $\gamma = \beta - \alpha = 195°$ whence $\beta = 225°$.
Iteration of equation (12.44) gives $\beta = 225.5° = 3.936$ rad
From equation (12.45)

$$V_o = \bar{I}_o R = \frac{\sqrt{2}\,V}{2\pi}(\cos\alpha - \cos\beta)$$

$$= \frac{\sqrt{2}\times 240}{2\pi}(\cos 30° - \cos 225°) = 85.0\text{V}$$

The average load current is

$$\bar{I}_o = V_o / R$$

$$= 85.0\text{V}/7.1\Omega = 12.0\text{A}$$

Alternatively, the average current can be extracted from figure 12.5b, which for $\phi = \tfrac{1}{4}\pi$ and $\alpha = \pi/6$ gives the normalised current as 0.35, thus

$$\bar{I}_o = \sqrt{2}V/Z \times 0.35$$

$$= \sqrt{2}\times 240\text{V}/10\Omega \times 0.35 = 11.9\text{A}$$

From equation (12.47), the rms current is

$$I_{rms} = \frac{V}{Z}\left[\frac{1}{2\pi}\left((\beta-\alpha) - \frac{\sin(\beta-\alpha)\cos(\alpha+\phi+\beta)}{\cos\phi}\right)\right]^{\frac{1}{2}}$$

$$= \frac{240\text{V}}{10\Omega}\times\left[\frac{1}{2\pi}\left((3.93-\pi/6) - \frac{\sin(3.93-\pi/6)\cos(\pi/6+\tfrac{1}{4}\pi+3.93)}{\cos\tfrac{1}{4}\pi}\right)\right]^{\frac{1}{2}} = 18.18\text{A}$$

The power delivered to the load resistor is

$$P_o = I_{rms}^2 R = 18.18\text{A}^2\times 7.1\Omega = 2346\text{W}$$

The load rms voltage, from equation (12.46), is

$$V_{rms} = V\left[\tfrac{1}{2\pi}\left\{(\beta-\alpha) - \tfrac{1}{2}(\sin 2\beta - \sin 2\alpha)\right\}\right]^{\frac{1}{2}}$$

$$= 240\text{V}\left[\tfrac{1}{2\pi}\left\{(3.94-\tfrac{1}{6}\pi) - \tfrac{1}{2}\times(\sin(2\times 3.94) - \sin(2\times\tfrac{1}{6}\pi))\right\}\right]^{\frac{1}{2}} = 175.1\text{V}$$

The load current and voltage ripple factors are

$$FF_i = \frac{18.18\text{A}}{12.0\text{A}} = 1.515 \qquad RF_i = \sqrt{FF_i^2 - 1} = 1.138$$

$$FF_v = \frac{175.1\text{V}}{85\text{V}} = 2.06 \qquad RF_v = \sqrt{FF_v^2 - 1} = 1.8$$

The supply power factor is

$$pf = \frac{2346\text{W}}{240\text{V}\times 18.18\text{A}} = 0.54$$

♣

12.2.2 Single-phase, half-wave half-controlled

The half-wave controlled converter waveform in figure 12.4b shows that when $\alpha < \omega t < \pi$, during the positive half of the supply cycle, energy is delivered to the load. But when $\pi < \omega t < 2\pi$, the supply reverses and some energy in the load circuit is returned to the supply. More energy can be retained by the load if the load voltage is prevented from reversing. A load freewheel diode facilitates this objective.

The single-phase half-wave converter can be controlled when a load commutating diode is incorporated as shown in figure 12.6a. The diode will prevent the instantaneous load voltage v_o from going negative, as with the single-phase half-controlled converters shown in figure 12.1.
The load current is defined by equation 11.31 for $\alpha \le \omega t \le \pi$ and equation 11.32 for $\pi \le \omega t \le 2\pi + \alpha$, namely:

$$L\frac{di}{dt} + Ri = \sqrt{2}V\sin\omega t \qquad \text{(A)} \qquad \alpha \le \omega t \le \pi$$
$$L\frac{di}{dt} + Ri = 0 \qquad \text{(A)} \qquad \pi \le \omega t \le 2\pi + \alpha \tag{12.68}$$

At $\omega t = \pi$ the thyristor is line commutated and the load current, and hence freewheel diode current, is of the form of equation 11.33. As shown in figure 12.6b, depending on the delay angle α and R-L load time constant (L/R), the load current may fall to zero, producing discontinuous load current.

The mean load voltage (hence mean output current) for all conduction cases, with a passive L-R load, is

$$V_o = \frac{1}{2\pi}\int_{\alpha}^{\pi}\sqrt{2}V\sin\omega t \;\; d\omega t$$
$$V_o = \bar{I}_o R = \frac{\sqrt{2}V}{2\pi}(1+\cos\alpha) \qquad \text{(V)} \tag{12.69}$$

which is half the mean voltage for a single-phase half-controlled converter, given by equation (12.1).

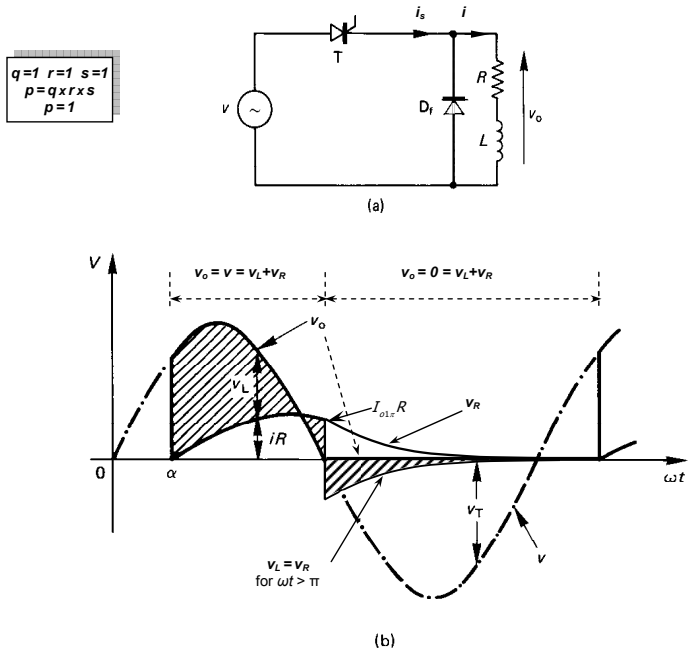


Figure 12.6. *Half-wave half-controlled converter:*
(a) circuit diagram and (b) circuit waveforms for an inductive load.

The maximum mean output voltage, $\hat{V}_o = \sqrt{2}V/\pi$ (equation 11.27), occurs at $\alpha = 0$. The normalised mean output voltage V_n is

$$V_n = V_o/\hat{V}_o = \tfrac{1}{2}(1+\cos\alpha) \tag{12.70}$$

The Fourier coefficients of the *1*-pulse output voltage are given by equation (12.128). For the single-phase, half-wave, half-controlled case, $p = 1$, thus the output voltage harmonics occur at $n = 1, 2, 3, \ldots$

The rms output voltage for both continuous and discontinuous load current is

$$V_{rms} = \left[\tfrac{1}{2\pi}\int_{\alpha}^{\pi}\left(\sqrt{2}V\right)^2\sin^2\omega t \; d\omega t\right]^{\frac{1}{2}}$$
$$= V\left[\tfrac{1}{2\pi}\left\{(\pi-\alpha)+\tfrac{1}{2}\sin 2\alpha)\right\}\right]^{\frac{1}{2}} \tag{12.71}$$

12.2.2i - For **discontinuous conduction** the load current is defined by equation (12.43) during thyristor conduction

$$i(\omega t) = i_s(\omega t) = \frac{\sqrt{2}V}{Z}\times\left(\sin(\omega t-\phi) - \sin(\alpha-\phi)\, e^{-\omega t+\alpha/\tan\phi}\right) \qquad \text{(A)}$$
$$\alpha \le \omega t \le \pi$$
$$i(\omega t) = i_{Df}(\omega t) = I_{01\pi}e^{-\omega t/\tan\phi} \tag{12.72}$$
$$= \left\{\frac{\sqrt{2}V}{Z}\times\sin(\phi-\alpha)(1-e^{-\pi/\tan\phi})\right\}e^{-\omega t+\pi/\tan\phi} \qquad \text{(A)}$$
$$\pi \le \omega t \le 2\pi + \alpha$$

The average thyristor current is

$$\bar{I}_T = \frac{V}{\sqrt{2}\pi R}\times\left(\cos^2\phi + \cos\alpha + \sin\phi\times\sin(\alpha-\phi)\times e^{\alpha-\pi/\tan\phi}\right) \tag{12.73}$$

while the average freewheel diode current is

$$\bar{I}_{Df} = \bar{I}_o - \bar{I}_T = \frac{V\sin\phi}{\sqrt{2}\pi R}\times\left(\sin\phi - \times\sin(\alpha-\phi)\times e^{\alpha-\pi/\tan\phi}\right) \tag{12.74}$$

12.2.2ii - For **continuous conduction** the load current is defined by

$$i(\omega t) = i_s(\omega t) = \frac{\sqrt{2}V}{Z}\times\left(\sin(\omega t-\phi) + \left(\frac{\sin\phi\; e^{-\alpha/\tan\phi} - \sin(\alpha-\phi)}{1-e^{-2\pi/\tan\phi}}\right)e^{-\omega t+\alpha/\tan\phi}\right)$$
$$\alpha \le \omega t \le \pi \qquad \text{(A)}$$
$$i(\omega t) = i_{Df}(\omega t) = I_{01\pi}e^{-\omega t/\tan\phi} \tag{12.75}$$
$$= \left\{\frac{\sqrt{2}V}{Z}\times\frac{\sin\phi - \sin(\alpha-\phi)\,e^{-\pi-\alpha/\tan\phi}}{1-e^{-\pi/\tan\phi}}\right\}e^{-\omega t+\pi/\tan\phi} \qquad \text{(A)}$$
$$\pi \le \omega t \le 2\pi + \alpha$$

The advantages of incorporating a load freewheel diode are

- the input power factor is improved and
- the load waveform is improved (less ripple) giving a better load performance

12.2.3 Single-phase, full-wave controlled rectifier circuit with an R-L load

Full-wave voltage control is possible with the circuits shown in figures 12.7a and b. The circuit in figure 12.7a uses a centre-tapped transformer and two thyristors which experience a reverse bias of twice the supply. At high powers where a transformer may not be applicable, a four-thyristor configuration as in figure 12.7b is suitable. The voltage ratings of the thyristors in figure 12.7b are half those of the devices in figure 12.7a, for a given converter input voltage.
Load voltage and current waveforms are shown in figure 12.7 parts c, d, and e for three different phase control angle conditions.
The load current waveform becomes continuous when the phase control angle α is given by

$$\alpha = \tan^{-1}\omega L/R = \phi \qquad \text{(rad)} \tag{12.76}$$

at which angle the output current is a rectified sine wave. For $\alpha > \phi$, discontinuous load current flows as shown in figure 12.7c. At $\alpha = \phi$ the load current becomes continuous as shown in figure 12.7d, whence $\beta = \alpha + \pi$. Further decrease in α, that is $\alpha < \phi$, results in continuous load current that is always greater than zero (no zero current periods), as shown in figure 12.7e.

12.2.3i - $\alpha > \phi$, $\beta - \alpha < \pi$ *, discontinuous load current*

The load current waveform is the same as for the half-wave situation considered in section 12.2.1, given by equation (12.43). That is

$$i(\omega t) = \frac{\sqrt{2}V}{Z}\left[\sin(\omega t - \phi) - \sin(\alpha - \phi)\, e^{\{(\alpha - \omega t)/\tan\phi\}}\right] \quad \text{(A)}$$
$$\alpha \le \omega t < \beta \quad \text{(rad)} \qquad (12.77)$$

The mean output voltage for this full-wave circuit will be twice that of the half-wave case in section 12.2.1, given by equation (12.45). That is

$$V_o = \bar{I}_o R = \frac{1}{\pi}\int_\alpha^\beta \sqrt{2}V \sin\omega t\, d\omega t$$
$$= \frac{\sqrt{2}V}{\pi}(\cos\alpha - \cos\beta) \quad \text{(V)} \qquad (12.78)$$

where β can be extracted from figure 12.5. For a purely resistive load, $\beta = \pi$. The average output current is given by $\bar{I}_o = V_o / R$ and the average and rms thyristor currents are $\frac{1}{2}\bar{I}_o$ and $I_{rms}/\sqrt{2}$, respectively.

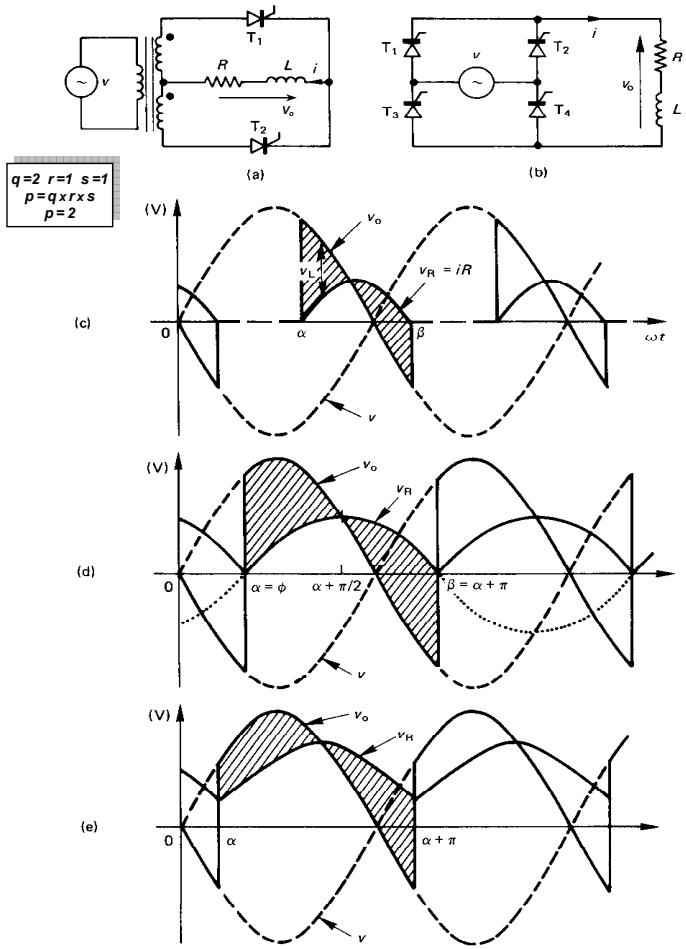


Figure 12.7. *Full-wave controlled converter:*
(a) and (b) circuit diagrams; (c) discontinuous load current; (d) verge of continuous load

current, when α = ø; and (e) continuous load current.

The rms load voltage is

$$V_{rms} = \left[\frac{1}{\pi}\int_\alpha^\beta 2V^2 \sin^2\omega t\, d\omega t\right]^{\frac{1}{2}}$$
$$= V\left[\frac{1}{\pi}\{(\beta - \alpha) - \tfrac{1}{2}(\sin 2\beta - \sin 2\alpha)\}\right]^{\frac{1}{2}} \qquad (12.79)$$

The rms load current is

$$I_{rms} = \frac{V}{Z}\left[\frac{1}{\pi}\left((\beta - \alpha) - \frac{\sin(\beta - \alpha)\cos(\alpha + \phi + \beta)}{\cos\phi}\right)\right]^{\frac{1}{2}} \qquad (12.80)$$

The load power is therefore $P = I_{rms}^2 R$.

12.2.3ii - $\alpha = \phi$, $\beta - \alpha = \pi$ *, verge of continuous load current*

When $\alpha = \phi = \tan^{-1}\omega L/R$, the load current given by equation (12.77) reduces to

$$i(\omega t) = \frac{\sqrt{2}V}{Z}\sin(\omega t - \phi) \quad \text{(A)}$$
$$\text{for } \phi \le \omega t \le \phi + \pi \quad \text{(rad)} \qquad (12.81)$$

and the mean output voltage, on reducing equation (12.78) using $\beta = \alpha + \pi$, is given by

$$V_o = \frac{2\sqrt{2}V}{\pi}\cos\alpha \quad \text{(V)} \qquad (12.82)$$

which is dependent on the load such that $\alpha = \phi = \tan^{-1}\omega L/R$. From equation (12.79), with $\beta - \alpha = \pi$, the rms output voltage is V, $I_{rms} = V/Z$, and power = $VI_{rms}\cos\phi$.

12.2.3iii - $\alpha < \phi$, $\beta - \pi = \alpha$*, continuous load current (and also a purely inductive load)*

Under a continuous load current conduction condition, a thyristor is still conducting when another is forward-biased and is turned on. The first device is instantaneously reverse-biased by the second device which has been turned on. The first device is commutated and load current is instantaneously transferred to the oncoming device.
The load current is given by

$$i(\omega t) = \frac{\sqrt{2}V}{Z}\left[\sin(\omega t - \phi) - \frac{2\sin(\alpha - \phi)}{1 - e^{-\pi/\tan\phi}}\, e^{\{(\alpha - \omega t)/\tan\phi\}}\right] \qquad (12.83)$$

This equation reduces to equation (12.81) for $\alpha = \phi$ and equation 11.52 for $\alpha = 0$.
The mean output voltage, whence mean output current, are defined by equation (12.82)

$$V_o = \bar{I}_o R = \frac{2\sqrt{2}V}{\pi}\cos\alpha \quad \text{(V)}$$

which is uniquely defined by α. The maximum mean output voltage, $\hat{V}_o = 2\sqrt{2}V/\pi$ (equation 11.54), occurs at $\alpha = 0$. Generally, for $\alpha > \frac{1}{2}\pi$, the average output voltage is negative, resulting in a net energy transfer from the load to the supply.

The normalised mean output voltage V_n is

$$V_n = V_o / \hat{V}_o = \cos\alpha \qquad (12.84)$$

The rms output voltage is equal to the rms input supply voltage and is given by

$$V_{rms} = \sqrt{\frac{1}{\pi}\int_\alpha^{\pi+\alpha}\left(\sqrt{2}V\right)^2 \sin^2\omega t\, d\omega t} = V \qquad (12.85)$$

The ac in the output voltage is

$$V_{ac} = \sqrt{V_{rms}^2 - V_o^2} = V\sqrt{1 + \tfrac{8}{\pi}\cos^2\alpha} \qquad (12.86)$$

The ac component harmonic magnitudes in the load are given by

$$V_n = \frac{\sqrt{2}V}{2\pi} \times \left(\frac{1}{(n-1)^2} + \frac{1}{(n+1)^2} - \frac{2\cos 2\alpha}{(n-1)(n+1)}\right) \qquad (12.87)$$

for n even, namely $n = 2, 4, 6\ldots$
The load voltage form factor, (thence ripple factor), is

$$FF_v = \frac{\pi}{2\sqrt{2}\cos\alpha} \qquad (12.88)$$

The current harmonics are obtained by division of the voltage harmonic by its load impedance at that frequency, that is

$$I_n = \frac{V_n}{Z_n} = \frac{V_n}{\sqrt{R^2 + (n\omega L)^2}} \qquad n = 0, 2, 4, 6, .. \tag{12.89}$$

Integration of equation (12.83), squared, yields the load rms current (or equation 11.53 for $\alpha = 0$)

$$I_{rms} = \frac{V}{Z}\left[\frac{1}{\pi}\left\{\pi + \left(\frac{2\sin(\alpha-\phi)}{1-e^{-\pi/\tan\phi}}\right)^2 \tan\phi\left(1-e^{-2\pi/\tan\phi}\right) - 4\left(\frac{2\sin(\alpha-\phi)}{1-e^{-\pi/\tan\phi}}\right)\sin\alpha\sin\phi\left(1-e^{-\pi/\tan\phi}\right)\right\}\right]^{½} \tag{12.90}$$

Thyristor average current is $½\bar{I}_o$, while thyristor rms current rating is $I_{rms}/\sqrt{2}$. The same thyristor current rating expressions are valid for both continuous and discontinuous load current conditions.

For a highly inductive load, constant load current, the supply power factor is $pf = \frac{2}{\pi}(1+\cos\alpha)\sqrt{\frac{\pi}{\pi-\alpha}}$.

The harmonic factor or voltage ripple factor for the output voltage is

$$RF_v = \frac{\sqrt{V_{rms}^2 - V_o^2}}{V_o} = \left[\frac{\pi^2}{8} - \cos^2\alpha\right]^{½} \tag{12.91}$$

which is a minimum of 0.483 for α=0 and a maximum of 1.11 when α=½π.

Critical load inductance (see figure 12.14)
The critical load inductance, to prevent the load current falling to zero, is given by

$$\frac{\omega L_{crit}}{R} = \frac{\pi}{2\cos\alpha}\left(\cos\theta + \frac{2}{\pi}\sin\alpha - \frac{2}{\pi}\cos\alpha\left(½\pi + \alpha + \theta\right)\right) \tag{12.92}$$

for $\alpha \le \theta$ where

$$\theta = \sin^{-1}\frac{V_o}{\sqrt{2}V} = \sin^{-1}\frac{2\cos\alpha}{\pi} \tag{12.93}$$

The minimum current occurs at the angle θ, where the mean output voltage V_o equals the instantaneous load voltage, v_o. When the phase delay angle α is greater than the critical angle θ, substituting $\alpha = \theta$ in equation (12.92) gives

$$\frac{\omega L_{crit}}{R} = -\tan\alpha \tag{12.94}$$

For a purely resistive load

$$V_o = \frac{\sqrt{2}V}{\pi}(1+\cos\alpha) \tag{12.95}$$

12.2.3iv *Resistive load, $\beta = \pi$*

When the load is purely resistive, that is $L = 0$, the average and rms output voltage and currents are given by substituting $Z = R$ and $\beta = \pi$ in to equations (12.78), (12.79) and (12.80). That is

$$V_o = \bar{I}_o R = \frac{\sqrt{2}V}{\pi}(\cos\alpha + 1) \qquad \text{(V)} \tag{12.96}$$

$$V_{ms} = I_{rms}R = V\left[\frac{1}{\pi}\{(\pi-\alpha) - ½\sin 2\alpha\}\right]^{½} \tag{12.97}$$

Example 12.2: ***Controlled full-wave converter – continuous and discontinuous conduction***

The fully controlled full-wave, single-phase converter in figure 12.7a has a source of 240V rms, 50Hz, and a 10Ω 50mH series load. If the delay angle is 45°, determine

i. the average output voltage and current, hence thyristor mean current
ii. the rms load voltage and current, hence thyristor rms current and load ripple factors
iii. the power absorbed by the load and the supply power factor

If the delay angle is increased to 75° determine

iv. the load current in the time domain
v. numerically solve the load current equation for β, the current extinction angle
vi. the load average current and voltage
vii. the load rms voltage and current hence load ripple factors and power dissipated
viii. the supply power factor

Solution

The load natural power factor angle is given by

$$\phi = \tan^{-1}\omega L/R = \tan^{-1}(2\pi 50 \times 50\text{mH}/10\Omega) = 57.5° = 1 \text{ rad}$$

Continuous conduction

Since $\alpha < \phi$ $(45° < 57.5°)$, continuous load current flows, which is given by equation (12.83).

$$i(\omega t) = \frac{\sqrt{2}\times 240\text{V}}{18.62\Omega}\left[\sin(\omega t - 1) - \frac{2\times\sin(1.31 - 1)}{1-e^{-\pi/1.56}}\, e^{-(1.31 - \omega t)/1.56}\right]$$
$$= 18.2\times\left[\sin(\omega t - 1) - 1.62\times e^{-\omega t/1.56}\right]$$

i. The average output current and voltage are given by equation (12.82)

$$V_o = \bar{I}_o R = \frac{2\sqrt{2}V}{\pi}\cos\alpha = \frac{2\sqrt{2}V}{\pi}\cos 45° = 152.8\text{V}$$
$$\bar{I}_o = V_o/R = 152.8\text{V}/10\Omega = 15.3\text{A}$$

Each thyristor conducts for 180°, hence thyristor mean current is ½ of 15.3A = 7.65A.

ii. The rms load current is determined by harmonic analysis. The voltage harmonics (peak magnitude) are given by equation (12.87)

$$V_n = \frac{\sqrt{2}\,V}{2\pi}\times\left(\frac{1}{(n-1)^2} + \frac{1}{(n+1)^2} - \frac{2\cos 2\alpha}{(n-1)(n+1)}\right) \qquad \text{for} \quad n = 2, 4, 6, ..$$

and the corresponding current is given from equation (12.89)

$$I_n = \frac{V_n}{Z_n} = \frac{V_n}{\sqrt{R^2 + (n\omega L)^2}}$$

harmonic *n*	V_n	$Z_n = \sqrt{R^2 + (n\omega L)^2}$	$I_n = \frac{V_n}{Z_n}$	$½I_n^2$
0	***(152.79)***	10.00	15.28	***(233.44)***
2	55.65	32.97	1.69	1.42
4	8.16	63.62	0.13	0.01
6	3.03	94.78	0.07	0.00
			$I_o^2 + \sum ½I_n^2 =$	234.4

The dc output voltage component is given by equation (12.82).
From the calculations in the table, the rms load current is

$$I_{rms} = \sqrt{I_o^2 + ½\sum I_n^2} = \sqrt{234.4} = 15.3\text{A}$$

Since each thyristor conducts for 180°, the thyristor rms current is $1/\sqrt{2}$ of 15.3A = 10.8A

The rms load voltage is given by equation (12.85), that is 240V.

$$FF_i = \frac{I_{rms}}{\bar{I}_o} = \frac{15.3\text{A}}{15.3\text{A}} = 1.0 \qquad RF_i = \sqrt{FF_i^2 - 1} = \sqrt{1.00^2 - 1} = 0.0$$
$$FF_v = \frac{V_{rms}}{\bar{V}_o} = \frac{240\text{V}}{152.8\text{V}} = 1.57 \qquad RF_v = \sqrt{FF_v^2 - 1} = \sqrt{1.57^2 - 1} = 1.21$$

iii. The power absorbed by the load is

$$P_L = I_{rms}^2 R = 15.3\text{A}^2 \times 10\Omega = 2344\text{W}$$

The supply power factor is

$$pf = \frac{P_L}{V_{rms} I_{rms}} = \frac{2344\text{W}}{240\text{V}\times 15.3\text{A}} = 0.64$$

Discontinuous conduction

iv. When the delay angle is increased to 75° (1.31 rad), discontinuous load current flows since the natural power factor angle $\phi = \tan^{-1}\omega L/R = \tan^{-1}(2\pi 50\times 50\text{mH}/10\Omega) = 57.5° \equiv 1\text{rad}$ is exceeded. The load current is given by equation (12.77)

$$i(\omega t) = \frac{\sqrt{2}\times 240\text{V}}{18.62\Omega}\ [\sin(\omega t\ -\ 1)\ -\ \sin(1.31\ -1)\ e^{\ -(1.31\ -\ \omega t)/1.56}]$$
$$= 18.2\times[\sin(\omega t\ -\ 1)\ -\ 0.71\times\ e^{\ -\omega t/1.56}]$$

v. Solving the equation in part iv for $\omega t = \beta$ and zero current, that is

$$0 = \sin(\beta\ -\ 1)\ -\ 0.71\times\ e^{\ -\beta/1.56}$$

gives β = 4.09 rad or 234.3°.

vi. The average load voltage from equation (12.78) is

$$V_o = \frac{\sqrt{2}\,240\text{V}}{\pi}\ (\cos\ 75°\ -\ \cos\ 234.5°) = 90.8\text{V}$$
$$\bar{I}_o = \frac{V_o}{R} = \frac{90.8V}{10\Omega} = 9.08\text{A}$$

vii. The rms load voltage is given by equation (12.79)

$$V_{rms} = 240\text{V}\times\left[\frac{1}{\pi}\{(4.09-1.31)-½(\sin 8.18-\sin 2.62)\}\right]^{½} = 216.46\text{V}$$

The rms current from equation (12.80) is

$$I_{rms} = \frac{240\text{V}}{18.62\Omega}\times\left[\frac{1}{\pi}\left((4.09-1.31)-\frac{\sin(4.09-1.31)\times\cos(1.31+1+4.09)}{\cos 1}\right)\right]^{½} = 13.55\text{A}$$

The load voltage and current form and ripple factors are

$$FF_i = \frac{I_{rms}}{\bar{I}_o} = \frac{13.55\text{A}}{9.08\text{A}} = 1.49 \qquad RF_i = \sqrt{FF_i^2-1} = \sqrt{1.49^2-1} = 1.11$$
$$FF_v = \frac{V_{rms}}{\bar{V}_o} = \frac{216.46\text{V}}{90.8\text{V}} = 2.38 \qquad RF_v = \sqrt{FF_v^2-1} = \sqrt{2.38^2-1} = 2.16$$

The power dissipated in the 10Ω load resistor is

$$P = I_{rms}^2 R = 13.55^2\times 10\Omega = 1836\text{W}$$

viii. The supply power factor is

$$pf = \frac{P_L}{V_{rms}I_{rms}} = \frac{1836\text{W}}{240\text{V}\times 13.55\text{A}} = 0.56$$

♣

12.2.4 Single-phase, full-wave, fully-controlled circuit with R-L and emf load, E

An emf source and *R-L* load can be encountered in dc machine modelling. The emf represents the machine speed back emf, defined by $E = k\phi\omega$. DC machines can be controlled by a fully controlled converter configuration as shown in figure 12.8a, where T_1-T_4 and T_2-T_3 are triggered alternately.
If in each half sine period the thyristor firing delay angle occurs after the rectified sine supply has fallen below the emf level *E*, then no load current flows since the bridge thyristors will always be reverse-biased. Thus the zero current firing angle $\hat{\alpha}$ is:

$$\hat{\alpha} = \sin^{-1}\left(E/\sqrt{2}V\right) \qquad \text{(rad)} \quad \text{for} \quad ½\pi < \hat{\alpha} < \pi \qquad (12.98)$$

where it has been assumed the emf has the polarity shown in figure 12.8a. With discontinuous output current, load current cannot flow until the supply voltage exceeds the back emf *E*. That is

$$\check{\alpha} = \sin^{-1}\left(E/\sqrt{2}V\right) \qquad \text{(rad)} \quad \text{for} \quad 0 < \check{\alpha} < ½\pi \qquad (12.99)$$

Load current can always flow with a firing angle defined by

$$\check{\alpha} \le \alpha \le \hat{\alpha} \qquad \text{(rad)} \qquad (12.100)$$

The load circuit current can be evaluated by solving

$$\sqrt{2}V\sin\omega t = L\frac{di}{dt} + Ri + E \qquad \text{(V)} \qquad (12.101)$$

The load voltage and current ripple are both at twice the supply frequency.

12.2.4i - Discontinuous load current

The load current is given by

$$i(\omega t) = \frac{\sqrt{2}V}{R}\ [\cos\phi\sin(\omega t\ -\ \phi) - \frac{E}{\sqrt{2}V} + \left\{\frac{E}{\sqrt{2}V} - \cos\phi\sin(\alpha\ -\phi)\right\}e^{\ -(\omega t\ -\ \alpha)/\tan\phi}] \qquad (12.102)$$
$$\check{\alpha} \le \omega t\ \le \beta < \pi + \alpha \qquad \text{(rad)}$$

For discontinuous load current conduction, the current extinction angle β, shown on figure 12.8b, is solved by iterative techniques for $i(\omega t{=}\beta) = 0$ in equation (12.102).

$$\cos\phi\sin(\beta\ -\ \phi) - \frac{E}{\sqrt{2}V} + \left\{\frac{E}{\sqrt{2}V} - \cos\phi\sin(\alpha\ -\phi)\right\}e^{\ -(\alpha\ -\ \beta)/\tan\phi} = 0 \qquad (12.103)$$

The mean output voltage can be obtained from equation (12.78), which is valid for $E = 0$.

In the interval between $\beta \le \omega t \le \pi+\alpha$ no current flows and the output voltage is the load back emf, $E = \sqrt{2}V_s\sin\check{\alpha}$. The average output voltage, hence current are given by

$$V_o = \int_{\alpha}^{\beta}\sqrt{2}V\sin\omega t\,d\omega t + \int_{\beta}^{\pi+\alpha}E\,d\omega t$$
$$= \int_{\alpha}^{\beta}\sqrt{2}V\sin\omega t\,d\omega t + \int_{\beta}^{\pi+\alpha}\sqrt{2}V\sin\check{\alpha}\,d\omega t$$
$$V_o = \frac{\sqrt{2}V}{\pi}\left[\cos\alpha - \cos\beta + (\pi+\alpha-\beta)\times\sin\check{\alpha}\right]$$
$$V_o = \frac{\sqrt{2}V}{\pi}\left(\cos\ \alpha - \cos\ \beta + (\pi+\alpha-\beta)\frac{E}{\sqrt{2}V}\right) \qquad \text{(V)} \qquad (12.104)$$
$$0 < \beta - \alpha < \pi \qquad \text{(rad)}$$

The current extinction angle β is load-dependent, being a function of *Z* and *E*, as well as α.

Since $V_o = E + \bar{I}_o R$, the mean load current is given by

$$\bar{I}_o = \frac{V_o - E}{R} = \frac{V_o - \sqrt{2}V\sin\check{\alpha}}{R}$$
$$\bar{I}_o = \frac{V_o - E}{R} = \frac{\sqrt{2}V}{\pi R}\left(\cos\ \alpha - \cos\ \beta - \frac{E}{\sqrt{2}V}(\beta-\alpha)\right) \qquad \text{(A)}$$
$$\bar{I}_o = \frac{\sqrt{2}V}{R}\frac{1}{\pi}\left[\cos\alpha - \cos\beta + (\alpha-\beta)\times\sin\check{\alpha}\right] \qquad (12.105)$$
$$0 < \beta - \alpha < \pi \qquad \text{(rad)}$$

The rms output voltage is given by

$$V_{rms} = \left(V^2\frac{\beta-\alpha}{\pi} + E^2(1-\frac{\beta-\alpha}{\pi}) - \frac{V^2}{2\pi}(\sin 2\beta - \sin 2\alpha)\right)^{½} \qquad \text{(V)} \qquad (12.106)$$

The rms voltage across the *R-L* part of the load is given by

$$V_{RLrms} = \sqrt{V_{rms}^2 - E^2} \qquad (12.107)$$

The total power delivered to the *R-L-E* load is

$$P_o = I_{rms}^2 R + \bar{I}_o E \qquad (12.108)$$

where the rms load current is found by integrating the current in equation (12.102), squared, etc.

The boundary for continuous current conduction is when $i > 0$ at the end of the conduction period when $\omega t = \pi+\alpha$, that is

$$-\sin(\alpha-\phi) - \frac{\sin\check{\alpha}}{\cos\phi} - \left(\sin(\alpha-\phi) - \frac{\sin\check{\alpha}}{\cos\phi}\right)e^{-\pi/\tan\phi} \ge 0$$

That is, continuous conduction occurs when

$$\sin(\alpha-\phi) \ge \frac{\sin\check{\alpha}}{\cos\phi}\times\tanh\left(\frac{-½\pi}{\tan\phi}\right) \qquad (12.109)$$

With **discontinuous conduction,** the output current is still given by equation (12.61), until the current falls to zero at the extinction angle β. The extinction angle β is found from the boundary condition $i(\omega t) =$

$i(\beta) = 0$, for $\pi+\alpha > \beta > \pi$, in equation(12.61). That is, β is found iteratively from:

$$\sin(\beta-\phi) - \frac{\sin\hat{\alpha}}{\cos\phi}\left(1 - e^{(\alpha-\beta)/\tan\phi}\right) - e^{(\alpha-\beta)/\tan\phi}\sin(\alpha-\phi) = 0 \qquad (12.110)$$

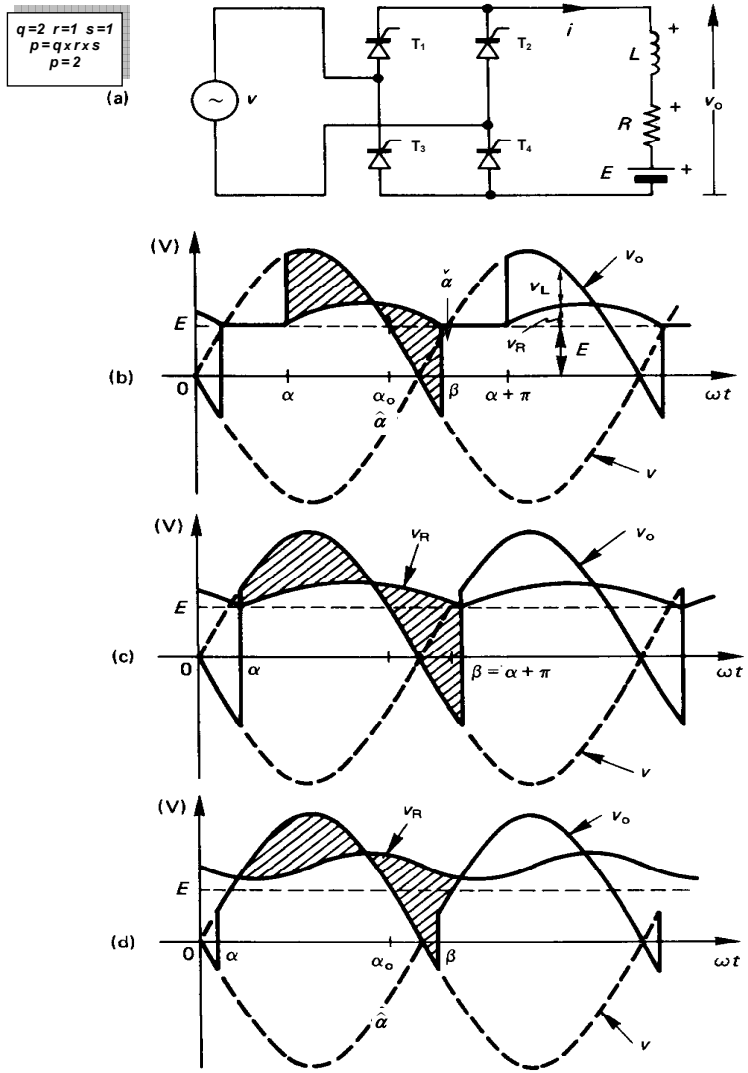


Figure 12.8. *A full-wave fully controlled converter with an inductive load which includes an emf source: (a) circuit diagram; (b) voltage waveforms with discontinuous load current; (c) verge of continuous load current; and (d) continuous load current.*

12.2.4ii - Continuous load current

With continuous load current conduction, the load rms voltage is *V*.

The load current is given by

$$i(\omega t) = \frac{\sqrt{2}V}{Z}\left[\sin(\omega t - \phi) - \frac{\sqrt{2}V/Z}{\cos\phi} + 2\frac{\sin(\alpha-\phi)}{e^{-\pi/\tan\phi}-1}e^{\{(\alpha-\omega t)/\tan\phi\}}\right] \qquad (12.111)$$

$$\alpha \le \omega t \le \pi + \alpha \qquad \text{(rad)}$$

The periodic minimum current is given by

$$\check{I} = \frac{\sqrt{2}V}{Z}\sin(\alpha-\phi)\frac{e^{-\pi/\tan\phi}+1}{e^{-\pi/\tan\phi}-1} - \frac{E}{R} = \frac{\sqrt{2}V}{Z}\sin(\alpha-\phi)\tanh\left(\frac{\frac{1}{2}\pi}{\tan\phi}\right) - \frac{E}{R} \qquad (12.112)$$

For continuous load current conditions, as shown in figures 12.8c and 12.8d, the mean output voltage is given by equation (12.104) with $\beta = \pi - \alpha$

$$V_o = \frac{1}{\pi}\int_{\alpha}^{\pi+\alpha}\sqrt{2}V\sin\omega t\, d\omega t \qquad \left(= E + \bar{I}_o R\right)$$

$$= \frac{2\sqrt{2}V}{\pi}\cos\alpha \qquad \text{(V)} \qquad (12.113)$$

$$\left(= \sqrt{2}V\frac{p}{\pi}\sin\frac{\pi}{p}\cos\alpha\right)$$

The average output voltage is dependent only on the phase delay angle α (independent of *E*), unlike the mean load current, which is given by

$$\bar{I}_o = \frac{V_o - E}{R} = \frac{\sqrt{2}V}{R}\left(\frac{2}{\pi}\cos\alpha - \frac{E}{\sqrt{2}V}\right) \qquad \text{(A)} \qquad (12.114)$$

The power absorbed by the emf source in the load is $P = \bar{I}_o E$, while the total power delivered to the *R-L-E* load is $P_o = I_{rms}^2 R + \bar{I}_o E$.

The output voltage can be expressed as a Fourier series.

$$v_o(t) = V_o + \sqrt{2}V\sum_{n=1}^{\infty}[a_n\cos 2n\omega t + b_n\sin 2n\omega t]$$

$$= \sqrt{2}V\left\{b_o + \sum_{n=1}^{\infty}[a_n\cos 2n\omega t + b_n\sin 2n\omega t]\right\}$$

$$V_o = \frac{1}{\pi}\int_0^{\pi}v_o(t)\,d\omega t = \frac{1}{\pi}\int_0^{\pi}\sqrt{2}V\sin\omega t\,d\omega t = \sqrt{2}V\frac{2}{\pi}\cos\alpha = \sqrt{2}V\times b_o \qquad (12.115)$$

$$a_n = \frac{2}{\pi}\left[\frac{\cos(2n+1)\alpha}{2n+1} - \frac{\cos(2n-1)\alpha}{2n-1}\right]$$

$$b_n = \frac{2}{\pi}\left[\frac{\sin(2n+1)\alpha}{2n+1} - \frac{\sin(2n-1)\alpha}{2n-1}\right] \qquad (12.116)$$

Each harmonic current can be found by dividing each harmonic current by its associated impedance, that is $i_n = V_n / Z_n$ where $Z_n = \sqrt{R^2 + (2n\omega L)^2}$.

The output current and voltage ripple is at multiples of twice the supply frequency. The output voltage harmonic magnitudes for continuous conduction, given by equation (12.87), are

$$V_n = \frac{\sqrt{2}V}{2\pi}\times\left(\frac{1}{(n-1)^2} + \frac{1}{(n+1)^2} - \frac{2\cos 2\alpha}{(n-1)(n+1)}\right) \quad \text{for} \quad n = 2,4,6,.. \qquad (12.117)$$

The dc component across the *R-L* (and just the resistor) part of the load is

$$V_{oR-L} = V_o - E$$

$$= \frac{2\sqrt{2}V}{\pi}\times\cos\alpha - E \qquad (12.118)$$

The ac component of the output voltage is

$$V_{ac} = \sqrt{V_{rms}^2 - \bar{V}_o^2} = \sqrt{1 - \left(\frac{2\sqrt{2}\cos\alpha}{\pi}\right)^2} \qquad (12.119)$$

and the output voltage form factor is

$$FF_v = \frac{\pi}{2\sqrt{2}\cos\alpha} \qquad (12.120)$$

Thyristor average current is $\frac{1}{2}\bar{I}_o$, while thyristor rms current rating is $I_{rms}/\sqrt{2}$. These same two thyristor expressions are valid for both continuous and discontinuous load current conditions.

Critical load inductance

From equation (12.112) set to zero (or $i = 0$ in equation (12.111)), the boundary between continuous and discontinuous inductor current must satisfy

$$\frac{R}{Z}\sin(\alpha-\phi)\tanh\left(\frac{\frac{1}{2}\pi}{\tan\phi}\right) > \frac{E}{\sqrt{2}V} \qquad (12.121)$$

Inversion

If the polarity of the back emf E is reversed as shown in figure 12.9a, waveforms as in parts b and c of figure 12.9 result. The emf supply can provide a forward bias across the bridge thyristors even after the supply polarity has gone negative. The zero current angle $\hat{\alpha}$ now satisfies $\pi < \hat{\alpha} < 3\pi/2$, as given by equation (12.98). Thus load and supply current can flow, even for $\alpha > \pi$.
The relationship between the mean output voltage and current is now given by

$$V_o = -E + \bar{I}_o R = \sqrt{2}V\frac{p}{\pi}\sin\frac{\pi}{p}\cos\alpha \text{ with } p = 2 \qquad (12.122)$$

That is, the emf term E in equations (12.98) to (12.121) is appropriately changed to $-E$.
The load current flows from the emf source and if $\alpha > \frac{1}{2}\pi$, the average load voltage is negative. Power is being delivered to the ac supply from the emf source in the load, which is an energy transfer process called *power inversion*. In general

$0 < \alpha < 90°$ $\rightarrow V_o > 0$ $P_o > 0$ $i_o > 0$ rectification
$90° < \alpha < 180°$ $\rightarrow V_o < 0$ $P_o < 0$ $i_o > 0$ inversion

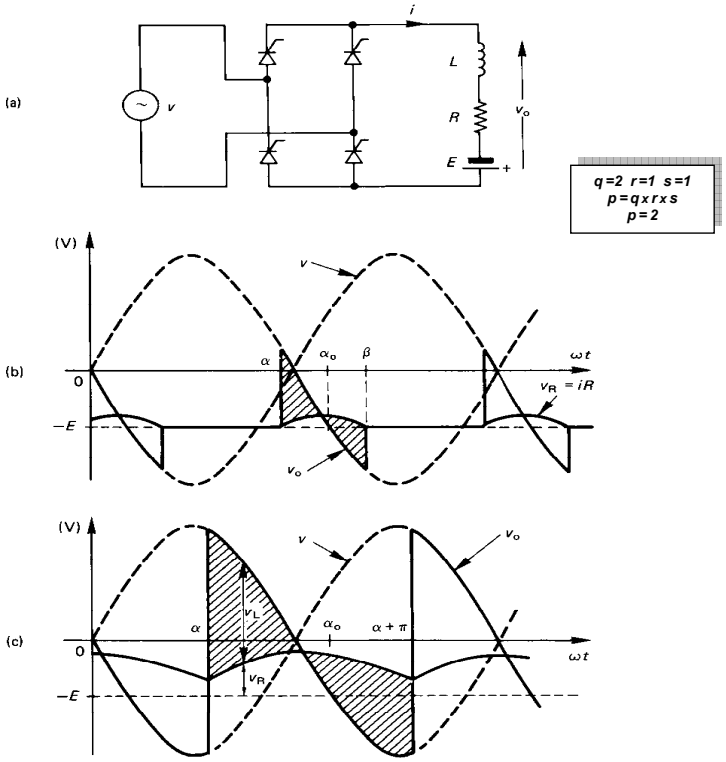


Figure 12.9. *A full-wave controlled converter with an inductive load and negative emf source: (a) circuit diagram; (b) voltage waveforms for discontinuous load current; and (c) continuous load current.*

Example 12.3: ***Single-phase, controlled converter – continuous conduction and back emf***

The fully controlled full-wave converter in figure 12.7a has a source of 240V rms, 50Hz, and a 10Ω, 50mH, 50V emf opposing series load. The delay angle is 45°.
Determine

i. the average output voltage and current
ii. the rms load voltage and the rms voltage across the *R-L* part of the load
iii. the power absorbed by the 50V load back emf
iv. the rms load current hence power dissipated in the resistive part of the load
v. the load efficiency, that is percentage of energy into the back emf and power factor
vi. the load voltage and current form and ripple factors

Solution

From example 12.2, continuous conduction is possible since $\alpha < \phi$ $(45° < 57.5°)$.

i. The average output voltage is given by equation (12.113)

$$V_o = \frac{2\sqrt{2}V}{\pi}\cos\alpha$$
$$= \frac{2\sqrt{2}\times 240}{\pi}\times\cos 45° = 152.8\text{V}$$

The average current, from equation (12.114) is

$$\bar{I}_o = \frac{V_o - E}{R} = \frac{152.8\text{V} - 50\text{V}}{10\Omega} = 10.28\text{A}$$

ii. From equation (12.85) the rms load voltage is 240V. The rms voltage across the *R-L* part of the load is

$$V_{RLrms} = \sqrt{V_{rms}^2 - E^2}$$
$$= \sqrt{240\text{V}^2 - 50\text{V}^2} = 234.7\text{V}$$

iii. The power absorbed by the 50V back emf load is

$$P = \bar{I}_o E = 10.28\text{A}\times 50\text{V} = 514\text{W}$$

iv. The *R-L* load voltage harmonics (which are even) are given by equations (12.117) and (12.118):

$$V_{oR-L} = \frac{2\sqrt{2}V}{\pi}\times\cos\alpha - E$$

$$V_n = \frac{\sqrt{2}V}{2\pi}\times\left(\frac{1}{(n-1)^2} + \frac{1}{(n+1)^2} - \frac{2\cos 2\alpha}{(n-1)(n+1)}\right) \quad \text{for} \quad n = 2,4,6,..$$

The harmonic currents and voltages are shown in the table to follow.

harmonic n	V_n	$Z_n = \sqrt{R^2 + (n\omega L)^2}$ (Ω)	$I_n = \frac{V_n}{Z_n}$ (A)	$\frac{1}{2}I_n^2$
0	***102.79***	10.00	10.28	***105.66***
2	60.02	32.97	1.82	1.66
4	8.16	63.62	0.13	0.01
6	3.26	94.78	0.04	0.00
			$I_o^2 + \sum\frac{1}{2}I_n^2 =$	107.33

From the table the rms load current is given by

$$I_{rms} = \sqrt{I_o^2 + \tfrac{1}{2}\sum I_n^2} = \sqrt{107.33} = 10.36\text{A}$$

The power absorbed by the 10Ω load resistor is

$$P_L = I_{rms}^2 R = 10.36\text{A}^2\times 10\Omega = 1073.3\text{W}$$

v. The load efficiency, that is, percentage energy into the back emf E is

$$\eta = \frac{514\text{W}}{514\text{W}+1073.3\text{W}} \times 100\% = 32.4\%$$

The power factor is

$$pf = \frac{P_L}{V_{rms} I_{rms}} = \frac{514\text{W}+1073.3\text{W}}{240\text{V} \times 10.36\text{A}} = 0.64$$

vi. The output performance factors are

$$FF_i = \frac{I_{rms}}{\bar{I}_o} = \frac{10.36\text{A}}{10.28\text{A}} = 1.011 \qquad RF_i = \sqrt{FF_i^2 - 1} = \sqrt{1.023^2 - 1} = 0.125$$

$$FF_v = \frac{V_{rms}}{\bar{V}_o} = \frac{240\text{V}}{152.8\text{V}} = 1.57 \qquad RF_v = \sqrt{FF_v^2 - 1} = \sqrt{1.57^2 - 1} = 1.211$$

Note that the voltage form factor (hence voltage ripple factor) agrees with that obtained by substitution into equation (12.120), 1.57.

♣

Example 12.4: ***Controlled converter – constant load current, back emf, and overlap***

The fully controlled single-phase full-wave converter in figure 12.7a has a source of 230V rms, 50Hz, and a series load composed of ½Ω, infinite inductance, 150V emf non-opposing. If the average load current is to be 200A, calculate the delay angle assuming the converter is operating in the inversion mode, taking into account 1mH of commutation inductance.

Solution

The mean load current is

$$\bar{I}_o = \frac{V_o(\alpha) - E}{R}$$

$$200\text{A} = \frac{V_o(\alpha) - -150\text{V}}{½\Omega}$$

which implies a load voltage $V_o(\alpha) = -50\text{V}$.

The output voltage is given by equation (12.82) $V_o = \frac{2\sqrt{2}V}{\pi}\cos\alpha$. Commutation of current from one rectifier to the other takes a finite time. The effect of commutation inductance is to reduce the output voltage, thus according to equation (12.194), the output voltage becomes

$$V_o = \frac{\sqrt{2}V}{\pi / n}\sin{}^{\pi}/_{n}\cos\alpha - n\omega L_c I_o / 2\pi \quad \text{where } n = 2$$

$$-50\text{V} = \frac{\sqrt{2}\times 230\text{V}}{\pi/2}\times\cos\alpha - 2\times 50\text{Hz}\times 1\text{mH}\times 200\text{A}$$

$$= 207\text{V}\times\cos\alpha - 20\text{V}$$

which yields $\alpha = 98.3°$. The commutation overlap causes the output voltage to reduce to zero volts and the overlap period γ is given by equation (12.195)

$$I_o = \frac{\sqrt{2}V}{2\pi f L_c}\left(\cos\alpha - \cos(\gamma+\alpha)\right)$$

$$200\text{A} = \frac{\sqrt{2}\,230\text{V}}{2\pi 50\text{Hz}\times 1\text{mH}}\left(\cos 93.8° - \cos(\gamma + 93.8°)\right)$$

This gives an overlap angle of $\gamma = 11.2°$.

♣

12.3 Three-phase half-controlled converter

Assuming three phase voltages

$$V_{an} = \vec{v}_a = V\angle 0° = V\sin\omega t$$

$$V_{bn} = \vec{v}_a = V\angle -120° = V\sin(\omega t - 120°)$$

$$V_{cn} = \vec{v}_a = V\angle +120° = V\sin(\omega t + 120°)$$

Then the line to line voltage phase a to phase b is

$$\vec{V}_{ab} = \vec{V}_a - \vec{V}_b$$

$$= \sqrt{2}V\sin\omega t - \sqrt{2}V\sin(\omega t - 120°)$$

$$= \sqrt{2}V\sin\omega t - ½\sqrt{2}V\sin\omega t - \frac{\sqrt{3}}{2}\sqrt{2}V\cos\omega t$$

$$= \sqrt{3}\,\sqrt{2}V\sin(\omega t + 30°) = \sqrt{3}\,\sqrt{2}V\sin(\omega t + {}^{\pi}/_{6})$$

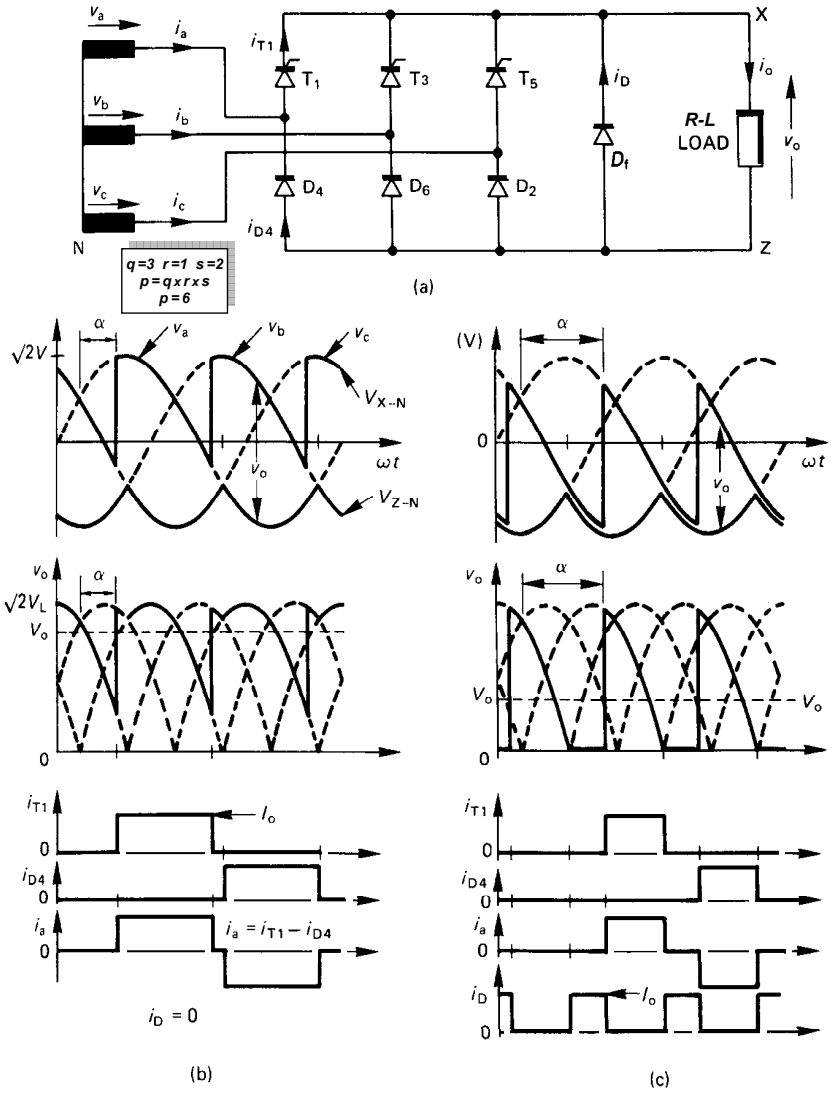


Figure 12.10. *Three-phase half-controlled bridge converter: (a) circuit connection; (b) voltage and current waveforms for a small firing delay angle α; and (c) waveforms for α large.*

The three-phase line input voltages are:

$$\begin{aligned} v_{ab} &= V_{max}\sin(\omega t + \tfrac{1}{6}\pi) \\ v_{bc} &= V_{max}\sin(\omega t - \tfrac{1}{2}\pi) \\ v_{ca} &= V_{max}\sin(\omega t + \tfrac{1}{2}\pi) \end{aligned} \qquad (12.123)$$

Figure 12.10a illustrates a half-controlled (semi-controlled) converter where half the devices are thyristors, the remainder being diodes. As in the single-phase case, a freewheeling diode can be added across the load so as to allow the bridge thyristors to commutate and decrease freewheeling losses. The output voltage expression consists of √2V 3√3/2π due to the uncontrolled half of the bridge and √2*V* 3√3 × cos α /2π due to the controlled half which is phase-controlled. The half-controlled bridge mean output is given by the sum, that is

$$\begin{aligned} V_o &= \sqrt{2}\,V\frac{3\sqrt{3}}{2\pi}(1+\cos\alpha) = \sqrt{2}\,V_L\frac{3}{2\pi}(1+\cos\alpha) \\ &= 2.34V(1+\cos\alpha) \qquad \text{(V)} \\ &\qquad 0 \le \alpha \le \tfrac{2}{3}\pi \qquad \text{(rad)} \end{aligned} \qquad (12.124)$$

$$V_o = \bar{I}_o R$$

At α = 0, $\hat{V}_o = \sqrt{2}\,V\,3\sqrt{3}/\pi = 1.35\,V_L$, as in equation (11.93). The normalised mean output voltage V_n is

$$V_n = V_o/\hat{V}_o = \tfrac{1}{2}(1+\cos\alpha) \qquad (12.125)$$

The diodes prevent any negative output, hence inversion cannot occur. Typical output voltage and current waveforms for a highly inductive load (constant current) are shown in figure 12.10b.

12.3i - α ≤ ⅓π

When the delay angle is less than ⅓π the output waveform contains six pulses per cycle, of alternating controlled and uncontrolled phases, as shown in figure 12.10b. The output current is always continuous (even for a resistive load) since no output voltage zeros occur. The rms output voltage is given by

$$\begin{aligned} V_{rms} &= \sqrt{\frac{3}{2\pi}\left\{\int_{\alpha+\pi/3}^{2\pi/3}\left(\sqrt{2}V_L\right)^2\sin^2\omega t\,d\omega t + \int_{\pi/3}^{\alpha+2\pi/3}\left(\sqrt{2}V_L\right)^2\sin^2\omega t\,d\omega t\right\}} \\ &= V_L\left(1+\frac{3\sqrt{3}}{4\pi}(1+\cos 2\alpha)\right)^{\frac{1}{2}} \end{aligned} \qquad (12.126)$$

for $0 \le \alpha \le \pi/3$

12.3ii - α ≥ ⅓π

For delay angles greater than ⅓π the output voltage waveform is made up of three controlled pulses per cycle, as shown in figure 12.10c. Although output voltage zeros result, continuous load current can flow through a diode and the conducting thyristor, or through the commutating diode if employed. The rms output voltage is given by

$$\begin{aligned} V_{rms} &= \sqrt{\frac{3}{2\pi}\int_{\alpha}^{\pi}\left(\sqrt{2}V_L\right)^2\sin^2\omega t\,d\omega t} \\ &= V_L\left(\frac{3}{2\pi}\left(\pi-\alpha+\tfrac{1}{2}\sin 2\alpha\right)\right)^{\frac{1}{2}} \end{aligned} \qquad (12.127)$$

for $\alpha \ge \pi/3$

The Fourier coefficients of the *p*-pulse output voltage are given by

$$\begin{aligned} a_n &= \frac{\sqrt{2}V}{2\pi/p}\left[\frac{-2}{n^2-1} - \frac{\cos(n+1)\alpha}{n+1} + \frac{\cos(n-1)\alpha}{n-1}\right] \\ b_n &= \frac{\sqrt{2}V}{2\pi/p}\left[\frac{\sin(n+1)\alpha}{n+1} - \frac{\sin(n-1)\alpha}{n-1}\right] \end{aligned} \qquad (12.128)$$

where $n = mp$ and m = 1, 2, 3, .. For the three-phase, full-wave, half-controlled case, p = 6, thus the output voltage harmonics occur at n = 6, 12, …

12.4 Three-phase, controlled thyristor converter circuits

The three-phase line input voltages are:

$$\begin{aligned} v_{ab} &= V_{max}\sin(\omega t + \tfrac{1}{6}\pi) \\ v_{bc} &= V_{max}\sin(\omega t - \tfrac{1}{2}\pi) \\ v_{ca} &= V_{max}\sin(\omega t + \tfrac{1}{2}\pi) \end{aligned} \qquad (12.129)$$

12.4.1 Three-phase, fully-controlled, half-wave circuit with an inductive load

When the diodes in the circuit of figure 11.10 are replaced by thyristors, as in figure 12.11a, a three-phase fully controlled half-wave converter results. The output voltage is controlled by the delay angle *α*. This angle is specified from the thyristor commutation angle, which is the earliest point the associated thyristor becomes forward-biased, as shown in parts b, c, and d of figure 12.11. (The reference is not the phase zero voltage cross-over point). The thyristor with the highest instantaneous anode potential will conduct when fired and in turning on will reverse bias and turn off any previously conducting thyristor. The output voltage ripple is three times the supply frequency and the supply currents contain dc components. Each phase progressively conducts for periods of ⅔π, displaced by *α*, as shown in figure 12.11b.

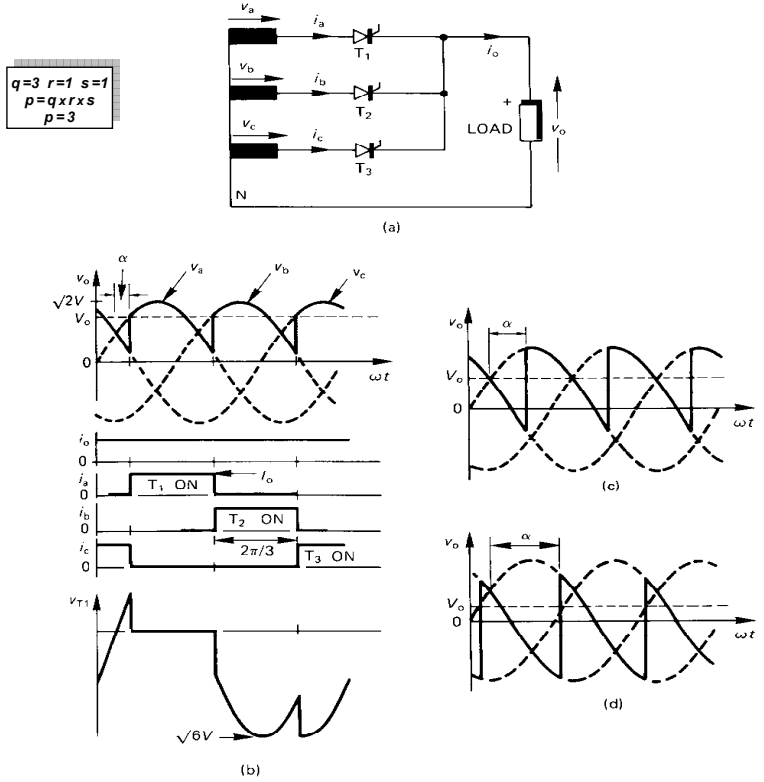


Figure 12.11. *Three-phase half-wave controlled converter:*
(a) circuit connection; (b) voltage and current waveforms for a small firing delay angle α; (c) and (d) load voltage waveforms for progressively larger delay angles.

The mean output voltage for an *n*-phase half-wave controlled converter is given by (see example 12.7)

$$\begin{aligned} V_o &= \frac{\sqrt{2}V}{2\pi/n}\int_{\alpha-\pi/n}^{\alpha+\pi/n}\cos\omega t\,d\omega t \\ &= \sqrt{2}V\frac{\sin(\pi/n)}{\pi/n}\cos\alpha \qquad \text{(V)} \end{aligned} \qquad (12.130)$$

which for the three-phase circuit considered with **continuous or discontinuous (*R*) load current** gives

$$V_o = \bar{I}_o R = \frac{3\sqrt{3}}{2\pi}\sqrt{2}V\cos\alpha = 1.17V\cos\alpha \qquad 0 \le \alpha \le \pi/6 \tag{12.131}$$

For **discontinuous conduction**, and a resistive load, the mean output voltage is

$$V_o = \bar{I}_o R = \frac{3}{2\pi}\sqrt{2}V\left(1+\cos\left(\alpha+\pi/6\right)\right) \qquad \pi/6 \le \alpha \le 5\pi/6 \tag{12.132}$$

The mean output voltage is zero for α = ½π. For 0 < α < $\frac{1}{6}$π, the instantaneous output voltage is always greater than zero. Negative average output voltage occurs when α > ½π as shown in figure 12.11d. Since the load current direction is unchanged, for α > ½π, power reversal occurs, with energy feeding from the load into the ac supply. Power inversion assumes a load with an emf to assist the current flow, as in figure 12.9. If α > π no reverse bias exists for natural commutation and continuous load current will freewheel.

The maximum mean output voltage $\hat{V}_o$ = √2*V* 3√3 /2π occurs at α = 0. The normalised mean output voltage V_n is

$$V_n = V_o/\hat{V}_o = \cos\alpha \tag{12.133}$$

With an *R-L* load, at V_o = 0, the load current falls to zero. Thus for α > ½π, continuous load current does not flow for an *R-L* load.

The rms output voltage (for inductive and resistive loads) is given by

$$V_{rms} = \sqrt{\frac{3}{2\pi}\int_{\alpha-\pi/3}^{\alpha+\pi/3}\left(\sqrt{2}V\right)^2\sin^2(\omega t)\,d\omega t}$$
$$= V\left(1+\frac{3\sqrt{3}}{4\pi}\sin 2\alpha\right)^{1/2} \qquad 0 \le \alpha \le 5\pi/6 \tag{12.134}$$

From equations (12.131) and (12.134), the ac in the output voltage is

$$V_{ac} = \sqrt{V_{rms}^2 - V_o^2} = V\left(1+\frac{3\sqrt{3}}{4\pi}\sin 2\alpha - \frac{9}{\pi}\cos^2\alpha\right)^{1/2} \tag{12.135}$$

The output voltage distortion ripple factor is

$$RF_v = \sqrt{\frac{2\pi^2}{27}+\frac{\sqrt{3}\pi}{18}\sin 2\alpha - \cos^2\alpha} \qquad (\min(\text{at } \alpha{=}0) = 0.173;\ \max(\text{at } \alpha{=}1/2n) = 0.66) \tag{12.136}$$

12.4.2 *Three-phase, half-wave converter with freewheel diode*

Figure 12.12 shows a three-phase, half-wave controlled rectifier converter circuit with a load freewheel diode, D_f. This diode prevents the load voltage from going negative, thus inversion is not possible.

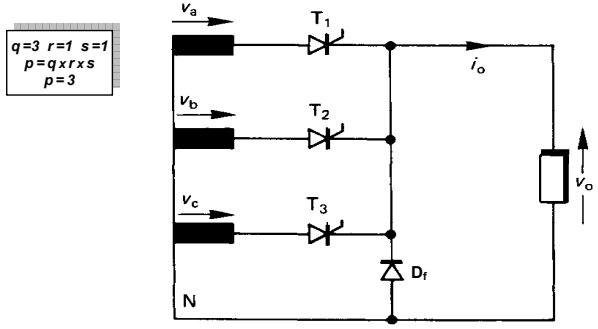


Figure 12.12. *A half-wave fully controlled three-phase converter with a load freewheel diode.*

12.4.2i - α < π/6. The output is as in figure 12.11b, with no voltage zeros occurring. The mean output voltage (and current) is given by equation (12.131), that is

$$V_o = \bar{I}_o R = \frac{3\sqrt{3}}{2\pi}\sqrt{2}V\cos\alpha = 1.17V\cos\alpha \qquad \text{(V)} \qquad 0 \le \alpha \le \pi/6 \qquad \text{(rad)} \tag{12.137}$$

The maximum mean output V_o = √2*V* 3√3/2π occurs at α = 0. The normalised mean output voltage, V_n is given by

$$V_n = V_o/\hat{V}_o = \cos\alpha \tag{12.138}$$

The Fourier coefficients of the 3-pulse output voltage are given by (12.128). For the three-phase, half-wave, half-controlled case, *p* = 3, thus the output voltage harmonics occur at *n* = 3, 6, 9, …

12.4.2ii - α > π/6. Because of the freewheel diode, *v*oltage zeros occur and the negative portions in the waveforms in parts c and d of figure 12.11 do not occur. The mean output voltage is given by

$$V_o = \bar{I}_o R = \frac{\sqrt{2}V}{2\pi/3}\int_{\alpha-\pi/6}^{\pi}\sin\omega t\,d\omega t$$
$$= \frac{\sqrt{2}V}{2\pi/3}\left(1+\cos\left(\alpha+\pi/6\right)\right) \qquad \text{(V)} \qquad \pi/6 \le \alpha \le 5\pi/6 \tag{12.139}$$

The normalised mean output voltage V_n is

$$V_n = V_o/\hat{V}_o = [1 + \cos(\alpha+\pi/6)]/\sqrt{3} \tag{12.140}$$

The average load current (with an emf *E* in the load) is given by

$$\bar{I}_o = \frac{V_o - E}{R} \tag{12.141}$$

These equations assume continuous load current.

12.4.2iii - α > 5π/6. A delay angle of greater than *5π/6* would imply a negative output voltage, clearly not possible with a freewheel load diode.

Example 12.5: ***Three-phase half-wave rectifier with freewheel diode***

The half-wave three-phase rectifier in figure 12.12 has a three-phase 415V 50Hz source (240V phase), and a 10Ω resistor and infinite series inductance as a load. If the delay angle is 60° determine the load current and output voltage if:

i. the phase commutation inductance is zero
ii. the phase commutation reactance is¼Ω

Solution

i. The output voltage, without any line commutation inductance and a 60° phase delay angle, is given by equation (12.139)

$$V_o = \bar{I}_o R = \frac{\sqrt{2}V}{2\pi/3}\left(1+\cos\left(\alpha+\pi/6\right)\right)$$
$$= \frac{\sqrt{2}\,240\text{V}}{2\pi/3}\left(1+\cos\left(60^\circ+\pi/6\right)\right) = 162\text{V}$$

The constant load current is therefore

$$I_o = \frac{V_o}{R} = \frac{162\text{V}}{10\Omega} = 16.2\text{A}$$

ii. When the current changes paths, any inductance will control the rate at which the commutation from one path to the next occurs. The voltage drops across the commutating inductors modifies the output voltage. Since the voltage across the freewheel diode is not associated with commutation inductance, the output voltage is not effected when the current swaps from a phase to the freewheel diode. But when the current transfers from the diode to a phase, while the commutation inductance current in the phase is building up to the constant load current level, the output remains clamped at the diode voltage level, viz. zero. The average voltage across the load during this overlap period is therefore reduced. The commutation current is defined by

$$\sqrt{2}V\sin\omega t = L_c\frac{di_c}{dt} = X_c\frac{di_c}{d\omega t}$$
$$i_c = \frac{\sqrt{2}V}{X_c}\left(\cos\left(\alpha+\frac{\pi}{6}\right)-\cos\omega t\right)$$

Solving for when the current rises to the load current I_o^γ gives

$$I_o^\gamma = \frac{\sqrt{2}V}{X_c}\left(\cos\left(\alpha+\frac{\pi}{6}\right)-\cos\left(\alpha+\frac{\pi}{6}+\gamma\right)\right)$$

but

$$I_o^\gamma = \frac{V_o^\gamma}{R} = \frac{\sqrt{2}V}{R\,2\pi/3}\left(1+\cos\left(\alpha+\pi/6+\gamma\right)\right)$$

$$\frac{\cos\left(\alpha+\frac{\pi}{6}\right)-\frac{X_c}{R\,2\pi/3}}{\frac{X_c}{R\,2\pi/3}+1} = \cos\left(\alpha+\frac{\pi}{6}+\gamma\right)$$

$$\gamma = \cos^{-1}\left(\frac{\cos\left(\alpha+\frac{\pi}{6}\right)-\frac{X_c}{R\,2\pi/3}}{\frac{X_c}{R\,2\pi/3}+1}\right)-\left(\alpha+\frac{\pi}{6}\right) = 0.68°$$

The load current and voltage are therefore

$$I_o^\gamma = \frac{\sqrt{2}V}{X_c}\left(\cos\left(\alpha+\frac{\pi}{6}\right)-\cos\left(\alpha+\frac{\pi}{6}+\gamma\right)\right) = \frac{\sqrt{2}\,240V}{¼\Omega}\left(\cos(90°)-\cos(90.68)\right) = 16.11\text{A}$$

$$V_o^\gamma = I_o^\gamma R = 16.11\text{A}\times 10\Omega = 161.1\text{V}$$

♣

12.4.3 *Three-phase, full-wave, fully-controlled circuit with an inductive load*

A three-phase bridge is fully controlled when all six bridge devices are thyristors, as shown in figure 12.13a. The frequency of the output ripple voltage is six times the supply frequency and each thyristor always conducts for ⅔π. Circuit waveforms are shown in figure 12.13b. The output voltage is continuous, and the mean output voltage for both inductive and resistive loads is given by

$$V_o = \frac{3}{\pi}\int_{\alpha+\pi/6}^{\alpha+½\pi}\sqrt{3}\sqrt{2}V\sin\left(\omega t+\pi/6\right)d\omega t$$

$$= \frac{3\sqrt{3}}{\pi}\sqrt{2}V\cos\alpha = 2.34V\cos\alpha \qquad \text{(V)} \qquad (12.142)$$

$$0\le\alpha\le 2\pi/3$$

which is twice the voltage given by equation (12.131) for the half-wave circuit, but for a purely resistive load the output voltage is discontinuous and equation (12.142) becomes

$$V_o = \frac{3}{\pi}V_{\max}\left[1+\cos\left(\alpha+\pi/6\right)\right] = \frac{3\sqrt{3}}{\pi}\sqrt{2}V\left[1+\cos\left(\alpha+\pi/6\right)\right] \qquad \text{(V)} \qquad (12.143)$$

$$\pi/3\le\alpha\le 2\pi/3$$

The average output current is given by $\bar{I}_o = V_o/R$ in each case. If a load back emf exists the average current becomes

$$\bar{I}_o = \frac{V_o - E}{R} \qquad (12.144)$$

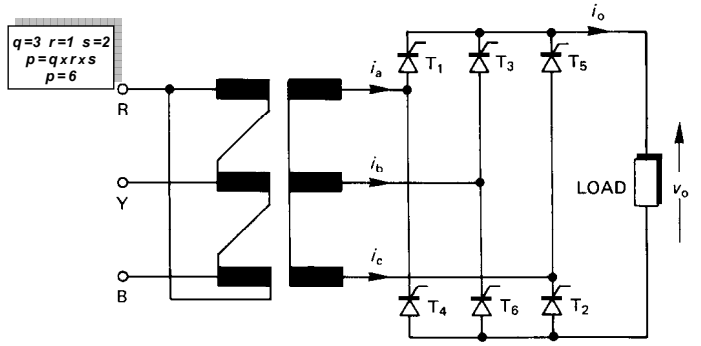
The maximum mean output voltage $\hat{V}_o = \sqrt{2}V\,3\sqrt{3}/\pi$ occurs at $\alpha = 0$. The normalised mean output V_n is

$$V_n = V_o/\hat{V}_o = \cos\alpha \qquad (12.145)$$

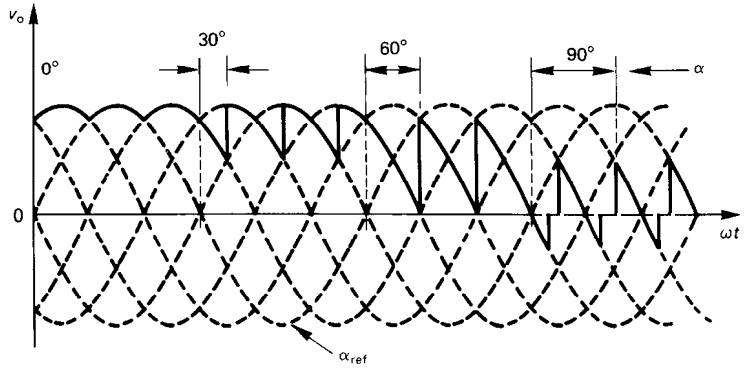
For delay angles up to ⅓π, the output voltage is at all instances non-zero, hence the load current is continuous for any passive load (both resistive and inductive). Beyond ⅓π the load current may be discontinuous (always discontinuous for a resistive load). For $\alpha > ½\pi$ the current is always discontinuous for passive loads (no back emf, E) and the average output voltage is less than zero

For **continuous load current**, the load current is given by

$$i(\omega t) = \frac{\sqrt{3}\sqrt{2}V}{Z}\sin\left(\omega t+\pi/6-\phi\right)-\frac{E}{R}+\frac{\sqrt{3}\sqrt{2}V}{Z}\sin(\alpha-\phi)\frac{e^{-\omega t+\pi/6+\alpha/\tan\phi}}{e^{-\pi/3\tan\phi}-1} \qquad (12.146)$$



(a)



(b)

Figure 12.13. *A three-phase fully controlled converter: (a) circuit connection and (b) load voltage waveform for four delay angles.*

The maximum and minimum ripple current magnitudes are

$$\hat{I} = \frac{\sqrt{3}\sqrt{2}V}{Z}\sin\left(\alpha+\pi/2-\phi\right)-\frac{E}{R}+\frac{\sqrt{3}\sqrt{2}V}{Z}\sin(\alpha-\phi)\frac{e^{\pi/6/\tan\phi}}{e^{\pi/3/\tan\phi}-1} \qquad (12.147)$$

at $\omega t = \alpha + \frac{1}{18}n\pi$ for n = 0, 6, 12, ..

$$\check{I} = \frac{\sqrt{3}\sqrt{2}V}{Z}\sin\left(\alpha+\pi/3-\phi\right)-\frac{E}{R}+\frac{\sqrt{3}\sqrt{2}V}{Z}\sin(\alpha-\phi)\frac{1}{e^{\pi/3/\tan\phi}-1} \qquad (12.148)$$

at $\omega t = \alpha - \frac{1}{6}\pi + \frac{1}{18}n\pi$ for n = 0, 6, 12, ..

With a load back emf the critical inductance for continuous load current must satisfy $\check{I} = 0$ in equation (12.148), that is

$$\frac{R}{Z}\times\left[\sin\left(\alpha-\phi+⅓\pi\right)+\frac{\sin(\alpha-\phi)}{e^{-\pi/3\tan\phi}-1}\right]\ge\frac{E}{\sqrt{3}\sqrt{2}V} \qquad (12.149)$$

where $\tan\phi = \omega L/R$.

The rms value of the output voltage for an inductive and purely resistive load is given by

$$V_{rms} = \left(\frac{3}{\pi}\int_{\alpha-\pi/6}^{\alpha+\pi/2}3\left(\sqrt{2}V\right)^2\sin^2(\omega t)\,d\omega t\right)^{½} \qquad 0\le\alpha\le\pi/3 \qquad (12.150)$$

$$= \sqrt{3}\sqrt{2}V\sqrt{½+\frac{3\sqrt{3}}{4\pi}\cos 2\alpha}$$

but for a purely resistive load

$$V_{rms} = \sqrt{3}\sqrt{2}\,V\sqrt{1-\frac{3\alpha}{2\pi}-\frac{3}{4\pi}\sin\left(2\alpha-\pi/3\right)} \qquad \pi/3 \le \alpha \le 2\pi/3 \qquad (12.151)$$

The output voltage ripple factor (with continuous current) is

$$RF_v = \sqrt{\frac{\pi^2}{18}+\frac{\sqrt{3}\,\pi}{12}\cos 2\alpha - \cos^2\alpha} \qquad (\min(\text{at } \alpha{=}0) = 0.025;\ \max(\text{at } \alpha{=}\tfrac{1}{2}n) = 0.3) \qquad (12.152)$$

The normalise voltage harmonic peak magnitudes in the output voltage, with continuous load current, are

$$V_{Ln} = \sqrt{2}\,V\,\frac{3\sqrt{3}}{\pi}\left(\frac{1}{(n-1)^2}+\frac{1}{(n+1)^2}-\frac{2\cos 2\alpha}{(n-1)(n+1)}\right)^{\frac{1}{2}} \qquad (12.153)$$

for n = 6, 12, 18…

The harmonics occur at multiples of six times the fundamental frequency.

For **discontinuous load current**, at high delay angles, when the output current becomes discontinuous with an inductive load, the output current is given by

$$i(\omega t) = \frac{\sqrt{3}\sqrt{2}\,V}{Z}\left[\sin\left(\omega t+\pi/6-\phi\right)-\sin\left(\alpha+\pi/3-\phi\right)e^{-\omega t-\pi/6+\alpha/\tan\phi}\right]-\frac{E}{R}\left[1-e^{-\omega t-\pi/6+\alpha/\tan\phi}\right] \qquad (12.154)$$

$$\alpha \le \omega t \le \alpha+\theta_c$$

where θ_c is the conduction period, which is found by solving the transcendental equation formed when in equation (12.154), $i(\omega t = \alpha+\frac{1}{6}\pi+\theta_c) = 0$. The average output voltage can then be found from

$$V_o = \frac{3\sqrt{3}\sqrt{2}\,V}{\pi}\left[\cos\left(\alpha+\pi/3\right)-\cos\left(\alpha+\pi/3+\theta_c\right)\right]-\frac{3E}{\pi}\left[\pi/3-\theta_c\right] \qquad (12.155)$$

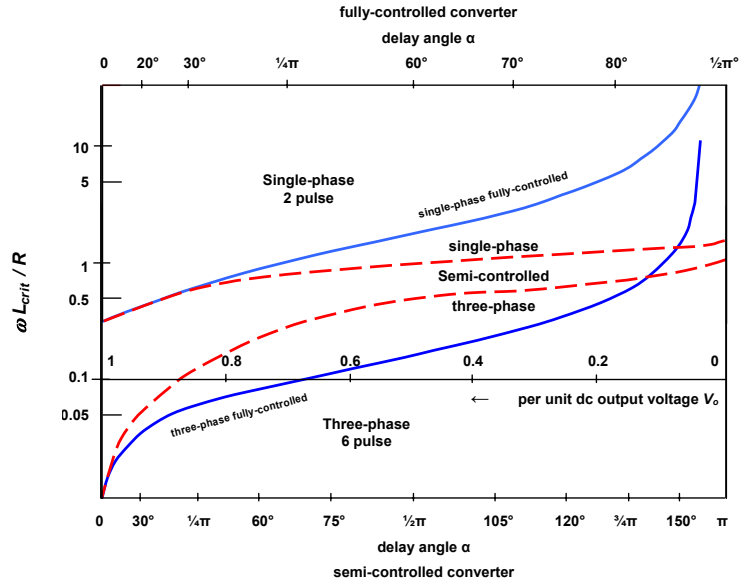


Figure 12.14. *Critical load inductance (reactance) of single-phase (two pulse) and three-phase (six pulse), semi-controlled and fully-controlled converters, as a function phase delay angle α whence dc output voltage V_o. For rectifier, α = 0.*

12.4.3i - Resistive load

For a resistive load, the load voltage harmonics for p pulses per cycle, are given by

$$a_n = \frac{\sqrt{2}V}{2\pi/p}\left[\frac{-2}{n^2+1}-\frac{\cos(n+1)\alpha}{n+1}-\frac{\cos(n-1)\alpha}{n-1}\right]$$
$$b_n = \frac{\sqrt{2}V}{2\pi/p}\left[\frac{\sin(n+1)\alpha}{n+1}-\frac{\sin(n-1)\alpha}{n-1}\right] \qquad (12.156)$$

for $n = pm$ and m = 1, 2, 3, .. The harmonics occur at multiples of six times the fundamental frequency, for a 6 - pulse (p = 6) per cycle output voltage.

12.4.3ii - Highly inductive load – constant load current

As with a continuous load current, with a constant load current the input current comprises ⅔π alternating polarity blocks of current, with each phase displaced relative to the others by ⅔π, independent of the thyristor triggering delay angle. At maximum voltage hence maximum power output, the delay angle is zero and the phase voltage and current fundamental are in phase. As the phase angle is increased, the inverter output voltage, hence power output is decreased, and the line current block of current (fundamental) shifts by α with respect to the line voltage. Reactive input power increases as the real power decreases. At α = ½π, the output voltage reduces to zero, the output power is zero, and the ⅔π current blocks in the ac input are shifted ½π with respect to the line voltage, producing only VAr's from the ac input. When the delay angle is increased above ½π, the inverter dc output reverses polarity and energy transfers back into the ac supply (inversion), with maximum inverted power reached at α = π, where the reactive VAr is reduced to zero, from a maximum at α = ½π.

For a highly inductive load, that is a constant load current $\bar{I}_o$:

the mean diode current is

$$\bar{I}_{Th} = \tfrac{1}{n}\bar{I}_o = \tfrac{1}{3}\bar{I}_o \qquad \text{(A)} \qquad (12.157)$$

and the rms diode current is

$$I_{Th\,rms} = \tfrac{1}{\sqrt{n}}I_{o\,rms} \approx \tfrac{1}{\sqrt{n}}\bar{I}_o = \tfrac{1}{\sqrt{3}}\bar{I}_o \qquad \text{(A)} \qquad (12.158)$$

The diode current form factor is

$$FF_{ITh} = I_{Th\,rms}/\bar{I}_{Th} = \sqrt{3} \qquad (12.159)$$

The diode current ripple factor is

$$RF_{ITh} = \sqrt{FF_{ITh}^2-1} = \sqrt{2} \qquad (12.160)$$

The rms input line currents are

$$I_{L\,rms} = \sqrt{\frac{2}{3}}\,I_{o\,rms} = 0.816\,I_{o\,rms} \qquad (12.161)$$

A phase voltage is given by

$$v_a = \sqrt{2}\,V\sin\omega t \qquad (12.162)$$

with phases b and c shifted by ⅔π. That is substitute ωt with $\omega t \pm ⅔\pi$.

The line current i_a is given by

$$i_a(\omega t-\phi_1) = \frac{2\sqrt{3}}{\pi}I_o\left(\sin\omega t-\frac{1}{5}\sin 5\omega t-\frac{1}{7}\sin 7\omega t+\frac{1}{11}\sin 11\omega t+\frac{1}{13}\sin 13\omega t\right.$$
$$\left.-\frac{1}{17}\sin 17\omega t-\frac{1}{19}\sin 19\omega t+.....\right) \qquad (12.163)$$

where Φ_1 is the angle between the supply v_a and the fundamental line current i_{a1}.

From Fourier coefficients of the line current harmonics are

$$a_n = \frac{1}{\pi}\int_0^{2\pi} i_l(t)\cos n\omega t\ d\omega t$$
$$= \frac{1}{\pi}\left[\int_{\frac{\pi}{6}+\alpha}^{\frac{5\pi}{6}+\alpha}\bar{I}_o\cos n\omega t\ d\omega t-\int_{\frac{7\pi}{6}+\alpha}^{\frac{11\pi}{6}+\alpha}\bar{I}_o\cos n\omega t\ d\omega t\right] \qquad (12.164)$$
$$= -\frac{4\bar{I}_o}{n\pi}\sin\frac{n\pi}{3}\sin n\alpha \qquad n = 1, 3, 5, \ldots \text{odd}$$

$$b_n = \frac{1}{\pi}\int_0^{2\pi} i_s(t)\sin n\omega t\, d\omega t$$
$$= \frac{1}{\pi}\left[\int_{\frac{\pi}{6}+\alpha}^{\frac{5\pi}{6}+\alpha} \bar{I}_o \sin n\omega t\, d\omega t - \int_{\frac{7\pi}{6}+\alpha}^{\frac{11\pi}{6}+\alpha} \bar{I}_o \sin n\omega t\, d\omega t\right] \qquad (12.165)$$
$$= \frac{4\bar{I}_o}{n\pi}\sin\frac{n\pi}{3}\cos n\alpha \qquad n = 1, 3, 5, \ldots \text{odd}$$

The input line current, which has no dc component ($a_n = 0$), is

$$i_l(t) = \sum_{n=1,3,5,\ldots}^{\infty} \frac{4\bar{I}_o}{n\pi}\sin\frac{n\pi}{3}\sin(n\omega t - n\alpha) \qquad (12.166)$$

The fundamental input current and rms are

$$I_{l1}(t) = \frac{2\sqrt{3}\,\bar{I}_o}{\pi}\sin(\omega t - \alpha) = \sqrt{2}\; I_{1rms}\sin(\omega t - \alpha) \qquad (12.167)$$

where $I_{1rms} = \sqrt{3}\frac{\sqrt{2}}{\pi}\bar{I}_o$

The power factor for a constant load current is

$$pf = \frac{\sqrt{3}V_{rms}I_{1rms}\cos\alpha}{\sqrt{3}V_{rms}\times\sqrt{\frac{2}{3}}\bar{I}_o} = \frac{\sqrt{3}V_{rms}\sqrt{3}\frac{\sqrt{2}}{\pi}\bar{I}_o\cos\alpha}{\sqrt{3}V_{rms}\times\sqrt{\frac{2}{3}}\bar{I}_o} = \frac{3}{\pi}\cos\alpha = 0.995\cos\alpha \qquad (12.168)$$

where $\cos\alpha$ is the displacement factor, DF, the cosine of the angle between the fundamental input voltage and current, Φ_1, as given in equation (12.167).
The supply fundamental apparent power, S_1, active power P and reactive power Q, are given by

$$S_1 = \sqrt{3}VI_{1rms} = 3V\frac{\sqrt{2}}{\pi}\bar{I}_o = \frac{3\sqrt{2}V}{\pi}\bar{I}_o = \sqrt{P_d^2 + Q_q^2}$$
$$P_d = P_1 = S_{\max}\cos\alpha \qquad (12.169)$$
$$Q_q = Q_1 = S_{\max}\sin\alpha \qquad \text{where } P_{\max} = \left|\frac{3\sqrt{2}V}{\pi}\bar{I}_o\right| = |Q_{\max}| = |S_{\max}|$$

The apparent power drawn by the 6-pulse converter, for a constant load current is

$$S_d = P_d + j\,Q_d$$
$$= S_{\max}(\cos\alpha + j\,\sin\alpha) \qquad (12.170)$$

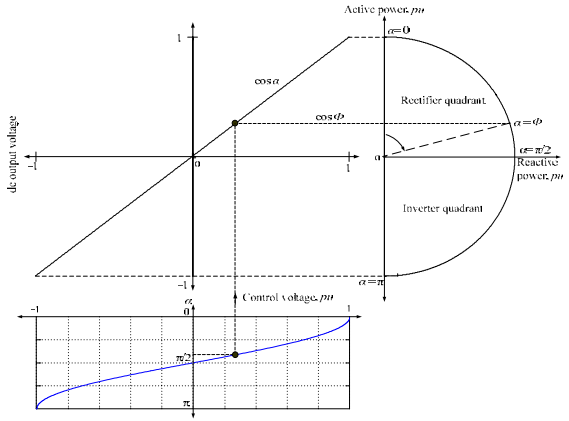


Figure 12.15. *Power locus of 6-pulse converter and per unit output voltage.*

The supply apparent power is constant for a given constant load current, independent of the thyristor turn-on delay angle. Maximum power is drawn for zero delay angle, while maximum apparent power is drawn at α = ½π. Diving equation (12.170) by S_1 gives the system power locus in per unit. The semicircle shown in figure 12.15 with centre '0' and a radius of 1 pu is the *P*-*Q* locus of the 6-pulse converter obtained by varying α from 0 to π in equation (12.170). The pu output voltage is cosα.

The supply power factor, equation (12.168), is defined as the ratio of the supply power delivered *P*, to apparent supply power *S*,

$$pf = \frac{P}{S} = \frac{I_{s1}}{I_s}\cos\alpha = \frac{3}{\pi}\cos\alpha = DF \times DPF = \frac{DPF}{\sqrt{1+THD^2}} \qquad (12.171)$$

where $DPF = \cos\alpha = \cos\Phi_1$, $DF = 3/\pi$, and the total harmonic input current distortion, *THD*, from equations (12.161) and (12.167) is

$$THD = \frac{\sqrt{I_s^2 - I_{s1}^2}}{I_{s1}} = \sqrt{\frac{I_s^2}{I_{s1}^2} - 1} = \sqrt{\left(\frac{\pi}{3}\right)^2 - 1} = 0.311 \qquad (12.172)$$

The dc-side voltage harmonics of the 6-pulse converter are generated at 6n times the fundamental line frequency. The output voltage, V_o, can be expressed as a Fourier series (see equation (12.153)):

$$V_o = \frac{3\sqrt{2}V_L}{\pi}\left(\cos\alpha + \sum_{n=1}^{\infty}\sqrt{\frac{1}{(6n-1)^2} + \frac{1}{(6n+1)^2} + \frac{2\cos 2\alpha}{(6n-1)(6n+1)}}\;\sin(6n\omega t + \lambda_{6n})\right) \qquad (12.173)$$

where

$$\lambda_{6n} = -n\pi + \tan^{-1}\left[\frac{\frac{\cos(6n+1)\alpha}{6n+1} - \frac{\cos(6n-1)\alpha}{6n-1}}{\frac{\sin(6n+1)\alpha}{6n+1} - \frac{\sin(6n-1)\alpha}{6n-1}}\right]$$

Undesirably, if triggering pulses to all the thyristors are removed, the dc current decays slowly and uncontrolled to zero through the last pair of thyristors that were triggered. Converter shut down is best achieved regeneratively by increasing (and controlling) the delay angle to greater than ½π such that the output voltage goes negative, which results in controlled power inversion back into the ac supply.
Series and parallel connection of fully-controlled, phase-shifted converters is considered in Chapter 19, in relation to HVDC transmission.

12.4.3iii - R-L load with load EMF, E

The load current during the interval $\alpha \le \omega t \le \alpha + ⅓\pi$ is defined by

$$Ri + L\frac{di}{dt} + E = \sqrt{2}V_L\sin(\omega t + ⅓\pi) \qquad (12.174)$$

which yields

$$i(t) = \frac{\sqrt{2}V_L}{Z}\left[\frac{\sin(\phi-\alpha)}{1-e^{\frac{-\pi}{3\tan\phi}}}e^{\frac{-\omega t+\alpha}{\tan\phi}} + \sin(\omega t + ⅓\pi - \phi) - \frac{\sin\alpha}{\cos\phi}\right] \qquad (12.175)$$

where $Z = \sqrt{R^2 + \omega^2L^2}$; $\tan\phi = \omega L/R$; $R = Z\cos\phi$; and $E = \sqrt{2}V_L\sin\alpha$.

Example 12.6: ***Three-phase full-wave controlled rectifier with constant output current***

The full-wave three-phase controlled rectifier in figure 12.13a has a three-phase 415V 50Hz source (240V phase), and provides a 100A constant current load.
Determine:

i. the average and rms thyristor current
ii. the rms and fundamental line current
iii. the apparent fundamental power S_1

If 25kW is delivered to the dc load, calculate:

iv. the supply power factor
iv. the dc output voltage, load resistance, hence the converter phase delay angle
v. the real active and reactive Q_1 ac supply power
vi. the delay angle range if the ac supply varies by ±5% (with 25kW and 100A dc).

Solution

i. From equations (12.157) and (12.158) the thyristor average and rms currents are

$$\bar{I}_{Th} = \tfrac{1}{3}\bar{I}_o = \tfrac{1}{3}\times 100\text{A} = 33\tfrac{1}{3}\text{A}$$

$$I_{Thrms} = \sqrt{\tfrac{1}{3}}\,\bar{I}_o = \sqrt{\tfrac{1}{3}}\times 100\text{A} = 57.7\text{A}$$

ii. The rms and fundamental line currents are

$$I_{L\,rms} = \sqrt{\frac{2}{3}}\,I_{o\,rms} = \sqrt{\frac{2}{3}}\times 100\text{A} = 81.6\text{A}$$

$$I_{1rms} = \sqrt{3}\,\frac{\sqrt{2}}{\pi}\,\bar{I}_o = \sqrt{3}\,\frac{\sqrt{2}}{\pi}\times 100\text{A} = 78.0\text{A}$$

iii. The apparent power is

$$S_1 = \sqrt{3}VI_{1rms} = \sqrt{3}\times 415\text{V}\times 78\text{A} = 56.1\text{kVA}$$

iv. The supply power factor, from equation (12.168), is

$$pf = \frac{P_L}{\sqrt{3}\,V_{rms}\,I_{rms}} = \frac{25\text{kW}}{\sqrt{3}\times 415\text{V}\times 81.6\text{A}} = 0.426 \qquad \left(=\frac{3}{\pi}\cos\alpha\right)$$

v. The output voltage is

$$V_o = \text{power}\big/\bar{I}_o = 25\text{kW}\big/100\text{A} = 250\text{V dc}$$

The load resistance is

$$R_L = \frac{V_o}{\bar{I}_o} = \frac{250\text{V}}{100\text{A}} = 2.5\Omega$$

Thyristor delay angle is given by equation (12.142), that is

$$V_o = 2.34V\cos\alpha$$

$$250\text{Vdc} = 2.34\times\frac{415\text{V}}{\sqrt{3}}\times\cos\alpha$$

which yields a delay angle of α = 1.11rad = 63.5°

vi. For a constant output power at 100A dc, the output voltage must be maintained at 250V dc independent of the ac input voltage magnitude, thus for equation (12.142)

$$\alpha = \cos^{-1}\frac{250\text{Vdc}}{2.34\times(415\pm 5\%)/\sqrt{3}}$$

$$\check{\alpha} = \cos^{-1}\frac{250\text{Vdc}}{2.34\times(415-5\%)/\sqrt{3}} = 1.08\,\text{rad} = 61.9°$$

$$\hat{\alpha} = \cos^{-1}\frac{250\text{Vdc}}{2.34\times(415+5\%)/\sqrt{3}} = 1.13\,\text{rad} = 64.9°$$

♣

12.4.4 *Three-phase, full-wave converter with freewheel diode*

Both half-controlled and fully controlled converters can employ a discrete load freewheel diode. These circuits have the voltage output characteristic that the output voltage can never go negative, hence power inversion is not possible. Figure 12.16 shows a fully controlled three-phase converter with a freewheel diode D. Thyristor/diode variations similar to those shown in figure 12.1 are possible.

- The freewheel diode is active for $\alpha > \tfrac{1}{3}\pi$. The output is as in figure 12.13b for $\alpha < \tfrac{1}{3}\pi$. The mean output voltage is

$$V_o = \bar{I}_o R = \frac{3\sqrt{3}}{\pi}\sqrt{2}V\cos\alpha = 2.34V\cos\alpha \qquad \text{(V)}$$

$$0 \le \alpha \le \pi/3 \qquad \text{(rad)} \qquad (12.176)$$

The maximum mean output voltage $\hat{V}_o = \sqrt{2}V\,3\sqrt{3}/\pi$ occurs at α = 0.

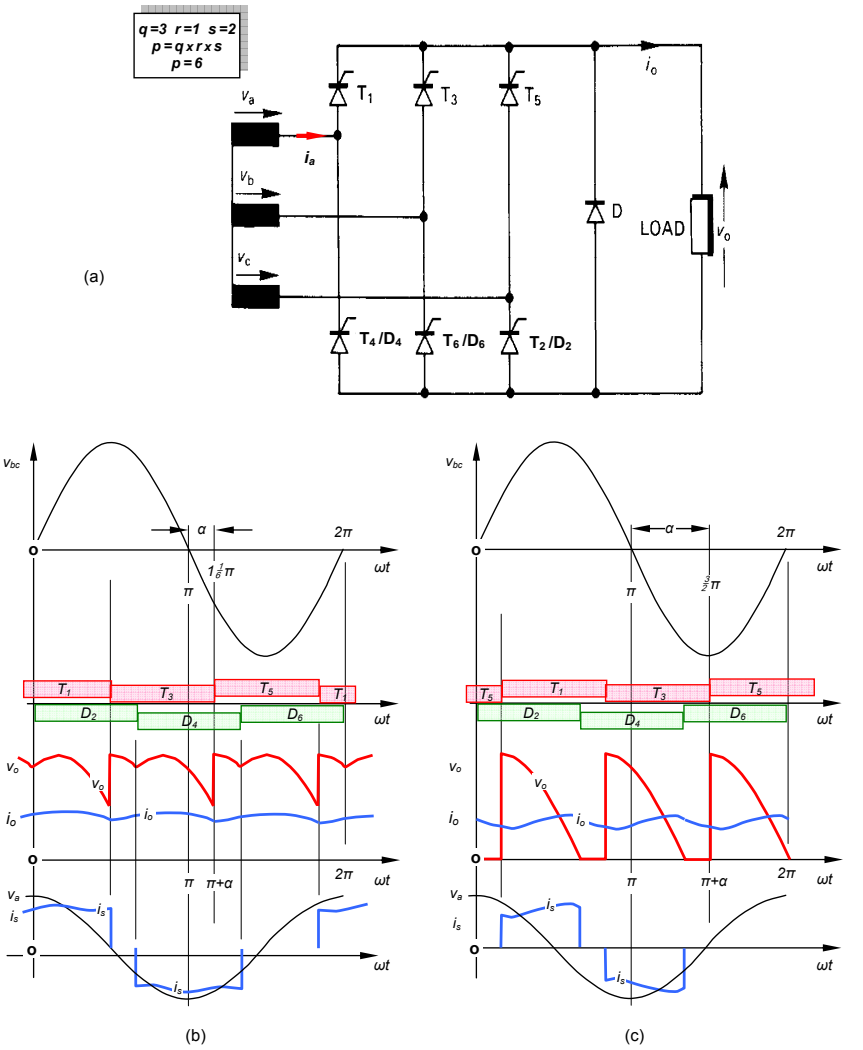


Figure 12.16. *A full-wave three-phase half-controlled converter with a load freewheeling diode: (a) circuit; (b) α = ⅙π; and (c) α = ½π.*

The normalised mean output voltage V_n is given by

$$V_n = V_o/\hat{V}_o = \cos\alpha \qquad (12.177)$$

The rms output voltage is

$$V_{rms} = \sqrt{3}\sqrt{2}V\sqrt{\frac{3}{\pi}\left(\pi-\alpha+\tfrac{1}{2}\sin 2\alpha\right)} \qquad (12.178)$$

- while

$$V_o = \bar{I}_o R = \frac{3\sqrt{3}}{\pi}\sqrt{2}V\left(1+\cos\left(\alpha+\pi/3\right)\right) \qquad \text{(V)}$$

$$\pi/3 \le \alpha \le 2\pi/3 \qquad \text{(rad)} \qquad (12.179)$$

The normalised mean output, V_n, is

$$V_n = V_o / \hat{V}_o = 1 + \cos(\alpha + \pi/3) \qquad (12.180)$$

The rms output voltage, assuming continuous conduction, is

$$V_{rms} = \sqrt{3}\sqrt{2}V\sqrt{\frac{3}{\pi}\left(2\pi/3 + \sqrt{3}\cos^2\alpha\right)} \qquad (12.181)$$

- while

$$V_o = 0 \quad \text{(V)}$$
$$2\pi/3 \le \alpha \quad \text{(rad)} \qquad (12.182)$$

In each case the average output current is given by $\bar{I}_o = V_o / R$, which can be modified to include any load back emf, that is, $\bar{I}_o = (V_o - E)/R$.

With continuous load current, with an *R-L-E* load, the closed form solution for the current is

For $0 \le \alpha \le \pi/3$

$$i(t) = \frac{\sqrt{2}V}{Z}\left[\frac{e^{\frac{-\omega t+\alpha+\pi/2}{\tan\phi}}}{1-e^{\frac{-2\pi}{3\tan\phi}}}\sin(\phi-\alpha) + \frac{e^{\frac{-\omega t}{\tan\phi}}}{1-e^{\frac{-2\pi}{3\tan\phi}}}\sin\phi + \sin(\omega t + \pi/3 - \phi) - \frac{\sin\alpha}{\cos\phi}\right] \qquad (12.183)$$

For $\pi/3 \le \alpha \le 2\pi/3$

$$i(t) = \frac{\sqrt{2}V}{Z}\left[\left\{e^{\frac{-\omega t+\alpha+\pi/2}{\tan\phi}} + \frac{e^{\frac{-\omega t+\alpha+\pi/2}{\tan\phi}}}{1-e^{\frac{-2\pi}{3\tan\phi}}}\right\}\sin(\phi-\alpha) + \frac{e^{\frac{-\omega t}{\tan\phi}}}{1-e^{\frac{-2\pi}{3\tan\phi}}}\sin\phi + \sin(\omega t - \phi) - \frac{\sin\alpha}{\cos\phi}\right] \qquad (12.184)$$

where $Z = \sqrt{R^2 + \omega^2 L^2}$; $\tan\phi = \omega L / R$; $R = Z\cos\phi$; and $E = \sqrt{2}V_L \sin\alpha$.

Example 12.7: ***Converter average load voltage***

Derive a general expression for the average load voltage of an *p*-pulse controlled converter.

Solution

Figure 12.17 defines the general output voltage waveform where *p* is the output pulse number per cycle of the ac supply. From the output voltage waveform

$$V_o = \frac{1}{2\pi/p}\int_{-\pi/n+\alpha}^{\pi/n+\alpha}\sqrt{2}V\cos\omega t\, d\omega t$$
$$= \frac{\sqrt{2}V}{2\pi/p}\left(\sin(\alpha+\pi/p) - \sin(\alpha-\pi/p)\right)$$
$$= \frac{\sqrt{2}V}{2\pi/p}2\sin(\pi/p)\cos\alpha$$
$$V_o = \frac{\sqrt{2}V}{\pi/p}\sin(\pi/p)\cos\alpha$$
$$= \hat{V}_o\cos\alpha \quad \text{(V)}$$

where

for *p* = 2 for the single-phase (*n* = 1) full-wave controlled converter in figure 12.7.
for *p* = 3 for the three-phase (*n* = 3) half-wave controlled converter in figure 12.11.
for *p* = 6 for the three-phase (*n* = 3) full-wave controlled converter in figure 12.13.

♣

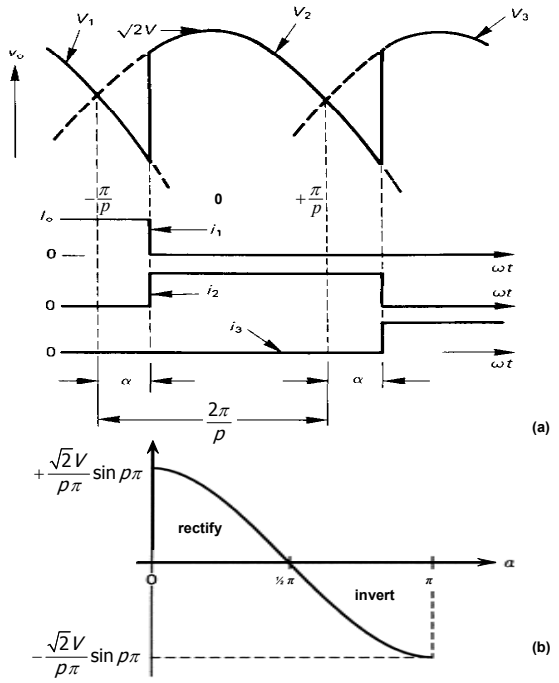


Figure 12.17. *A half-wave n-phase controlled converter: (a) output voltage and current waveform and (b) transfer function of voltage versus delay angle α.*

12.5 Overlap

In the previous sections of this chapter, impedance of the ac source has been neglected, such that current transfers or *commutates* instantly from one switch to the other with higher anode potential, when triggered. However, in practice the source has inductive reactance X_c and current takes a finite time to fall in the device turning off and rise in the device turning on.

Consider the three-phase half-wave controlled rectifying converter in figure 12.11a, where it is assumed that a continuous dc load current, I_o, flows. When thyristor T_1 is conducting and T_2 (which is forward biased) is turned on after delay α, the equivalent circuit is shown in figure 12.18a. The source reactances X_1 and X_2 limit the rate of change of current in T_1 as i_1 decreases from I_o to 0 and in T_2 as i_2 increases from 0 to I_o. These current transitions in T_1 and T_2 are shown in the waveforms of figure 12.18d. A circulating current, i, flows between the two thyristors. If the line reactances are identical, for a constant output current, the inductor di/dt currents are equal and opposite, the output voltage during commutation, v_γ, is mid-way between the conducting phase voltages v_1 and v_2, as shown in figure 12.18b. That is $v_\gamma = \frac{1}{2}(v_1 + v_2)$, creating a series of notches in the output voltage waveform as shown in figure 12.18c. This interval during which both T_1 and T_2 conduct ($i \ne 0$) is termed the *overlap period* and is defined by the *overlap angle* γ. Ignoring thyristor voltage drops, the overlap angle is calculated as follows:

$$v_2 - v_1 = 2L\frac{di}{dt}$$

With reference $t = 0$ when T_2 is triggered

$$v_2 - v_1 = v_L = \sqrt{3}\,v_{phase} = \sqrt{3}\sqrt{2}\,V\sin(\omega t + \alpha)$$

where V is the line to neutral rms voltage.
Equating these two equations

$$2L\,di/dt = \sqrt{3}\sqrt{2}\,V\sin(\omega t + \alpha)$$

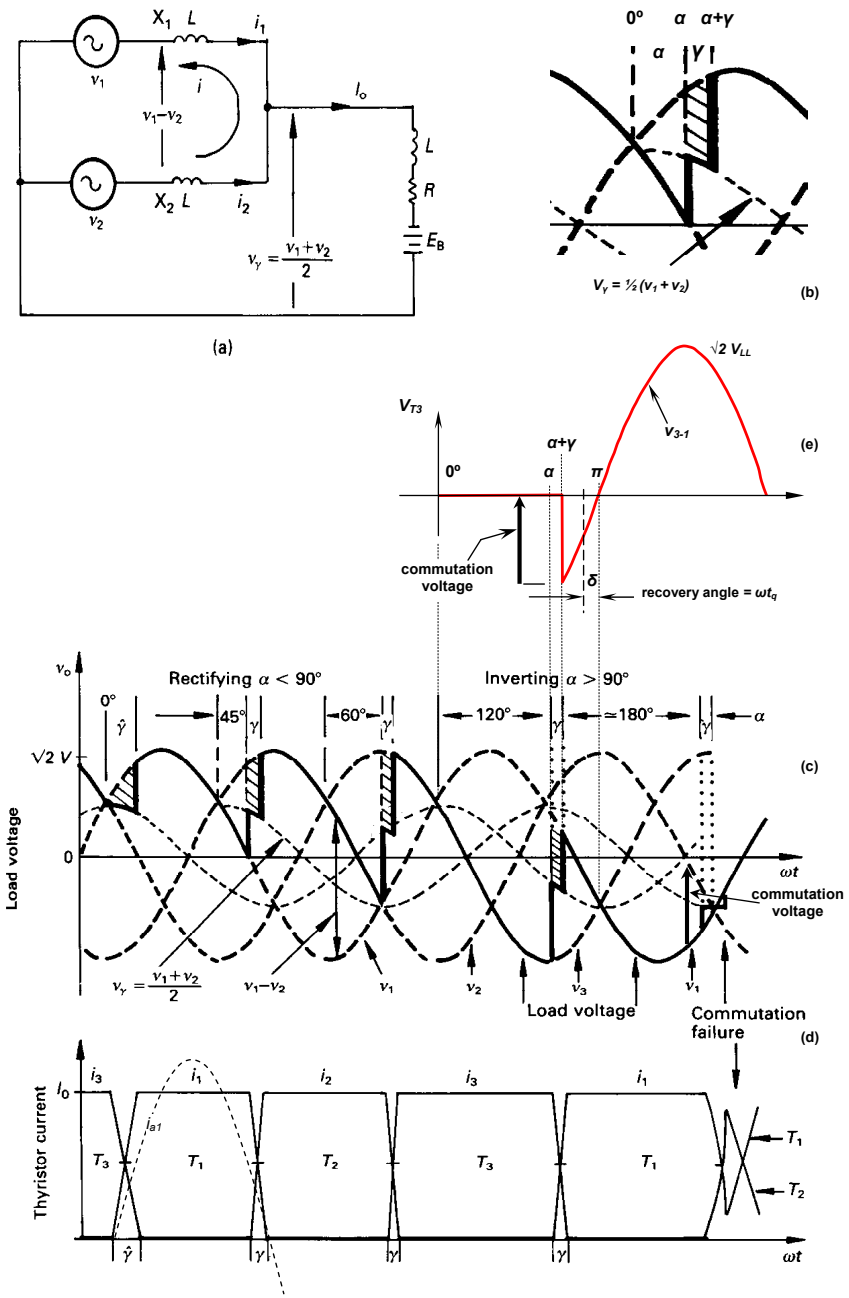


Figure 12.18. *Overlap: (a) equivalent circuit during overlap; (b) angle relationships; (c) load voltage for different delay angles α (hatched areas equal to I_oL; last overlap shows commutation failure); (d) thyristor currents showing eventual failure; and (e) voltage across a thyristor in the inversion mode, α >90°.*

Rearranging and integrating gives

$$i(\omega t) = \frac{\sqrt{3}\sqrt{2}V}{2\omega L}\left(\cos\alpha - \cos(\omega t + \alpha)\right) \qquad (12.185)$$

Commutation from T_1 to T_2 is complete when $i = I_o$, at $\omega t = \gamma$, that is

$$I_o = \frac{\sqrt{3}\sqrt{2}V}{2\omega L}\left(\cos\alpha - \cos(\gamma + \alpha)\right) = \frac{2\pi V_o}{3\omega L}\left(\cos\alpha - \cos(\gamma + \alpha)\right) \qquad \text{(A)} \qquad (12.186)$$

This equation holds for $\gamma \le ⅓\pi$, provided

$$I_o \le \frac{\sqrt{3}V}{\sqrt{2}\,\omega L}\cos(\alpha - ⅓\pi) \qquad \left[I_o \le \frac{2\pi V_o}{3\omega L}\cos(\alpha - ⅓\pi)\right]$$

Figure 12.18b shows that the load voltage comprises the phase voltage v_2 when no source inductance exists minus the voltage due to circulating current v_γ (= ½(v_1 + v_2)) during commutation.

The mean output voltage V_o^γ is therefore

$$V_o^\gamma = V_o - \bar{v}_\gamma$$
$$= \frac{1}{2\pi/3}\left[\int_{\alpha+\pi/6}^{\alpha+5\pi/6} v_2\, d\omega t - \int_{\alpha+\pi/6}^{\gamma+\alpha+\pi/6} v_\gamma\, d\omega t\right]$$

where v_γ = ½(v_1 + v_2)

$$V_o^\gamma = \frac{3}{2\pi}\left[\int_{\alpha+\pi/6}^{\alpha+5\pi/6}\sqrt{2}\,V\sin(\omega t + \alpha)\,d\omega t - \int_{\alpha+\pi/6}^{\gamma+\alpha+\pi/6}\sqrt{2}\,V\left\{\sin\left(\omega t + 2\pi/3\right) + \sin\omega t\right\}d\omega t\right]$$

$v_2 - v_1 = 2L\,di/dt$

$$V_o^\gamma = \frac{3}{2\pi}\sqrt{3}\sqrt{2}V\cos\alpha - \frac{3}{2\pi}\frac{\sqrt{3}}{2}\sqrt{2}V\left(\cos\alpha - \cos(\alpha - \gamma)\right) \qquad (12.187)$$

$$V_o^\gamma = \frac{3\sqrt{3}}{4\pi}\sqrt{2}V\left[\cos\alpha + \cos(\alpha + \gamma)\right] = ½V_o\left[\cos\alpha + \cos(\alpha + \gamma)\right] \qquad (12.188)$$

which reduces to equation (12.131) when $\gamma = 0$. Substituting $\cos\alpha - \cos(\alpha + \gamma)$ from equation (12.186) into equation (12.187) yields

$$V_o^\gamma = \frac{3\sqrt{3}}{2\pi}\sqrt{2}\,V\cos\alpha - \frac{3}{2\pi}\omega L I_o \quad = V_o - \frac{3}{2\pi}\omega L I_o \quad \text{where } V_o = \frac{3\sqrt{3}}{2\pi}\sqrt{2}\,V \qquad (12.189)$$

From equation (12.186) the commutation angle γ is

$$\gamma = \cos^{-1}\left(\cos\alpha - \frac{2\omega L}{\sqrt{3}\sqrt{2}V}I_o\right) - \alpha = \cos^{-1}\left(\cos\alpha - \frac{3\omega L}{2\pi V_o}I_o\right) - \alpha \qquad (12.190)$$

The displacement power factor angle is displaced by half the overlap angle, that is

$$DPF = \cos\phi_1 = \cos(\alpha + ½\gamma) \qquad (12.191)$$

while the overall power factor is

$$pf = DPF \times DF = \frac{\cos(\alpha + ½\gamma)}{\sqrt{1 + THD^2}} \qquad (12.192)$$

The mean output voltage V_o is reduced or regulated by the commutation reactance $X_c = \omega L$ and this regulation varies with load current magnitude I_o. Converter semiconductor voltage drops also regulate (decrease) the output voltage.
The component $3\omega L/2\pi$ is called the *equivalent internal resistance*. Being an inductive phenomenon, it does not represent a power loss component.
As shown in figure 12.19, the overlap occurs immediately after the delay α. The commutation voltage, $v_2 - v_1$, is $\sqrt{3}\,\sqrt{2}\,V \sin\alpha$. The commutation time is inversely proportional to the commutation voltage $v_2 - v_1$.

For rectification, as α increases from zero to ½π, the commutation voltage increases from a minimum of zero volts to a maximum of $\sqrt{3}\,\sqrt{2}\,V$ at ½π, whence the overlap angle γ decreases from a maximum of $\hat{\gamma}$ at $\alpha = 0$ to a minimum of $\breve{\gamma}$ at ½π.

[For inversion, the overlap angle γ decreases from a minimum of $\breve{\gamma}$ at ½π to a maximum of $\hat{\gamma}$ at π, as the commutation voltage reduces from a maximum, back to zero volts.]

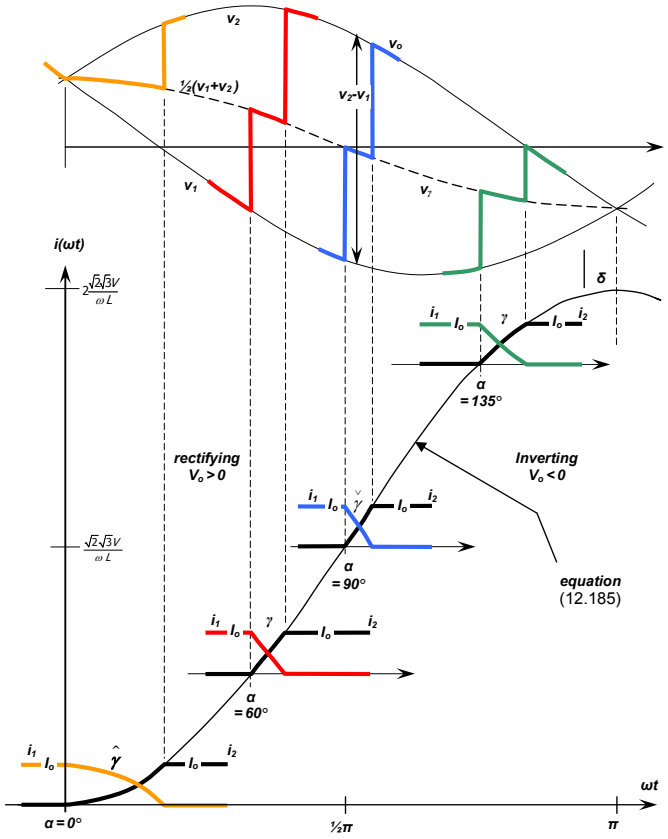


Figure 12.19. Overlap γ *for current commutation from thyristor 1 to thyristor 2, at delay angle α.*

From equation (12.186), with $\alpha = \pi$

$$\check{\gamma} = arc\ \sin(2\omega L I_o / \sqrt{2}\sqrt{3}\ V)$$

The general expressions for the mean load voltage V_o^{γ} of an *n*-pulse, fully-controlled converter, with underlap, are given by

$$V_o^{\gamma} = \frac{\sqrt{2}V}{2\pi/n} \sin{}^{\pi}\!/_{n} \left[\cos\alpha + \cos(\alpha+\gamma)\right] \qquad (12.193)$$

and

$$V_o^{\gamma} = \frac{\sqrt{2}V}{\pi/n} \sin{}^{\pi}\!/_{n} \cos\alpha \ \mp\ nX_c I_o / 2\pi \qquad (12.194)$$

where *V* is the line voltage for a full-wave converter and the phase voltage for a half-wave converter and the plus sign in equation (12.194) accounts for inversion operation.

Effectively, as shown in figure 12.20, during rectification, overlap reduces the mean output voltage by $nfLI_o$ or as if α were increased. The supply voltage is effectively distorted and the harmonic content of the output is increased. Equating equations (12.193) and (12.194) gives the mean output current

$$I_o = \frac{\sqrt{2}V}{X_c} \sin{}^{\pi}\!/_{n} \left(\cos\alpha - \cos(\gamma+\alpha)\right) \qquad \text{(A)} \qquad (12.195)$$

which reduces to equation (12.186) when $n = 3$.
Harmonic input current magnitudes are decreased by a factor $\sin(½n\gamma)/½n\gamma$.

In the three-phase case, for a constant dc link current I_o, without commutation effects, the rms phase current and the magnitude of the n[th] current harmonic are

$$I_{rms} = \frac{\sqrt{2}I_o}{\sqrt{3}} \qquad I_{hn} = I_o \frac{2\sqrt{3}}{n\pi} \qquad (12.196)$$

When accounting for commutation reactance effects the fundamental current is

$$I_{h1} = I_o \frac{2\sqrt{3}}{\pi} \frac{\left[\left[\cos 2\alpha - \cos 2(\alpha+\gamma)\right]^2 + \left[2\gamma + \sin 2\alpha - \sin 2(\alpha+\gamma)\right]^2\right]^{-½}}{4\left[\cos\alpha - \cos(\alpha+\gamma)\right]} \qquad (12.197)$$

The single-phase, full-wave, converter voltage drop is $2\omega LI_o/\pi$ and the overlap output voltage is zero.

The general effects of line inductance, which causes current overlap are:

- the average output voltage is reduced
- the input voltage is distorted – notching in the ac voltage
- the inversion safety angle to allow for thyristor commutation, is increased
- the output voltage spectrum component frequencies are unchanged but there magnitudes are decreased slightly
- thyristor *di/dt* is reduced.

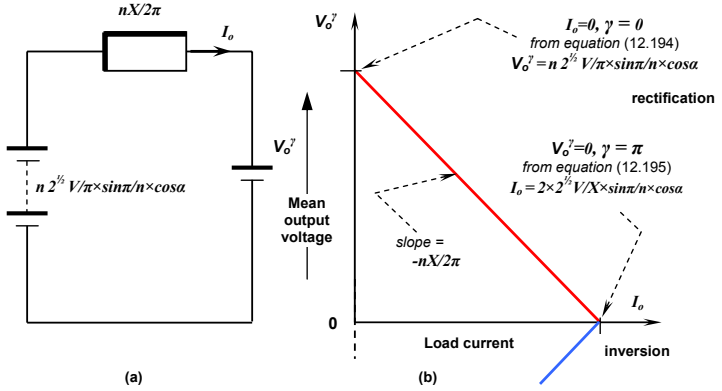


Figure 12.20. *Overlap regulation model: (a) equivalent circuit and (b) load plot of overlap model.*

Table 12.2. Summary of overlap effects on rectifier circuits

configuration	single-phase full-wave	single-phase bridge	three-phase half-wave	three-phase bridge	*m*-pulse rectifier
ΔV_c	$\frac{X_c}{\pi} I_o$	$\frac{2X_c}{\pi} I_o$	$\frac{3X_c}{2\pi} I_o$	$\frac{3X_c}{\pi} I_o$	$\frac{nX_c}{2\pi} I_o$
$\cos\alpha - \cos(\gamma+\alpha)$	$\frac{I_o X_c}{\sqrt{2}V}$	$\frac{2I_o X_c}{\sqrt{2}V}$	$\frac{2I_o X_c}{\sqrt{6}V}$	$\frac{2I_o X_c}{\sqrt{6}V}$	$\frac{I_o X_c}{\sqrt{2}V \sin\frac{\pi}{n}}$

12.6 Overlap - inversion

A fully controlled converter operates in the inversion mode when $\alpha > 90°$ and the mean output voltage is negative and less than the load back emf shown in figure 12.18a. Since the direction of the load current I_o is from the supply and the output voltage is negative, energy is being returned, *regenerated* into the supply from the load. Figure 12.21 shows the power flow differences between rectification and inversion. As α increases, the returned energy magnitude increases. If α plus the necessary overlap γ exceeds $\omega t = \pi$, commutation failure occurs. The output goes positive and the load current builds up uncontrolled.

The last commutation with $\alpha \approx \pi$ in figures 12.18c and d results in a commutation failure of thyristor T_1. Before the circulating inductor current i has reduced to zero, the incoming thyristor T_2 experiences an anode potential which is less positive than that of the thyristor to be commutated T_1, $v_1 - v_2 < 0$. The incoming device T_2 fails to stay on and conduction continues through T_1, impressing positive supply cycles across the load. This positive converter voltage aids the load back emf and the load current builds up uncontrolled.

Equations (12.193) and (12.194) are valid provided a commutation failure does not occur. The controllable delay angle range is curtailed to

$$0 \le \alpha \le \pi - \gamma$$

The maximum allowable delay angle $\hat{\alpha}$ occurs when $\hat{\alpha} + \gamma = \pi$ and from equations (12.193) and (12.194) with $\alpha + \gamma = \pi$ gives

$$\hat{\alpha} = \cos^{-1}\left\{\frac{XI_o}{\sqrt{2}V\sin\pi/n} - 1\right\} < \pi \qquad \text{(rad)} \qquad (12.198)$$

In practice commutation must be complete δ rad before $\omega t = \pi$, in order to allow the outgoing thyristor to regain a forward blocking state. That is $\alpha + \gamma + \delta \le \pi$. δ is known as the *recovery* or *extinction angle*, and is shown in figure 12.23e. The thyristor recovery period increases with increased anode current and temperature, and decreases with increased voltage.

The input power is equal to the dc power

$$P = \sqrt{3}VI\cos\phi = V_o I_o \qquad (12.199)$$

The input power factor is therefore

$$\cos\phi = \frac{V_o I_o}{\sqrt{3}VI} \approx \tfrac{1}{2}\left[\cos\alpha + \cos(\alpha+\gamma)\right] \qquad (12.200)$$

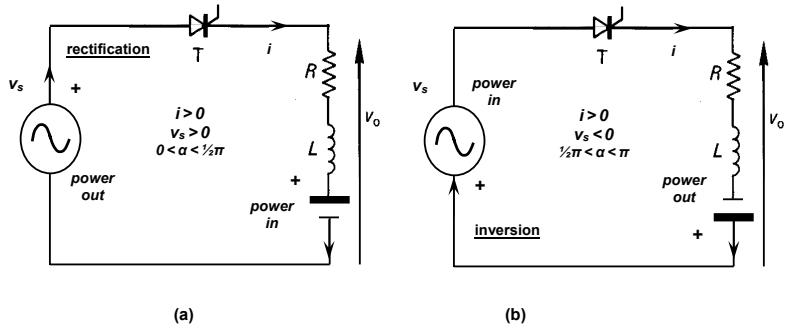


Figure 12.21. *Controlled converter model showing: (a) rectification and (b) inversion.*

Example 12.8: ***Converter overlap***

A three-phase full-wave converter is supplied from the 415 V ac, 50 Hz mains with phase source inductance of 0.1 mH. If the average load current is 100 A continuous, for phase delay angles of (i) 0° and (ii) 60° determine

i. the supply reactance voltage drop,
ii. mean output voltage (with and without commutation overlap), load resistance, and output power, and
iii. the overlap angle

Ignoring thyristor forward blocking recovery time requirements, determine the maximum allowable delay angle.

Solution

Using equations (12.193) and (12.194) with n = 6 and V = 415 V ac, the mean supply reactance voltage

$$\bar{v}_\gamma = \frac{n}{2\pi}2\pi f LI_o = \frac{6}{2\pi}\times 2\pi 50\times 10^{-4}\times 10^2$$
$$= 3\text{V}$$

(i) α = 0° - as for uncontrolled rectifiers. From equation (12.194), the maximum output voltage is

$$V_o^\gamma = \frac{\sqrt{2}V}{2\pi/n}\sin\%\cos\alpha - nX_c I_o/2\pi$$
$$= \frac{\sqrt{2}\times 415}{2\pi/6}\sin\%\times\cos 0 - 3\text{V} = 560.44\text{V} - 3\text{V} = 557.44\text{V}$$

where the mean output voltage without commutation inductance effects is 560.4V.
The power output for 100A is 560.4V×100A = 56.04kW and the load resistance is 560.4V/100A = 5.6Ω.
From equation (12.193)

$$V_o^\gamma = \frac{\sqrt{2}V}{2\pi/n}\sin\pi/n\left[\cos\alpha + \cos(\alpha+\gamma)\right]$$
$$557.44 = \frac{\sqrt{2}\times 415}{2\pi/6}\times\sin\pi/6\times\left[1+\cos\gamma\right]$$

that is $\gamma = 8.4°$

(ii) α = 60°

$$V_o^\gamma = \frac{\sqrt{2}V}{2\pi/n}\sin\%\cos\alpha - nX_c I_o/2\pi$$
$$= \frac{\sqrt{2}\times 415}{2\pi/6}\sin\%\times\cos 60° - 3\text{V} = 280.22\text{V} - 3\text{V} = 277.22\text{V}$$

where the mean output voltage without commutation inductance effects is 280.2V.
The power output for 100A is 280.2V×100A = 28.02kW and the load resistance is 280.2V/100A = 2.8Ω.

$$V_o^\gamma = \frac{\sqrt{2}V}{2\pi/n}\sin\pi/n\left[\cos\alpha + \cos(\alpha+\gamma)\right]$$
$$277.22 = \frac{\sqrt{2}\times 415}{2\pi/6}\times\tfrac{1}{2}\times\left[\cos 60° + \cos(60°+\gamma)\right]$$

that is $\gamma = 0.71°$

Equation (12.198) gives the maximum allowable delay angle as

$$\hat{\alpha} = \cos^{-1}\left\{\frac{XI_o}{\sqrt{2}V\sin\pi/n} - 1\right\}$$
$$= \cos^{-1}\left\{\frac{2\pi 50\times 10^{-4}\times 10^2}{\sqrt{2}\times 415\times\tfrac{1}{2}} - 1\right\}$$
$$= 171.56° \text{ and } V_o^\gamma = -557.41\text{V}$$

♣

12.7 Summary

General expressions for n-phase converter mean output voltage, V_o

(i) Half-wave and full-wave, fully-controlled converter

$$V_o = \sqrt{2}V\frac{\sin(\pi/n)}{\pi/n}\cos\alpha$$

where V is
the rms line voltage for a full-wave converter or
the rms phase voltage for a half-wave converter.
$\cos\alpha = \cos\psi$, the supply displacement factor
From L'Hopital's rule, for $n\to\infty$, $V_o = \sqrt{2}V\cos\alpha$

(ii) Full-wave, half-controlled converter

$$V_o = \sqrt{2}\,V\,\frac{\sin(\pi/n)}{\pi/n}\,(1+\cos\alpha)$$

where V is the rms line voltage.

(iii) Half-wave and full-wave controlled converter with load freewheel diode

$$V_o = \sqrt{2}\,V\,\frac{\sin(\pi/n)}{\pi/n}\cos\alpha \qquad 0 < \alpha < \tfrac{1}{2}\pi - \pi/n$$

$$V_o = \sqrt{2}\,V\,\frac{1+\cos\left(\alpha+\tfrac{1}{2}\pi-\pi/n\right)}{2\pi/n} \qquad \tfrac{1}{2}\pi-\pi/n < \alpha < \tfrac{1}{2}\pi+\pi/n$$

the output rms voltage is given by

$$V_{rms} = V\sqrt{1+\frac{\cos 2\alpha \sin 2\pi/n}{2\pi/n}} \qquad \alpha+\pi/n \le \tfrac{1}{2}\pi$$

$$V_{rms} = V\sqrt{\tfrac{1}{2}+\frac{n}{4}-\frac{\alpha}{2\pi/n}-\frac{\cos\left(2\alpha-2\pi/n\right)}{4\pi/n}} \qquad \alpha+\pi/n > \tfrac{1}{2}\pi$$

where V is
the rms line voltage for a full-wave converter or
the rms phase voltage for a half-wave converter.
n = 0 for single-phase and three-phase half-controlled converters
= $\frac{1}{6}$π for three-phase half-wave converters
= ⅓π for three-phase fully controlled converters

These voltage output characteristics are shown in figure 12.22 and the main converter circuit characteristics are shown in table 12.2.

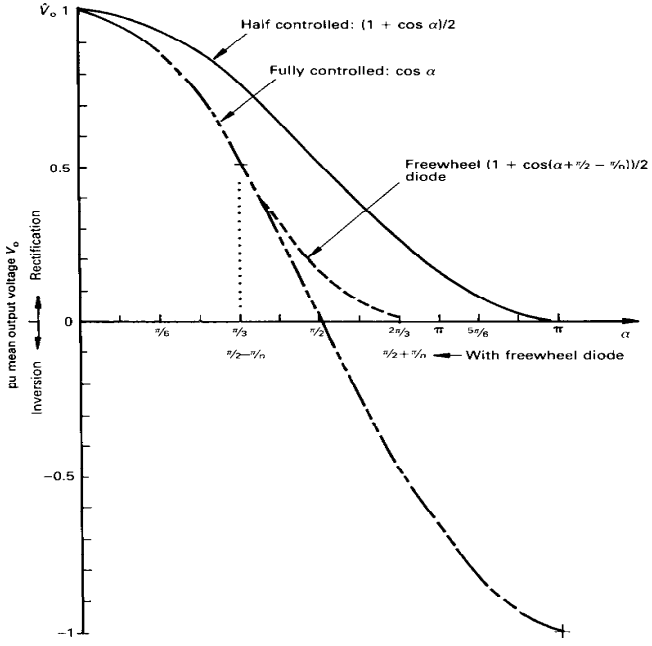


Figure 12.22. *Converter normalised output voltage characteristics as a function of firing delay angle α.*

12.8 Definitions (see Chapter 11.6)

V_o	average output voltage	$\bar{I}_o$	average output current
V_{rms}	rms output voltage	I_{rms}	rms output current
$\hat{V}$	peak output voltage	$\hat{I}$	peak output current

Load voltage form factor $= FF_v = V_{rms}/V_o$ Load voltage crest factor $= CF_v = \hat{V}/V_{rms}$

Load current form factor $= FF_i = I_{rms}/\bar{I}_o$ Load current crest factor $= CF_i = \hat{I}/I_{rms}$

$$\text{Rectification efficiency} = \eta = \frac{\text{dc load power}}{\text{ac load power + rectifier losses}} = \frac{V_o I_o}{V_{rms} I_{rms} + \text{Loss}_{\text{rectifier}}}$$

$$\text{Waveform smoothness = Ripple factor} = RF_v = \frac{\text{effective values of ac } V \text{ (or } I\text{)}}{\text{average value of } V \text{ (or } I\text{)}} = \frac{V_{Ri}}{V_o} = \sqrt{\frac{V_{rms}^2 - V_o^2}{V_o^2}} = \sqrt{FF_v^2 - 1}$$

where $V_{Ri} = \left[\sum_{n=1}^{\infty} \tfrac{1}{2}\left(v_{an}^2 + v_{bn}^2\right)\right]^{1/2}$

similarly the current ripple factor is $RF_i = \frac{I_{Ri}}{\bar{I}_o} = \sqrt{FF_i^2 - 1}$

$RF_i = RF_v$ for a resistive load

12.9 Output pulse number

Output pulse number p is the number of pulses in the output voltage that occur during one ac input cycle, of frequency f_s. The pulse number p therefore specifies the output harmonics, which occur at p x f_s, and multiples of that frequency, $m \times p \times f_s$, for m = 1, 2, 3, ...

$$p = \frac{\text{period of input supply voltage}}{\text{period of minimum order harmonic in the output } V \text{ or } I \text{ waveform}}$$

The pulse number p is specified in terms of

q	the number of elements in the commutation group
r	the number of parallel connected commutation groups
s	the number of series connected (phase displaced) commutating groups

Parallel connected commutation groups, r, are usually associated with (and identified by) intergroup reactors (to reduce circulating current), with transformers where at least one secondary is effectively star connected while another is delta connected. The rectified output voltages associated with each transformer secondary, are connected in parallel.

Series connected commutation groups, s, are usually associated with (and identified by) transformers where at least one secondary is effectively star while another is delta connected, with the rectified output associated with each transformer secondary, connected in series.

q=3 r=2 s=2
p=qxrxs
p=12

The mean converter output voltage V_o can be specified by

$$V_o = s\frac{q}{\pi}\sqrt{2}V_\phi \times \sin\frac{\pi}{q} \times \cos\alpha \qquad (12.201)$$

For a full-wave fully-controlled single-phase converter, r = 1, q = 2, and s = 1, whence p = 2

$$V_o = 1 \times \frac{2}{\pi}\sqrt{2}V_\phi \times \sin\frac{\pi}{2} \times \cos\alpha = \frac{2\sqrt{2}V_\phi}{\pi} \times \cos\alpha$$

For a full-wave, fully-controlled, three-phase converter, r = 1, q = 3, and s = 2, whence p = 6

$$V_o = 2 \times \frac{3}{\pi}\sqrt{2}V_\phi \times \sin\frac{\pi}{3} \times \cos\alpha = \frac{3\sqrt{2}V_\phi}{\pi} \times \cos\alpha$$

Table 12.3: Main characteristics of controllable converter circuits

Ouput phase number n and ripple frequency (xf_s):			*1*	*2*		*3*	*6*	
Type of controlled circuit:				*Single-phase bridge*			*Three-phase bridge*	
Text figure number:			*Single-phase half-wave* **12.2a**	*Two-phase half-wave* **12.7a**	*Fully controlled* **12.6b**	*Half controlled* **12.1**	*Three-phase half-wave* **12.11a**	*Fully controlled* **12.12a**	*Half controlled* **12.10a**
Mean output voltage	Maximum output voltage $\hat{V}_o$ $\alpha = 0$ or diode bridge V is rms phase voltage		$\sqrt{2}\,V/\pi$ (0.45 V)	$2\sqrt{2}\,V/\pi$	(0.9 V)		$3\sqrt{3}\,\sqrt{2}\,V/2\pi$ (1.17 V)	$3\sqrt{3}\sqrt{2}\,V/\pi$	(2.33 V)
	Normalised controlled mean output voltage $V_o/\hat{V}_o$	Pure resistive load or with freewheel diode D_f	$\frac{1+\cos\alpha}{2}$	$\frac{1+\cos\alpha}{2}$			$0 \leqslant \alpha \leqslant \pi/6$: $\cos\alpha$; $\pi/6 < \alpha \leqslant 5\pi/6$: $\frac{1+\cos(\alpha+\pi/6)}{\sqrt{3}}$	$0 \leqslant \alpha \leqslant \pi/3$: $\cos\alpha$; $\pi/3 < \alpha < 2\pi/3$: $1+\cos(\alpha+\pi/3)$	$\frac{1+\cos\alpha}{2}$
		Text figure no.	**12.6a**	**12.1**			**12.12**	**12.15**	**12.10a**
		Inductive load without D_f		$\cos\alpha$	$\frac{1+\cos\alpha}{2}$		$\cos\alpha$	$\cos\alpha$	$\frac{1+\cos\alpha}{2}$
Equivalent internal resistance $i\,nX/2\pi$ $X = \omega L$				$0.318X$	$0.637X$		$0.477X$	$0.955X$	
Output voltage ripple ratio (per cent) ($\alpha = 0$, $\gamma = 0$)			121	48			19	4.2	
Rectifying device	Average current I_o/n		I_o	$I_o/2$			$I_o/3$	$I_o/3$	
	Peak voltage, × V		$\sqrt{2}$	$2\sqrt{2}$	$\sqrt{2}$		$\sqrt{3}\sqrt{2}$	$\sqrt{3}\sqrt{2}$	I_s
Supply rms currents	Fundamental I_1				$2\sqrt{2}\,I_o/\pi$	$\frac{2\sqrt{2}\,I_o}{\pi}\cos\alpha/2$		$\sqrt{6}\,I_o/\pi$	$\alpha \leq \frac{1}{3}\pi$: $\sqrt{\frac{2}{3}}\,I_o$
	Total I_s				I_o	$I_o\sqrt{1-\alpha/\pi}$		$\sqrt{\frac{2}{3}}\,I_o$	$\alpha \geq \frac{1}{3}\pi$: $\sqrt{\frac{\pi-\alpha}{\pi}}\,I_o$
Supply factors	Harmonic factor ρ				$\sqrt{\frac{\pi^2}{8}-1}$	$\sqrt{\frac{\pi(\pi-\alpha)}{4(1+\cos\alpha)}-1}$		$\sqrt{\left(\frac{\pi}{3}\right)^2-1}$	λ $\alpha \leq \frac{1}{3}\pi$
	Displacement factor, $\cos\psi$				$\cos-\alpha$	$\cos-\alpha/2$		$\cos-\alpha$	$\frac{3}{2\pi}(1+\cos\alpha)$ $\alpha \geq \frac{1}{3}\pi$
	Power factor λ				$\frac{2\sqrt{2}}{\pi}\cos\alpha$	$\frac{\sqrt{2}(1+\cos\alpha)}{\sqrt{\pi(\pi-\alpha)}}$		$\frac{3}{\pi}\cos\alpha$	$\frac{\sqrt{3}\,(1+\cos\alpha)}{\sqrt{\pi-\alpha}}$

12.10 AC-dc converter generalised equations

Alternating sinusoidal voltages

$$V_1 = \sqrt{2}\,V\sin\omega t$$
$$V_2 = \sqrt{2}\,V\sin\left(\omega t - \tfrac{2\pi}{q}\right)$$
$$\cdot$$
$$V_q = \sqrt{2}\,V\sin\left(\omega t - (q-1)\tfrac{2\pi}{q}\right)$$

where q is the number of phases (number of voltage sources)

On the secondary or converter side of any transformer, if the load current is assumed constant I_o then the power factor is determined by the load voltage harmonics.
Voltage form factor

$$FF_v = \frac{V_{rms}}{V_o}$$

whence the voltage ripple factor is

$$RF_v = \frac{1}{V_o}\left[V_{rms}^2 - V_o^2\right]^{\frac{1}{2}} = \left[FF_v^2 - 1\right]^{\frac{1}{2}}$$

The power factor on the secondary side of any transformer is related to the voltage ripple factor by

$$pf = \frac{P_d}{S} = \frac{V_o I_o}{qVI_{rms}} = \frac{1}{\sqrt{RF_v^2+1}}$$

On the primary side of a transformer the power factor is related to the secondary power factor, but since the supply is assumed sinusoidal, the power factor is related to the primary current harmonics.

Relationship between current ripple factor and power factor

$$RF_i = \frac{1}{I_1}\sqrt{\sum_{h=3}^{\infty} I_h^2} = \frac{1}{I_1}\sqrt{I_{rms}^2 - I_1^2}$$
$$pf = \frac{I_1}{I_{rms}} = \frac{1}{\sqrt{1+RF_i^2}}$$

The supply power factor is related to the primary power factor and is dependent of the supply connection, star or delta, etc.

Half-wave controlled rectifiers – star connected secondary supply [see figures 12.4, 12.11]

q phases and q thyristors, and a phase delay angle of α. The pulse number is *p (=q)*.
Mean output voltage is

$$V_o = \frac{q}{2\pi}\int_{\frac{1}{2}\pi-\frac{\pi}{q}+\alpha}^{\frac{1}{2}\pi+\frac{\pi}{q}+\alpha}\sqrt{2}\,V\sin\omega t\;d\omega t$$
$$= \frac{q}{\pi}\sqrt{2}\,V\sin\frac{\pi}{q}\cos\alpha$$
$$= V_o(\alpha=0)\;\cos\alpha$$

The rms output voltage is

$$V_{rms} = \frac{q}{2\pi}\int_{\frac{1}{2}\pi-\frac{\pi}{q}+\alpha}^{\frac{1}{2}\pi+\frac{\pi}{q}+\alpha}\left(\sqrt{2}\,V\sin\omega t\right)^2\;d\omega t$$
$$= \sqrt{2}\,V\left[\tfrac{1}{2} + \frac{q}{4\pi}\sin\frac{2\pi}{q}\cos\alpha\right]^{\frac{1}{2}}$$

The maximum and minimum voltages in the output are

$$\hat{v} = \sqrt{2}\,V \qquad \text{for } 0 < \alpha < \frac{\pi}{q}$$
$$= \sqrt{2}\,V\cos\left(-\frac{\pi}{q}+\alpha\right) \qquad \text{for } \alpha > \frac{\pi}{q}$$
$$\check{v} = \sqrt{2}\,V\cos\left(\frac{\pi}{q}+\alpha\right) \qquad \text{for } \alpha < \frac{3\pi}{2}+\frac{\pi}{q}$$
$$= -\sqrt{2}\,V \qquad \text{for } \alpha > \frac{3\pi}{2}+\frac{\pi}{q}$$

Normalised peak to peak ripple voltage and amplitude of the output harmonics are

$$V_{n_{p-p}} = \frac{v_{p-p}}{V_o} = \frac{\pi}{q}\,\frac{1-\cos\frac{\pi}{q}}{\sin\frac{\pi}{q}}$$

$$V_h = \frac{q}{4\pi}\sqrt{2}V\sin\frac{2\pi}{q}\left|\cos\alpha\right|\frac{2}{k^2q^2-1}\sqrt{1+k^2q^2\tan^2\alpha}$$

Diode reverse voltage

$$\hat{V}_{D_R} = 2\sqrt{2}V \qquad q \text{ even}$$

$$\hat{V}_{D_R} = 2\sqrt{2}V\cos\frac{\pi}{2q} \qquad q \text{ odd}$$

The thyristor currents are the same as the equivalent diode circuit

Power factor (is related to the equivalent diode circuit)

$$pf = \frac{\sqrt{2q}}{\pi}\sin\frac{\pi}{q}\cos\alpha = pf_{\alpha=0}\cos\alpha$$

Overlap angle and inductive voltage

$$\cos\alpha - \cos(\alpha+\mu) = \frac{\omega L_c I_o}{\sqrt{2}V\sin\frac{\pi}{q}}$$

$$v_{com} = \frac{q}{2\pi}\omega L_c I_o$$

Time domain equations, for an *R-L* load

$$L\frac{di}{dt} + Ri = \sqrt{2}V\cos\omega t \qquad \text{(V)}$$

Continuous current

$$i(\omega t) = \frac{\sqrt{2}\,V}{Z}\cos(\omega t - \phi) + \{\, i_o - \frac{\sqrt{2}\,V}{Z}\cos(-\frac{\pi}{p}+\alpha - \phi)\, e^{(-\frac{\pi}{p}+\alpha - \omega t)/\tan\phi}\}$$

where $Z = \sqrt{R^2 + (\omega L)^2}$ (ohms)

$\tan\phi = \omega L / R$

where

$$i_o = \frac{\sqrt{2}\,V}{R\sec^2\phi}\times\frac{\cos(\frac{\pi}{p}+\alpha)+\tan\phi\sin(\frac{\pi}{p}+\alpha)-\left[\cos(\frac{\pi}{p}-\alpha)-\tan\phi\sin(\frac{\pi}{p}-\alpha)\right]e^{-\frac{2\pi}{p\tan\phi}}}{1-e^{-\frac{2\pi}{p\tan\phi}}}$$

with an average value of

$$I_o = \frac{V_o}{R} = \frac{p}{\pi}\frac{\sqrt{2}V}{R}\sin\frac{\pi}{p}\cos\alpha$$

Discontinuous current
Boundary condition

$$\tan\alpha = \frac{\tan\phi\tan\frac{\pi}{p}+\tanh\frac{\pi}{p\tan\phi}}{\tan\frac{\pi}{p}-\tan\phi\tanh\frac{\pi}{p\tan\phi}}$$

$$i(\omega t) = \frac{\sqrt{2}V}{R\sec^2\phi}\left\{\cos\omega t + \tan\phi\sin\omega t + \left[\cos\left(\frac{\pi}{p}-\alpha\right)-\tan\phi\sin\left(\frac{\pi}{p}-\alpha\right)\right]e^{(-\frac{\pi}{p}+\alpha - \omega t)/\tan\phi}\right\}$$

The average output voltage is dependent on the current extinction angle, β

$$V_o = \frac{p}{2\pi}\sqrt{2}V\left[\sin\beta - \sin\left(-\frac{\pi}{p}+\alpha\right)\right]$$

Half-wave controlled rectifiers, with freewheel diode [see figures 12.6, 12.12]

q phases *q* thyristors and 1 diode

$$v(\omega t) = \sqrt{2}V\cos\omega t \qquad \text{for } -\frac{\pi}{p}+\alpha < \omega t < \frac{\pi}{p}+\alpha$$

$$= 0 \qquad \text{for } \frac{\pi}{2} < \omega t < \frac{\pi}{p}+\alpha$$

where the earliest conduction point is

$$\alpha > ½\pi - \frac{\pi}{p}$$

Mean rectified output voltage is

$$V_o = \frac{p}{2\pi}\int_{-\frac{\pi}{q}+\alpha}^{½\pi}\sqrt{2}V\cos\omega t\; d\omega t$$

$$= \frac{p}{2\pi}\sqrt{2}V\left[1-\sin\left(-\frac{\pi}{p}+\alpha\right)\right]$$

RMS voltage

$$V_{rms} = \sqrt{2}V\left[¼ - \frac{p}{8} - \frac{p\alpha}{4\pi} + \frac{p}{8\pi}\sin\left(2\alpha - \frac{2\pi}{q}\right)\right]^{½}$$

The maximum ripple occurs at $\omega t = -\frac{\pi}{p}+\alpha$, with zero volts during diode freewheeling, thus

$$v_{p-p} = \sqrt{2}V\cos\left(\frac{\pi}{p}-\alpha\right) - 0$$

$$V_{n_{p-p}} = \frac{v_{p-p}}{\hat{V}_o} = \frac{\sqrt{2}V\cos\left(\frac{\pi}{p}-\alpha\right)}{\frac{p}{2\pi}\sqrt{2}V\left(1-\sin\left(\frac{\pi}{p}-\alpha\right)\right)} = \frac{2\pi}{p}\,\frac{\cos\left(\frac{\pi}{p}-\alpha\right)}{1+\sin\left(\frac{\pi}{p}-\alpha\right)}$$

The freewheel diodes conduct for p periods of duration $\pi/p+\alpha$, and the currents are

$$\bar{I}_{Df} = \bar{I}_o\left(½ + \frac{p\alpha}{2\pi} - ¼p\right)$$

$$I_{rmsDf} = \bar{I}_o\left(½ + \frac{p\alpha}{2\pi} - ¼p\right)^{½}$$

The thyristor conductor for $2\pi/p$ without a load freewheel diode and $2\pi/p-(\pi/p+\alpha+½\pi)$ when the diode is present. The thyristor rms current is

$$I_{rmsTh} = \frac{I_o}{\sqrt{q}}\left[½ - \frac{p\alpha}{2\pi} + ¼p\right]^{½}$$

Full-wave fully controlled thyristor converters–star connected supply [see figures 12.7, 12.8, 12.9, 12.13]

q phases 2*q* thyristors
Pulse number, *p*

p=q if *q* is even
p=2q if *q* is odd

Mean voltage

$$V_o = \frac{p}{\pi}\hat{V}\sin\frac{\pi}{p}\cos\alpha = V_o'\cos\alpha$$

The rms output voltage is

$$V_{o\,rms} = \hat{V}\left[½ + \frac{p}{4\pi}\sin\frac{2\pi}{p}\cos\alpha\right]^{½}$$

The maximum and minimum voltages in the output are

$$\hat{v} = \hat{V} \qquad \text{for } 0 < \alpha < \frac{\pi}{p}$$

$$= \hat{V}\cos\left(-\frac{\pi}{p} + \alpha\right) \qquad \text{for } \alpha > \frac{\pi}{p}$$

$$\overset{\vee}{v} = \hat{V}\cos\left(\frac{\pi}{p} + \alpha\right) \qquad \text{for } \alpha < \frac{3\pi}{2} + \frac{\pi}{p}$$

$$= -\hat{V} \qquad \text{for } \alpha > \frac{3\pi}{2} + \frac{\pi}{p}$$

where

$$\hat{V} = \frac{\pi}{p}\frac{1}{\sin\frac{\pi}{p}} \times V_o'$$

$$V_o' = \frac{2q}{\pi}\sqrt{2}V\sin\frac{\pi}{q}$$

Thyristor maximum reverse and forward voltage

$$\hat{V}_R = 2\sqrt{2}V \qquad q \text{ even}$$

$$\hat{V}_R = 2\sqrt{2}V\cos\frac{\pi}{2q} \qquad q \text{ odd}$$

RMS currents and power factor

$$I_{TH} = \frac{I_o}{q} \qquad I_{TH\,rms} = I_o\sqrt{\frac{2}{q}}$$

$$pf = \frac{q}{\pi}\sin\frac{\pi}{q}\cos\alpha = pf_{\alpha=0}\cos\alpha$$

Overlap angle and inductive voltage

$$\cos\alpha - \cos(\alpha - \mu) = \frac{\omega L_c I_o}{\sqrt{2}V\sin\frac{\pi}{q}}$$

$$v_{com} = \frac{q}{\pi}\omega L_c I_o$$

Half-controlled full bridges – star connected secondary [see figures 12.1, 12.10, 12.16]

q phases *q* thyristors and *q* diodes, each of which conduct for 2π/*q*
Pulse number

p=*q* for all *q*, odd or even

Mean voltage

$$V_o = \frac{q}{\pi}\sqrt{2}V\sin\frac{\pi}{q}(1+\cos\alpha)$$

$$= V_o'\frac{1+\cos\alpha}{2}$$

where $V_o' = \frac{2q}{\pi}\sqrt{2}V\sin\frac{\pi}{q}$

Thyristor and diode reverse voltage

$$\hat{V}_R = 2\sqrt{2}V \qquad q \text{ even}$$

$$\hat{V}_R = 2\sqrt{2}V\cos\frac{\pi}{2q} \qquad q \text{ odd}$$

Power factor

$$I_s = I_o\sqrt{\frac{2}{q}} \qquad pf = \frac{2}{\pi}\sqrt{q}\sin\frac{\pi}{q}\frac{1+\cos\alpha}{2} \qquad \alpha < \pi - \frac{2\pi}{q}$$

$$I_s = I_o\left(1-\frac{\alpha}{\pi}\right)^{1/2} \qquad pf = \frac{\sqrt{2}}{\pi}\sin\frac{\pi}{q}(1+\cos\alpha)\left(\frac{\pi}{\pi-\alpha}\right)^{1/2} \qquad \alpha > \pi - \frac{2\pi}{q}$$

Full-wave fully-controlled bridges – delta connected secondary supply

Pulse number in the rectified output is

p=*q* for *q* even
p=2*q* for *q* odd

Thyristor maximum forward and reverse voltages

$$\hat{V}_R = \frac{2\sqrt{2}V}{\frac{q}{\pi}\sin\frac{\pi}{q}} \qquad q \text{ even}$$

$$\hat{V}_R = \frac{2\sqrt{2}V\cos\frac{\pi}{2q}}{\frac{q}{\pi}\sin\frac{\pi}{q}} \qquad q \text{ odd}$$

The power factor is the same as for the star case.

Half-controlled full bridges – delta connected secondary supply

In terms of the semiconductors and rectified voltage star and mesh behave the same.
Mean voltage, for all *q* is

$$V_o = \frac{q}{\pi}\sqrt{2}V\sin\frac{\pi}{q}(1+\cos\alpha)$$

$$= V_o'\frac{1+\cos\alpha}{2}$$

where $V_o' = \frac{2q}{\pi}\sqrt{2}V\sin\frac{\pi}{q}$

For a 3-phase half-controlled converter, the secondary current is

$$I_s = \frac{I_o}{\sqrt{3}}\left(1-\frac{\alpha}{\pi}\right)^{1/2}$$

For large *q*, *q*>6

$$I_s = \frac{I_o}{\sqrt{3}}\left(1-\left(\frac{\alpha}{\pi}\right)^2\right)^{1/2}$$

$$pf = \frac{\sqrt{2}}{\pi}\frac{1+\cos\alpha}{\pi\left(1-\left(\frac{\alpha}{\pi}\right)^2\right)^{1/2}}$$

Reading list

Dewan, S. B. and Straughen, A., *Power Semiconductor Circuits,* John Wiley and Sons, New York, 1975.

Sen, P.C., *Power Electronics*, McGraw-Hill, 5th reprint, 1992.

Shepherd, W *et al*. *Power Electronics and motor control*, Cambridge University Press, 2nd Edition 1995.

http://www.ipes.ethz.ch/

Problems

12.1. For the circuit shown in figure 12.23, if the thyristor is fired at $\alpha = \frac{1}{3}\pi$
- i. derive an expression for the load current, *i*
- ii. determine the current extinction angle, β
- iii. determine the peak value and the time at which it occurs
- iv. sketch to scale on the same ωt axis the supply voltage, load voltage, thyristor voltage, and load current.

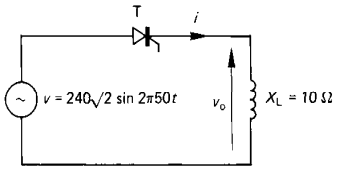


Figure 12.23. *Problem 12.1.*

12.2. For the circuit shown in figure 12.24, if the thyristor is fired at $\alpha = \frac{1}{4}\pi$ determine
- i. the current extinction angle, β
- ii. the mean and rms values of the output current
- iii. the power delivered to the source *E*.
- iv. sketch the load current and load voltage v_o.

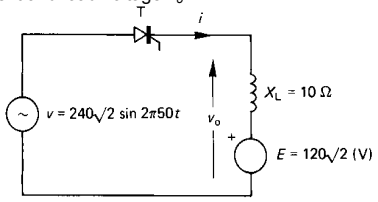


Figure 12.24. *Problem 12.2.*

12.3. Assuming a constant load current derive an expression for the mean and rms device current and the device form factor, for the circuit in figure 12.1.

12.4. Plot load ripple voltage K_{RI} and load voltage ripple factor RF_v, against the thyristor phase delay angle α for the circuit in figure 12.1.

12.5. Show that the average output voltage of a *n*-phase half-wave controlled converter with a freewheel diode is characterised by

$$V_o = \sqrt{2}V\frac{\sin(\pi/n)}{\pi/n}\cos\alpha \qquad \text{(V)}$$

$$0 < \alpha < \tfrac{1}{2}\text{-}\pi/n$$

$$V_o = \sqrt{2}V\frac{1+\cos\alpha+\frac{1}{2}\pi-\frac{1}{n}\pi}{2\pi/n} \qquad \text{(V)}$$

$$\tfrac{1}{2}\pi-\tfrac{1}{n}\pi < \alpha < \tfrac{1}{2}\pi+\tfrac{1}{n}\pi$$

12.6. Draw the load voltage and current waveforms for the circuit in figure 12.6a when a freewheel diode is connected across the load. Specify the load rms voltage.

12.7. The converter in figure 12.6a, with a freewheel diode, is operated from the 240 V, 50 Hz supply. The load consists of, series connected, a 10 Ω resister, a 5 mH inductor and a 40 V battery. Derive the load voltage expression in the form of a Fourier series. Determine the rms value of the fundamental of the load current.

12.8. Show that the average output voltage of a single-phase fully controlled converter is given by

$$V_o = \frac{2\sqrt{2}V}{\pi}\cos\alpha$$

Assume that the output current I_o is constant.
Prove that the supply current Fourier coefficients are given by

$$a_n = -\frac{4I_o}{n\pi}\sin n\alpha \qquad b_n = \frac{4I_o}{n\pi}\cos n\alpha$$

for *n* odd.
Hence or otherwise determine (see section 12.6)
- i. the displacement factor, cos ψ
- ii. the distortion factor, μ
- iii. the total supply power factor λ.

Determine the supply harmonic factor, ρ, if

$$\rho = I_h / I_\ell$$

where I_h is the total harmonic current and I_ℓ is the fundamental current.

12.9. Show that the average output voltage of a single-phase half-controlled converter is given by

$$v_o = \frac{\sqrt{2}V}{\pi}(1+\cos\alpha)$$

Assume that the output current I_o is constant.

Determine
- i. the displacement factor, cos ψ
- ii. the distortion factor, μ
- iii. the total supply power factor, λ.

Show that the supply harmonic factor, ρ (see problem 12.8), is given by

$$\rho = \sqrt{\left[\frac{\pi(\pi-\alpha)}{4(1+\cos\alpha)}-1\right]}$$

12.10. A centre tapped transformer, single-phase, full-wave converter (figure 12.7a) with a load freewheel diode is supplied from the 240 V ac, 50 Hz supply with source inductance of 0.25 mH. The continuous load current is *5* A. Find the overlap angles for
- i. the transfer of current form a conducting thyristor to the load freewheel diode and
- ii. from the freewheel diode to a thyristor when the delay angle α is 30°.

$$\gamma_{t-d} = \cos^{-1}\left\{1-\frac{\omega LI_o}{\sqrt{2}V}\right\} = 2.76°;$$

$$\gamma_{t-d} = \cos^{-1}\left\{\cos\alpha-\frac{\omega LI_o}{\sqrt{2}V}\right\}-\alpha = 0.13°$$

12.11. The circuit in figure 12.4a, with $v = \sqrt{2}V\sin(\omega t+\alpha)$, has a steady-state time response of

$$i(\omega t) = \frac{\sqrt{2}V}{Z}\left\{\sin(\omega t+\alpha-\phi)-\sin(\alpha-\phi)e^{-Rt/L}\right\}$$

where α is the trigger phase delay angle after voltage crossover and $\phi = \tan^{-1}(\omega L/R)$

Sketch the current waveform for $\alpha = \frac{1}{4}\pi$ and *Z* with
- i. $R >> \omega L$
- ii. $R = \omega L$
- iii. $R << \omega L$.

[$(\sqrt{2}\,V/R)\sin(\omega t+\frac{1}{4}\pi)$; $(V/R)\sin\omega t$; $(V/\omega L)(\sin\omega t-\cos\omega t+1)$]

12.12. A three-phase, fully-controlled converter is connected to the 415 V supply, which has a reactance of 0.25 Ω/phase and resistance of 0.05 Ω/phase. The converter is operating in the inverter mode with α = 150° and a continuous 50 A load current. Assuming a thyristor voltage drop of 1.5 V at 50 A, determine the mean output voltage, overlap angle, and available recovery angle.

[-485.36 V -3 V -5 V -11.94 V = -505.3 V; 6.7°; 23.3°]

12.13. For the converter system in problem 12.12, what is the maximum dc current that can be accommodated at a phase delay of 165°, allowing for a recovery angle of 5°?

[35.53 A]

12.14. The single-phase half-wave controlled converter in figure 12.4 is operated from the 240 V, 50 Hz supply and a 10 Ω resistive load. If the mean load voltage is 50 per cent of the maximum mean voltage, determine the (a) delay angle, α, (b) mean and rms load current, and (c) the input power factor.

12.15. The converter in figure 12.1a is operated from the 240 V, 50 Hz supply with a load consisting of the series connection of a 10 Ω resistor, a 5 mH inductor, and a 40 V battery. Derive the load voltage expression in the form of a Fourier series. Determine the rms value of the fundamental of the load current.

12.16. The converter in figure 12.12 is operated from a Y-connected, 415 V, 50 Hz supply. If the load is 100 A continuous with a phase delay angle of π/6, calculate the (a) harmonic factor of the supply current, (b) displacement factor cos ψ, and (c) supply power factor, λ.

12.17. The converter in figure 12.12 is operated from the 415 V line-to-line voltage, 50 Hz supply, with a series load of 10 Ω + 5 mH + 40 V battery. Derive the load voltage expression in terms of a Fourier series. Determine the rms value of the fundamental of the load current.

12.18. Repeat problem 12.17 for the three-phase, half-controlled converter in figure 12.10.

12.19. Repeat problem 12.17 for the three-phase, fully-controlled converter in figure 12.13.

12.21. The three-phase, half-controlled converter in figure 12.10 is operated from the 415 V, 50 Hz supply, with a 100 A continuous load current. If the line inductance is 0.5 mH/phase, determine the overlap angle γ if (a) $\alpha = \pi/6$ and (b) $\alpha = \frac{2}{3}\pi$.

12.22. Repeat *example 12.1* using a 100Vac 60Hz supply.

12.23. A fully controlled half-wave rectifier has a resistive 30Ω and a 240V ac 50Hz voltage source.

(a) If the delay angle $\alpha = 60^o$, determine:
   i. the average voltage across the load resistor
   ii. the power absorbed by the load resistor
   iii. the ac source power factor.

(b) If the average load current is 5A determine
   i. the average voltage across the load
   ii. the power absorbed by the load
   iii. the supply power factor.

12.24. A fully controlled half-wave rectifier has a 240V ac 50Hz source and a series *R*-*L* load of R = 20Ω and L = 50mH. If the delay angle is $\alpha = 60^o$, determine

   i. an expression for the load current
   ii. the average load current
   iii. the power absorbed by the load
   iv. the supply power factor

12.25. An electromagnet is modelled by a series *R*-*L* circuit with R=10Ω and L=100mH, and supplied from a 50Hz 240V ac voltage source.

   i. when supplied from a full-wave uncontrolled rectifier, the average current must be 20A to active the magnetic field. What series resistance must be added to increase the average current to 20A?
   ii. when supplied from a full-wave fully-controlled converter, what delay angle will produce the necessary average current of 20A to activate the electro-magnet?

12.26. Show that the power factor for a fully-controlled single-phase full-wave converter with a purely resistive load is given by

$$pf = \sqrt{1 - \frac{\alpha}{\pi} + \frac{\sin 2\alpha}{2\pi}}$$

12.27. A fully controlled single-phase full-wave bridge converter has a 60 Ω resistive load and a 240V ac 50Hz voltage source. If the firing delay angle is $\alpha = 60^o$, determine:
   i. the average load current
   ii. the rms load current
   iii. the rms source current
   iv. the ac source power factor

12.28. A three-phase, fully-controlled, converter is supplied from a 3.3kV ac 50Hz source. If the load is a 110Ω resistor determine:
   i. the delay angle which results is an average load current of 20A
   ii. the amplitude of the first voltage harmonic (at 300Hz)

12.29. A three-phase, fully-controlled, converter is supplied from a 3.3kV ac 50Hz source. If the *R*-*L* load is R = 100Ω and L = 100mH, and the delay angle is $\alpha = 30^o$, determine:
   i. the average load current
   ii. the amplitude of the first current harmonic (at 300Hz)
   iii. the rms phase current from the ac voltage source.

13

AC Voltage Regulators

AC voltage regulators have a constant voltage ac supply input and incorporate semiconductor switches which vary the rms voltage impressed across the ac load. These regulators generally fall into the category of naturally commutating converters since their thyristor switches are naturally commutated by the alternating supply. This converter turn-off process is termed *line commutation*.
The regulator output current, hence supply current, may be discontinuous or non-sinusoidal and as a consequence input power factor correction and harmonic reduction are usually necessary, particularly at low output voltage levels (relative to the input ac voltage magnitude).
A feature of *direction conversion* of ac to ac is the absence of any intermediate energy stage, such as a capacitive dc link or energy storage inductor. Therefore ac to ac converters are potentially more efficient but usually involve a larger number of switching devices and output is lost if the input supply is temporarily lost.
There are three basic ac regulator categories, depending on the relationship between the input supply frequency f_s, which is usually assumed single frequency sinusoidal, possibly multi-phased, and the output frequency f_o. Without the use of transformers (or boost inductors), the output voltage rms magnitude V_{Orms} is less than or equal to the input voltage rms magnitude V_s, $V_{o\,rms} \le V_s$.

- output frequency increased, $f_o > f_s$, for example, the matrix converter
- output frequency decreased, $f_o < f_s$, for example, the cycloconverter
- output frequency fundamental = supply frequency, $f_o = f_s$, for example, a phase controller

13.1 Single-phase ac regulator

Figure 13.1a shows a single-phase thyristor ac regulator supplying an *L-R* load. The two inverse parallel connected thyristors, possibly in the form of an ac output solid-state relay, SSR, can be replaced by any of the bidirectional conducting and blocking switch arrangements shown in figure 13.1c or figure 6.11. Equally, in low power applications the two thyristors are usually replaced by a triac.
The ac regulator in figure 13.1a can be controlled by two methods

- phase angle control – using symmetrical delay angles
- integral (or half integral) cycle control – using zero phase angle delay

13.1.1 Single-phase ac regulator – phase control with line commutation

For control by phase angle delay, the thyristor gate trigger delay angle is α, where $0 \le \alpha \le \pi$, as indicated in figure 13.1b. The fundamental of the output angular frequency is the same as the input angular frequency, $\omega = 2\pi f_s$. The thyristor current, shown in figure 13.1b, is defined by equation (11.76); that is

$$L\frac{di}{dt} + Ri \quad \begin{cases} = \sqrt{2}V \sin\omega t & \text{(V)} \quad \alpha \le \omega t \le \beta \quad \text{(rad)} \\ = 0 & \text{otherwise} \end{cases} \qquad (13.1)$$

The solution to this first order differential equation has two solutions, depending on the delay angle α relative to the load natural power factor angle, $\phi = \tan^{-1}\omega L/R$.
Because of symmetry, the mean supply and load, voltages and currents, are zero.

Case 1: $\alpha > \phi$
When the delay angle exceeds the load power factor angle the load current always reaches zero before $\pi+\varphi$, thus the differential equation boundary conditions are zero. The solution for the current i is

$$i(\omega t) = \frac{\sqrt{2}V}{Z}\left\{\sin(\omega t - \phi) - \sin(\alpha - \phi)\, e^{-(\omega t+\alpha)/\tan\phi}\right\} \quad \text{(A)} \qquad \alpha \le \omega t \le \beta \quad \text{(rad)} \qquad (13.2)$$

$$i(\omega t) = 0 \quad \text{(A)} \qquad \pi \le \beta \le \omega t \le \pi + \alpha \quad \text{(rad)} \qquad (13.3)$$

where $Z = \sqrt{(R^2 + \omega^2 L^2)}$ (ohms) and $\tan\phi = \omega L/R = 1/Q$
Provided $\alpha > \phi$ both ac regulator thyristors will conduct and load current flows symmetrically as shown in figure 13.1b. The thyristor conduction period is given by the angle $\theta = \beta - \alpha$.
The thyristor current extinction angle β for discontinuous load current can be determined with the aid of figure 11.9a, but with the restriction that $\beta - \alpha \le \pi$, or figure 13.1d, or by solving equation 11.78, that is:

$$\sin(\beta - \phi) = \sin(\alpha - \phi)\, e^{(\alpha-\beta)/\tan\phi} \qquad (13.4)$$

From figure 13.1b the rms output voltage is

$$V_{rms} = \left[\tfrac{1}{2\pi}\int_\alpha^\beta \left(\sqrt{2}V\right)^2 \sin^2\omega t\, d\omega t\right]^{\frac{1}{2}} = \sqrt{2}V\left[\tfrac{1}{2\pi}\int_\alpha^\beta (1-\cos 2\omega t)\, d\omega t\right]^{\frac{1}{2}}$$
$$= V\left[\tfrac{1}{2\pi}\left\{(\beta-\alpha) - \tfrac{1}{2}(\sin 2\beta - \sin 2\alpha)\right\}\right]^{\frac{1}{2}} = V\left[\tfrac{1}{2\pi}\left\{(\beta-\alpha) - \sin(\beta-\alpha)\cos(\alpha+\beta)\right\}\right]^{\frac{1}{2}} \qquad (13.5)$$

The maximum rms output voltage is when $\alpha = \varphi$ and $\beta = \varphi + \pi$ in equation (13.5), giving $\hat{V}_{rms} = V$.
The rms load current is found by the appropriate integration of equation (13.2) squared, namely

$$I_{rms} = \left[\frac{1}{\pi}\int_\alpha^\beta \left(\frac{\sqrt{2}V}{Z}\right)^2 \left\{\sin(\omega t - \phi) - \sin(\alpha-\phi)\, e^{-(\omega t+\alpha)/\tan\phi}\right\}^2 d\omega t\right]^{\frac{1}{2}}$$
$$= \frac{V}{Z}\left[\frac{1}{\pi}\left(\beta - \alpha - \frac{\sin(\beta-\alpha)}{\cos\phi}\cos(\beta+\alpha+\phi)\right)\right]^{\frac{1}{2}} \qquad (13.6)$$

The maximum rms output current is when $\alpha = \varphi$ in equation (13.6), giving $\hat{I}_{rms} = V/Z$.

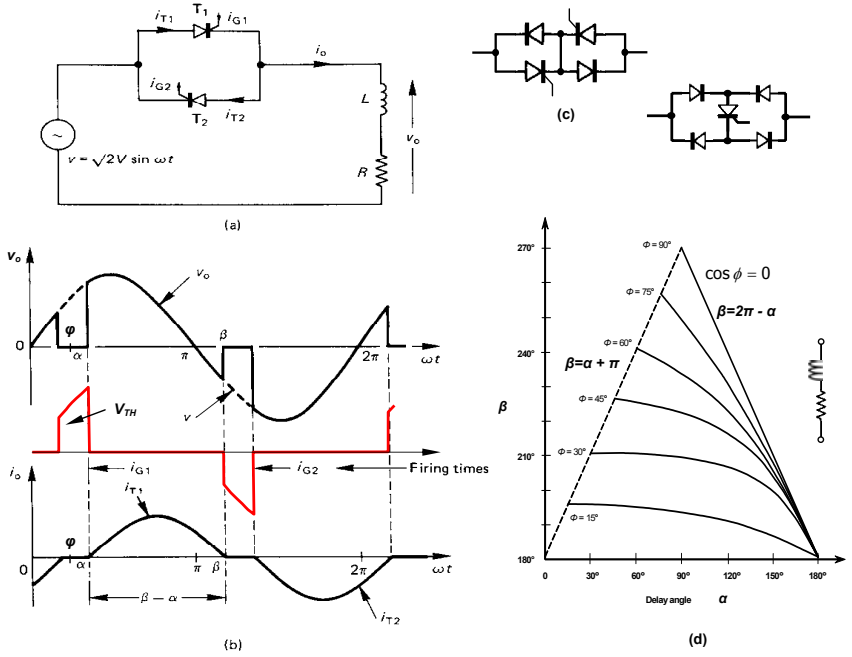


Figure 13.1. *Single-phase full-wave symmetrical thyristor ac regulator with an R-L load: (a) circuit connection; (b) load current and voltage waveforms for α>φ; (c) asymmetrical voltage blocking thyristor alternatives; and (d) current extinction angle β versus triggering delay angle α.*

From equation (13.6), the thyristor rms current is given by $I_{Th_{rms}} = I_{rms}/\sqrt{2}$ and is a maximum when $\alpha \le \phi$, that is

$$\hat{I}_{Th\,rms} = \hat{I}_{rms}\Big/\sqrt{2} = V\Big/\sqrt{2}\,Z \tag{13.7}$$

Using the fact that the average voltage across the load inductor is zero, the rectified mean voltage (hence current) can be used to determine the thyristor mean current rating.

$$\begin{aligned}\overline{V}_o &= \overline{I}_o R = \tfrac{1}{2\pi}\int_\alpha^\beta \sqrt{2}\,V\sin\omega t\; d\omega t \\ &= \sqrt{2}\,V\left[\tfrac{1}{2\pi}\{\cos\alpha - \cos\beta\}\right] \qquad \text{(V)}\end{aligned} \tag{13.8}$$

The mean thyristor current $\overline{I}_{Th} = \tfrac{1}{2}\overline{I}_o = \tfrac{1}{2}\overline{V}_o/R$, that is

$$\overline{I}_{Th} = \frac{\tfrac{1}{2}\overline{V}_o}{R} = \frac{\sqrt{2}\,V}{2R}\left[\tfrac{1}{2\pi}(\cos\alpha - \cos\beta)\right] \qquad \text{(A)} \tag{13.9}$$

The maximum mean thyristor current is for a load $\alpha = \varphi$ and $\beta = \pi+\varphi$, that is

$$\hat{\overline{I}}_{Th} = \sqrt{2}\,V\cos\phi\Big/\pi R = \sqrt{2}\,V\Big/\pi Z \tag{13.10}$$

The thyristor forward and reverse voltage blocking ratings are both $\sqrt{2}V$.
The load current form factor, using $\cos\varphi = R/Z$, is

$$\begin{aligned}FF_{i\,load} &= \frac{I_{rms}}{\overline{I}_o} \\ &= \frac{\cos\phi\left[\frac{1}{\pi}\left(\beta - \alpha - \frac{\sin(\beta-\alpha)}{\cos\phi}\cos(\beta+\alpha+\phi)\right)\right]^{1/2}}{\sqrt{2}\left[\tfrac{1}{2\pi}(\cos\alpha - \cos\beta)\right]}\end{aligned} \tag{13.11}$$

which is a maximum when $\alpha = \varphi$, giving $\hat{FF}_{i\,Load} = \pi/2\sqrt{2}$.

The thyristor current form factor is $FF_{i\,Th} = \sqrt{2}\,FF_{i\,Load}$, which is a maximum when $\alpha = \varphi$, $\hat{FF}_{i\,Th} = \tfrac{1}{2}\pi$.

The load power is

$$\begin{aligned}P_o &= I_{rms}^2 R \\ &= \frac{(V\cos\phi)^2}{R}\left[\frac{\beta-\alpha}{\pi} - \frac{\sin(\beta-\alpha)}{\pi\cos\phi}\cos(\alpha+\phi+\beta)\right]\end{aligned} \tag{13.12}$$

which is a maximum when $\alpha = \varphi$, giving $\hat{P}_o = \frac{V^2}{Z}\cos\phi = \left(\frac{V}{Z}\right)^2 R$.

The supply power factor is

$$\begin{aligned}pf &= \frac{P_o}{V\,I_{rms}} \\ &= \left[\frac{\beta-\alpha}{\pi} - \frac{\sin(\beta-\alpha)}{\pi\cos\phi}\cos(\alpha+\phi+\beta)\right]^{1/2} \times \cos\phi\end{aligned} \tag{13.13}$$

which is a maximum when $\alpha = \varphi$, giving $\hat{pf} = \cos\phi$.

For an inductive *L*-*R* load, the fundamental load voltage components (cos and sin respectively) are

$$\begin{aligned}a_1 &= \frac{\sqrt{2}\,V}{2\pi}(\cos 2\alpha - \cos 2\beta) \\ b_1 &= \frac{\sqrt{2}\,V}{2\pi}\left(2(\beta-\alpha) - (\sin 2\beta - \sin 2\alpha)\right)\end{aligned} \tag{13.14}$$

$$\begin{aligned}a_n &= \frac{\sqrt{2}V}{\pi}\left[\frac{\cos(n+1)\alpha - \cos(n+1)\beta}{n+1} - \frac{\cos(n-1)\alpha - \cos(n-1)\beta}{n-1}\right] \\ b_n &= \frac{\sqrt{2}V}{\pi}\left[\frac{\sin(n+1)\alpha - \sin(n+1)\beta}{n+1} - \frac{\sin(n-1)\alpha - \sin(n-1)\beta}{n-1}\right]\end{aligned} \tag{13.15}$$

for n = 3, 5, 7, .. odd.

The Fourier component magnitudes and phases are given by

$$c_n = \sqrt{a_n^2 + b_n^2} \qquad \text{and} \qquad \phi_n = \tan^{-1} a_n\Big/b_n \tag{13.16}$$

If $\alpha = \phi$, then continuous ac load current flows, and equation (13.14) reduces to $a_1 = 0$ and $b_1 = \sqrt{2}V$, when $\beta = \alpha + \pi = \varphi + \alpha$ and $\alpha = \phi$ are substituted.

The supply apparent power can be grouped into a component at the fundamental frequency plus components at the harmonic frequencies.

$$\begin{aligned}S^2 &= V^2I_1^2 + V^2I_3^2 + V^2I_5^2 + \ldots \\ &= V^2I_1^2 + V^2I_1^2 + V^2I_3^2 + V^2I_5^2 + \ldots - V^2I_1^2 \\ &= V^2I_1^2 + V^2I_{rms}^2 - V^2I_1^2 \\ &= V^2I_1^2 + V^2\left(I_{rms}^2 - I_1^2\right) \\ &= S_1^2 + D^2 \\ &= P^2 + Q_1^2 + D^2 \\ S^2 &= (VI_1\cos\phi_1)^2 + (VI_1\sin\phi_1)^2 + D^2\end{aligned} \tag{13.17}$$

where D is the supply current distortion due to the harmonic currents.

The current harmonic components are found by dividing the load Fourier voltage components by the load impedance at that frequency. Equation (13.16) gives the current harmonic angles ϕ_n and magnitudes according to

$$I_n = \frac{V_n}{Z_n} = \frac{V_n}{\sqrt{R^2 + (\omega L)^2}} = \frac{c_n\Big/\sqrt{2}}{\sqrt{R^2 + (\omega L)^2}} \tag{13.18}$$

Case 2: $\alpha \le \phi$ (continuous gate pulses)
When $\alpha \le \phi$, a pure sinusoidal load current flows, and substitution of $\alpha = \phi$ in equation (13.2) results in

$$i(\omega t) = \frac{\sqrt{2}V}{Z}\sin(\omega t - \phi) \qquad \text{(A)} \qquad \alpha \le \phi \qquad \text{(rad)} \tag{13.19}$$

If a short duration gate trigger pulse is used and $\alpha < \phi$, unidirectional load current may result. The device to be turned on is reverse-biased by the conducting device. Thus if the gate pulse ceases before the previous half-cycle load current has fallen to zero, only one device conducts. It is therefore usual to employ a continuous gate pulse, or stream of pulses, from α until π, then for $\alpha < \phi$ a sine wave output current results.

For both delay angle conditions, equations (13.5) to (13.14) are valid, except the simplification $\beta=\alpha+\pi$ is used when $\alpha \le \phi$, which gives the maximum values for those equations. That is, for $\alpha \le \phi$, substituting $\alpha = \phi$

$$\begin{aligned}&\hat{V}_{rms} = V_{rms} = V \qquad \hat{I}_{rms} = I_{rms} = V/Z \qquad \hat{I}_{Th\,rms} = I_{Th\,rms} = V\Big/\sqrt{2}\,Z \\ &\hat{FF}_{iLoad} = FF_{iLoad} = \pi/2\sqrt{2} \qquad \hat{P}_o = P_o = I_{rms}^2 R = V^2\cos\phi/Z \qquad \hat{pf} = pf = \cos\phi \qquad \overline{I}_{Th} = \sqrt{2}\,V/\pi Z\end{aligned} \tag{13.20}$$

13.1.1i - Resistive load

For a purely resistive load, the load voltage and current are related according to

$$i_o(\omega t) = v_o(\omega t)\Big/R = \begin{cases}\sqrt{2}\,V\sin(\omega t)\Big/R & \alpha \le \omega t \le \pi, \quad \alpha+\pi \le \omega t \le 2\pi \\ 0 & \text{otherwise}\end{cases} \tag{13.21}$$

The equations (13.1) to (13.20) can be simplified if the load is purely resistive. Continuous output current only flows for $\alpha = 0$, since $\phi = \tan^{-1}0 = 0°$. Therefore the output equations are derived from the discontinuous equations (13.2) to (13.14), with $\varphi = 0$.

The average output voltage and current are zero. The mean half-cycle output voltage, used to determine the thyristor mean current rating, is found by integrating the supply voltage over the interval α to π, ($\beta = \pi$).

$$V_o = \tfrac{1}{2\pi}\int_{\alpha}^{\pi} \sqrt{2}V \sin\omega t \, d\omega t$$
$$= \frac{\sqrt{2}V}{\pi}(1+\cos\alpha) \qquad \text{(V)} \qquad (13.22)$$

whence $\bar{I}_o = \dfrac{V_o}{R} = \dfrac{\sqrt{2}V}{\pi R}(1+\cos\alpha) = 2\bar{I}_T$ (A)

The average thyristor current is $\bar{I}_T = \tfrac{1}{2}\bar{I}_o$, which has a maximum value of $\hat{I}_T = \sqrt{2}V/\pi R$ when $\alpha = 0$.
From equation (13.5) the rms output voltage for a delay angle α is

$$V_{rms} = \sqrt{\tfrac{1}{2\pi}\int_{\alpha}^{\pi}\left(\sqrt{2}V\sin\omega t\right)^2 d\omega t}$$
$$= V\sqrt{\frac{2(\pi-\alpha)+\sin 2\alpha}{2\pi}} \qquad \text{(V)} \qquad (13.23)$$

which has a maximum of $\hat{V}_{rms} = V$ when $\alpha = 0$.

The rms output current and supply current from $I_{rms} = V_{rms}/R$ is

$$I_{rms} = \frac{V_{rms}}{R} = \frac{V}{R}\sqrt{1-\frac{2\alpha-\sin 2\alpha}{2\pi}} = \sqrt{2}\,I_{T\,rms} \qquad \text{(A)} \qquad (13.24)$$

and $I_{T\,rms} = I_{rms}/\sqrt{2}$

The maximum rms supply current is $I_{rms} = V/R$ at α = 0 when the maximum rms thyristor current is $\hat{I}_{T\,rms} = \hat{I}_{rms}/\sqrt{2} = V/\sqrt{2}R$.
Therefore the output power, with $V_{rms} = R\,I_{rms}$, for a resistive load, is

$$P_o = I_{rms}^2 R = \frac{V_{rms}^2}{R} = \frac{V^2}{R}\left\{1-\frac{2\alpha-\sin 2\alpha}{2\pi}\right\} \qquad \text{(W)} \qquad (13.25)$$

The input power is $P_{in} = VI_1\cos\phi_1 \quad (= P_{out})$.
The supply power factor λ is defined as the ratio of the real power to the apparent power, that is

$$pf = \lambda = \frac{P_o}{S} = \frac{V_{rms}\,I}{V\,I} = \frac{V_{rms}}{V} = \sqrt{\frac{2(\pi-\alpha)+\sin 2\alpha}{2\pi}} \qquad (13.26)$$

where the apparent power is

$$S = VI_{rms} = \frac{V^2}{R}\left[1-\frac{\alpha}{\pi}+\frac{\sin 2\alpha}{2\pi}\right]^{\frac{1}{2}} \qquad (13.27)$$

and $Q = \sqrt{S^2 - P^2}$. The fundament reactive power is

$$Q_1 = \frac{V^2}{R}\frac{\cos 2\alpha - 1}{2\pi} \qquad (13.28)$$

The thyristor current (and voltage for a resistive load) form factor (rms to mean), shown in figure 13.2, is

$$FF_{i_{Th}} = \frac{I_{Th\,rms}}{\bar{I}_{Th}} = \frac{\left[\pi\left(\pi-\alpha+\tfrac{1}{2}\sin 2\alpha\right)\right]^{\frac{1}{2}}}{1+\cos\alpha} \qquad (13.29)$$

From equation (13.161) , the thyristor current crest factor is

$$\delta = \frac{\hat{I}_T}{\bar{I}_T} = \begin{cases} \dfrac{2\pi}{1+\cos\alpha} & 0 \le \alpha \le \tfrac{1}{2}\pi \\ \dfrac{2\pi\sin\alpha}{1+\cos\alpha} & \tfrac{1}{2}\pi \le \alpha \le \pi \end{cases} \qquad (13.30)$$

The Fourier voltage components for a resistive load (with $\beta = \pi$ in equations (13.14) and (13.15)) are

$$a_1 = \frac{\sqrt{2}V}{\pi}\left(\tfrac{1}{2}\cos 2\alpha - \tfrac{1}{2}\right) \qquad b_1 = \frac{\sqrt{2}V}{\pi}\left(\pi-\alpha+\tfrac{1}{2}\sin 2\alpha\right)$$
$$a_n = \frac{\sqrt{2}V}{\pi}\left(\frac{\cos(n+1)\alpha - 1}{n+1} - \frac{\cos(n-1)\alpha - 1}{n-1}\right) \qquad b_n = \frac{\sqrt{2}V}{\pi}\left(\frac{\sin(n+1)\alpha}{n+1} - \frac{\sin(n-1)\alpha}{n-1}\right) \qquad (13.31)$$

for n = 3, 5, 7, … odd. Figure 13.2 show the relative harmonic rms magnitudes and dependence on α.
The load current harmonics are found by dividing the voltage components by R, since $i(\omega t) = v(\omega t)/R$.

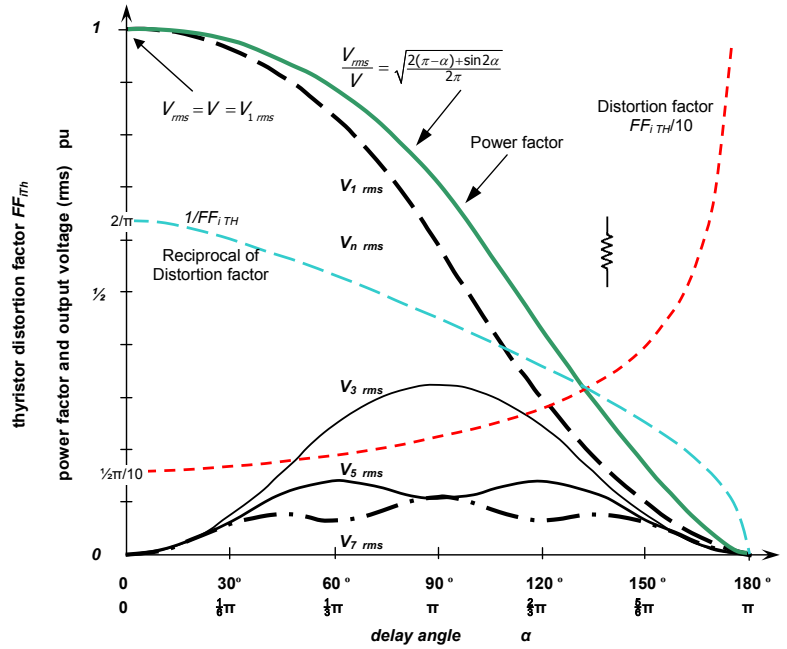


Figure 13.2. *Normalised RMS harmonics (voltage and current) for a single-phase full-wave ac regulator with a pure resistive load.*

The fundamental supply current is

$$i_{s1} = \frac{v_{o1}}{R} = \frac{1}{R}\left[b_1\sin\omega t + a_1\cos\omega t\right]$$
$$i_{s1} = \frac{\sqrt{2}V}{\pi R}\left[\left(\pi-\alpha+\tfrac{1}{2}\sin 2\alpha\right)\sin\omega t - \left(\tfrac{1}{2}\cos 2\alpha - \tfrac{1}{2}\right)\cos\omega t\right] \qquad (13.32)$$

which has an rms value of

$$I_{s1} = \frac{V}{2\pi R}\left[\left(\cos 2\alpha - 1\right)^2 + \left(2\pi - 2\alpha + \sin 2\alpha\right)^2\right]^{\frac{1}{2}} \qquad (13.33)$$

When the power factor λ in equation (13.26) can be expressed in terms of the distortion factor and displacement factor, that is

$$\lambda = \frac{\text{average power}}{\text{apparent } VA} = \frac{P}{V\,I_{rms}} = \frac{V\,i_{s1}\cos\phi_1}{V\,I_{rms}} = \frac{i_{s1}}{I_{rms}}\cos\phi_1$$
$$= \text{distortion factor} \times \text{diplacement factor}$$

The current distortion factor is equation (13.33) divided by equation (13.24), while the fundamental current displacement factor from the fundamental components in equation (13.31) yields

$$\text{diplacement factor} = \cos\phi_1 \text{ where } \phi_1 = \tan^{-1}\left(a_1/b_1\right)$$

If the thyristors are modelled by

$$v_{TH} = v_o + i \times r_o$$

Then the thyristor losses are given by

$$P_{TH} = v_o\bar{I}_{Th} + r_o \times I_{Th\,rms}^2 = v_o\bar{I}_{TH} + r_o \times \bar{I}_{Th}^2 \times FF_{i\,Th}(\alpha) \qquad \text{(W)} \qquad (13.34)$$

13.1.1ii - Pure inductive load

For a purely inductive load, the load power factor angle is $\phi = \tfrac{1}{2}\pi$. Since the inductor voltage average is zero, current conduction will be symmetrical about π. Thus equations (13.2) to (13.14) apply except they can be simplified since $\beta = 2\pi - \alpha$. These bounds imply that the delay angle should be greater than ½π,

but less than π. Therefore, if the delay angle is less than ½π, conduction extends into the next half cycle, and with short gate pulses, preventing the reverse direction thyristor from conducting, as shown in figure 13.3c. The output is then a series of half-wave rectified current pulses as with the case $\alpha \le \phi$ considered in 13.3iii. For the purely inductive load case, the equations and waveforms for the half-wave controlled rectifier in section 11.3.1ii, apply. Kirchhoff's voltage law gives

$$L\frac{di}{dt} = \sqrt{2}V\sin\omega t \qquad (13.35)$$

The load current is given by:

$$i(\omega t) = \frac{\sqrt{2}V}{\omega L}\left(\cos\alpha - \cos(\omega t)\right) \qquad \alpha \le \omega t \le 2\pi - \alpha \qquad (13.36)$$

The current waveform is symmetrical about π.

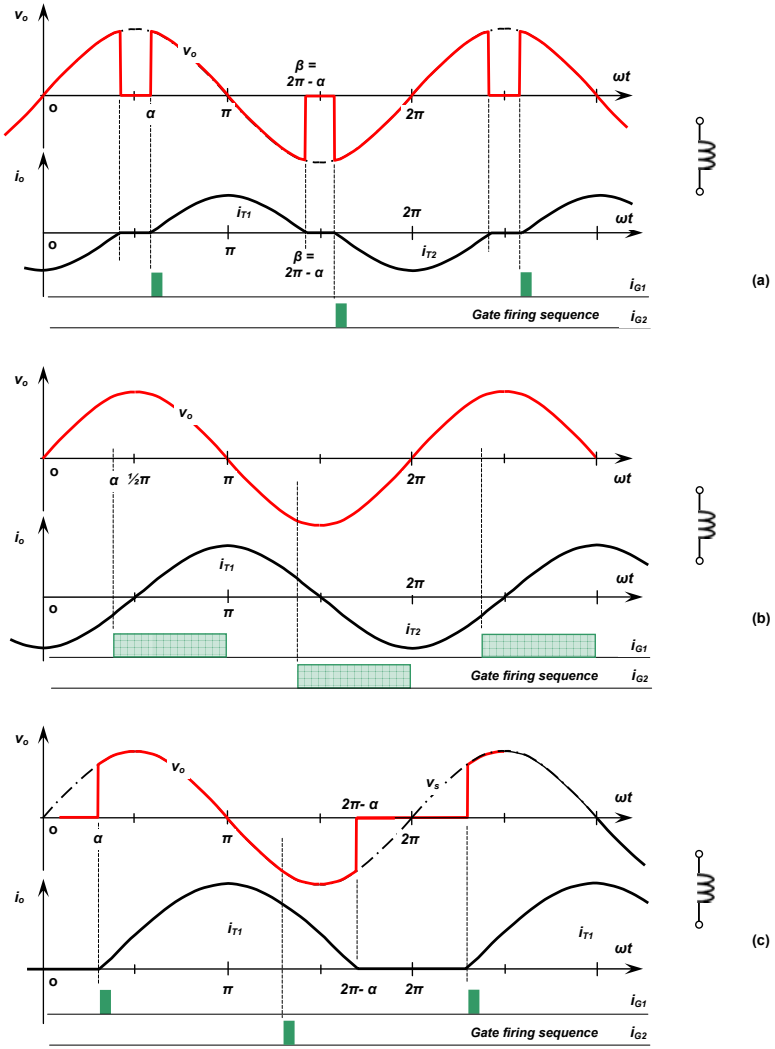


Figure 13.3. *Single-phase full-wave thyristor ac regulator with a pure inductor load: (a) α > ½π; (b) α < ½π, gate pulse until π; and (c) α < ½π, short gate pulse.*

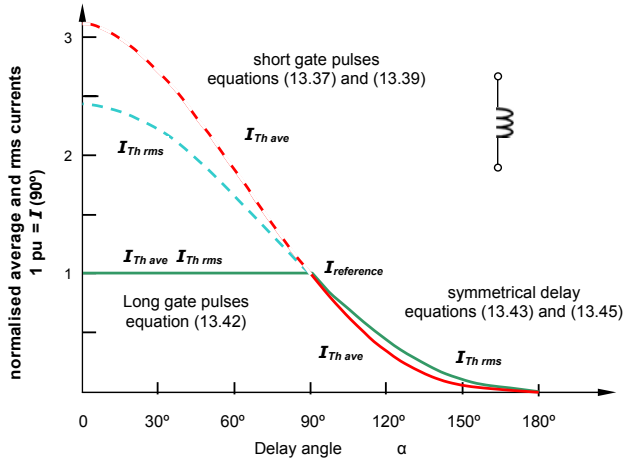
α < ½π

i. **short** gate pulse period

With a purely inductive load, the average output voltage is zero. If uni-directional current flows (due to the uses of a narrow gate pulse), as shown in figure 13.3c, the average load current, hence average thyristor current, for the conducting thyristor, is

$$\bar{I}_o = \bar{I}_T = \tfrac{1}{2\pi}\int_{\alpha}^{2\pi-\alpha}\frac{\sqrt{2}\,V}{\omega L}\{\cos\alpha - \cos\omega t\}\,d\omega t$$
$$= \frac{\sqrt{2}\,V}{\pi\omega L}\left[(\pi-\alpha)\cos\alpha + \sin\alpha\right] \qquad (13.37)$$

which with uni-polar pulses has a maximum of $\sqrt{2}\,V/\omega L$ at $\alpha = 0$.



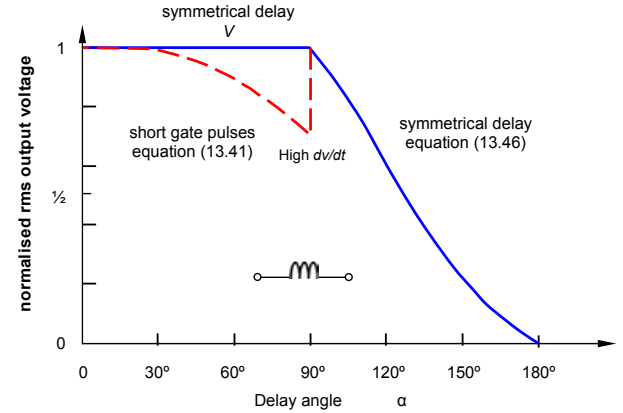


Figure 13.4. *Normalised ac-chopper purely inductive load control characteristics of: (a) thyristor average and rms currents and (b) rectified average output voltage.*

The rectified average load voltage is

$$V_o = \frac{\sqrt{2}V}{\pi}(1+\cos\alpha) \qquad 0 \le \alpha \le ½\pi \qquad (13.38)$$

The rms load and supply (and one thyristor) current is

$$I_{rms} = \frac{\sqrt{2}\,V}{\omega L}\left[\tfrac{1}{2\pi}\int_{\alpha}^{2\pi-\alpha}(\cos\alpha - \cos\omega t)^2\,d\omega t\right]^{\frac{1}{2}} = \frac{V}{X}\left[\frac{1}{\pi}\left((\pi-\alpha)(2+\cos 2\alpha)+\frac{3}{2}\sin 2\alpha\right)\right]^{\frac{1}{2}} \quad (13.39)$$

The thyristor current form factor is

$$FF_{iT} = \frac{\tfrac{1}{2}\pi(\pi-\alpha)(2+\cos 2\alpha)+\tfrac{3}{4}\pi\sin 2\alpha}{\sin\alpha+(\pi-\alpha)\cos\alpha} \quad (13.40)$$

which has a maximum value of ½π when $\alpha = \frac{1}{2}\pi$.
The rms load voltage is

$$V_{rms} = \left[\tfrac{1}{2\pi}\int_{\alpha}^{2\pi-\alpha}\left(\sqrt{2}V\right)^2\sin^2\omega t\,d\omega t\right]^{\frac{1}{2}} = V\left[\tfrac{1}{2\pi}\left\{(\pi-\alpha)+\tfrac{1}{2}\sin 2\alpha\right\}\right]^{\frac{1}{2}} \quad (13.41)$$

ii. **extended** gate pulse period
When the gate pulses are extended to π, continuous current flows, as shown in figure 13.3b, given by $I_{rms}=V/\omega L$, lagging V by ½π. Each thyristor conducts an average current and rms current of

$$\bar{I}_{Th} = \frac{\sqrt{2}V}{\pi\omega L} \qquad I_{Th\,rms} = \frac{I_{rms}}{\sqrt{2}} = \frac{V}{\pi\omega L} \quad (13.42)$$

π ≥ α ≥ ½π (symmetrical gate pulses)
The output voltage and current are symmetrical, as shown in figure 13.3a, hence the average output voltage and current are both zero, as is the average input current. The average thyristor current is given by

$$\bar{I}_T = \tfrac{1}{2\pi}\int_{\alpha}^{2\pi-\alpha}\frac{\sqrt{2}\,V}{\omega L}\{\cos\alpha - \cos\omega t\}\,d\omega t = \frac{\sqrt{2}\,V}{\pi\omega L}\left[(\pi-\alpha)\cos\alpha+\sin\alpha\right] \quad (13.43)$$

which has a maximum of $\sqrt{2}\,V/\pi\omega L$ at $\alpha = \frac{1}{2}\pi$.

The rectified average load voltage over half a cycle is

$$V_o = \frac{2\sqrt{2}V}{\pi}(1+\cos\alpha) \quad (13.44)$$

The rms load and supply current is

$$I_{rms} = \sqrt{2}I_{Th\,rms} = \frac{V}{\omega L}\left[\tfrac{2}{\pi}\int_{\alpha}^{2\pi-\alpha}(\cos\alpha - \cos\omega t)^2\,d\omega t\right]^{\frac{1}{2}} = \frac{V}{X}\left[\frac{2}{\pi}\left((\pi-\alpha)(2+\cos 2\alpha)+\frac{3}{2}\sin 2\alpha\right)\right]^{\frac{1}{2}} \quad (13.45)$$

The rms load voltage is

$$V_{rms} = \left[\tfrac{1}{2\pi}\int_{\alpha}^{2\pi-\alpha}\left(\sqrt{2}V\right)^2\sin^2\omega t\,d\omega t\right]^{\frac{1}{2}} = V\left[\tfrac{2}{2\pi}\left\{(\pi-\alpha)+\tfrac{1}{2}\sin 2\alpha\right\}\right]^{\frac{1}{2}} \quad (13.46)$$

The maximum rms voltage and current are $\hat{V}_{rms} = V$ and $\hat{I}_{rms} = V/X$ at $\alpha = \frac{1}{2}\pi$.

The rms equations for α greater than and less than ½π are basically the same except the maximum period over which a given thyristor conducts changes from π to 2π (respectively), hence the rms values differ by √2. Since the output power is zero, the supply power factor is zero, for bidirectional current.
If the controller in figure 13.1a is use in the half-controlled mode (thyristor and anti-parallel diode), the resultant dc component precludes its use in ac transformer applications. The controller is limited to low power ac applications because of dc restrictions on the ac mains supply.

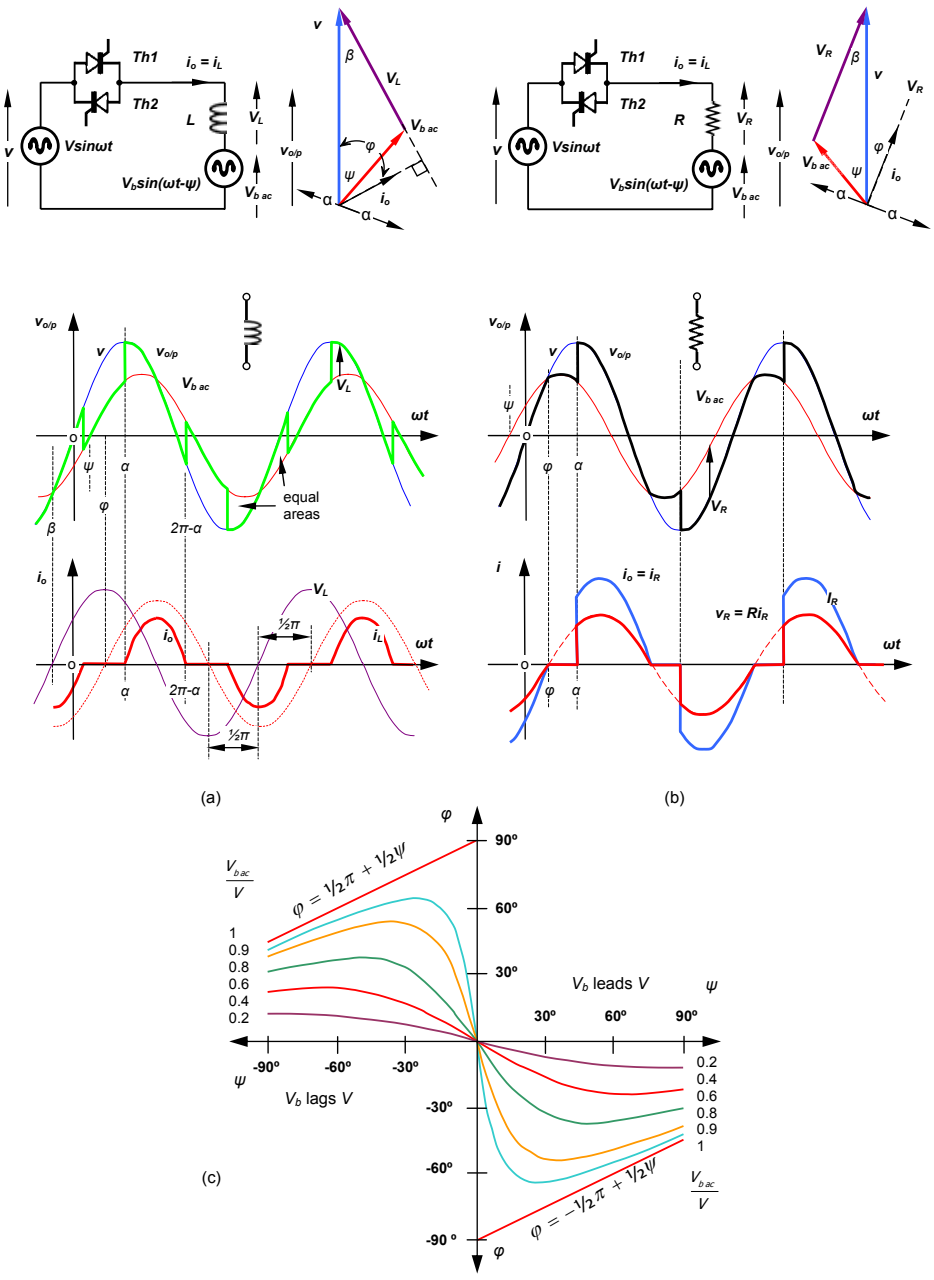


Figure 13.5. *AC-chopper characteristics with ac back emf and purely resistive or inductive load. Circuit, phasor diagram and circuit waveforms for: (a) purely resistive and ac source load and (b) purely inductive and ac source load. (c) Phase displacement of resultant voltage of an ac emf opposing the ac mains.*

13.1.1iii - Load sinusoidal back emf

When the ac controller load comprises an ac back emf $v_{b\,ac}$ of the same frequency as the ac supply v, as with embedded generation, then, when the thyristors conduct, the load effectively sees the vector difference between the two ac voltages, v-$v_{b\,ac}$, as shown in figure 13.5.

$$\begin{aligned} v_{R,L} &= v - v_{b\,ac} \\ &= V\angle 0 - V_{b\,ac}\angle\psi = V + 0j - V_{b\,ac}\left(\cos\psi + j\sin\psi\right) \\ &= V - V_{b\,ac}\cos\psi - jV_{b\,ac}\sin\psi \\ &= v_{R,L}\angle\varphi \end{aligned} \qquad (13.47)$$

where

$$\begin{aligned} v_{R,L} &= \sqrt{\left(V - V_{b\,ac}\cos\psi\right)^2 + \left(V_{b\,ac}\sin\psi\right)^2} = \sqrt{V^2 + V_{b\,ac}^2 - 2VV_{b\,ac}\cos\psi} \\ \varphi &= \tan^{-1}\frac{-V_{b\,ac}\sin\psi}{V - V_{b\,ac}\cos\psi} \end{aligned} \qquad (13.48)$$

The passive part of the load can now be analysed as in sections 13.1.1i and ii, but the thyristor phase triggering delay angles are shifted by φ with respect to the original ac supply reference, as shown in the phasor diagrams in figure 13.5.
If the voltage is normalised with respect to the ac supply V, then the normalised curves in figure 13.5 can be used to obtain the phase angle φ, with respect to the ac mains reference. Therefore curves give the angle of the voltage (and the current in the case of a resistor load) across the passive part of the load.
As seen in the waveforms in figure 13.5, the load current is dependent on the relative magnitudes and angle between the two ac sources, the type of load, and the thyristor phase delay angle. Performance features with a resistive load and inductive load are illustrated in Example 13.1d.

13.1.1iv - Semi-controlled single-phase ac regulator

A semi-controlled single-phase ac regulator is formed by replacing one thyristor in figure 13.1a with a diode. A dc component results in the load current and voltage. For a resistive load, the diode average and rms currents are found by substituting α = 0 in equations (13.22) and (13.24). Using these equations, the load resistance average and rms currents (hence voltages) are

$$\begin{aligned} \bar{I}_R &= \bar{I}_D - \bar{I}_T = \frac{2\sqrt{2}}{\pi R} - \frac{2\sqrt{2}}{\pi R}\left(1+\cos\alpha\right) = \frac{2\sqrt{2}}{\pi R}\left(1-\cos\alpha\right) = \frac{V_o}{R} \\ I_{R\,rms} &= \sqrt{I_{D\,rms}^2 + I_{T\,rms}^2} = \frac{V}{2R}\left[\frac{2\alpha - \sin 2\alpha}{\pi}\right]^{1/2} = \frac{V_{rms}}{R} \end{aligned} \qquad (13.49)$$

The power dissipated in the resistive load is

$$P_R = \frac{V^2}{R}\left[\frac{2\alpha - \sin 2\alpha}{4\pi}\right] \qquad (13.50)$$

Example 13.1a: ***Single-phase ac regulator – 1***

If the load of the 50 Hz 240V ac voltage regulator shown in figure 13.1 is Z = 7.1+j7.1 Ω, calculate the load natural power factor angle, ϕ. Then, assuming bipolar load current conduction, calculate

(a) the rms output voltage, thence
(b) the output power and rms current, whence input power factor and supply current distortion factor, μ

for

i. α = ⅙π
ii. α = ⅓π

Solution

From equation (13.3) the load natural power factor angle is

$$\phi = \tan^{-1}\omega L/R = \tan^{-1}X_L/R = \tan^{-1}7.1/7.1 = \tfrac{1}{4}\pi \text{ (rad)}$$
$$Z = \sqrt{R^2 + (\omega L)^2} = \sqrt{7.1^2 + 7.1^2} = 10\Omega$$

i. $\alpha = \frac{1}{6}\pi$

(a) Since $\alpha = \pi/6 < \phi = \pi/4$, the load current is continuous and bidirectional, ac. The rms load voltage is 240V.

(b) From equation (13.20) the power delivered to the load is

$$P_o = I_{rms}^2 R = \frac{V^2}{Z}\cos\phi$$
$$= \frac{240^2}{10\Omega}\cos\tfrac{1}{4}\pi = 4.07\text{kW}$$

The rms output current and supply current are both given by

$$I_{rms} = \sqrt{P_o/R}$$
$$= \sqrt{4.07\text{kW}/7.1\Omega} = 23.8\text{A}$$

The input power factor is the load natural power factor, that is

$$pf = \frac{P_o}{S} = \frac{4.07\text{kW}}{240\text{V}\times 23.8\text{A}} = 0.70$$
$$= \mu\cos\phi = \mu/\sqrt{2}$$

Thus the current input distortion factor is μ = 1, for this sinusoidal current case.

ii. α = ⅓π

(a) Since $\alpha = \pi/3 > \phi = \tfrac{1}{4}\pi$, the load hence supply current is discontinuous. For $\alpha = \pi/3 > \phi = \tfrac{1}{4}\pi$ the extinction angle β = 3.91 rad or 224.15° can be extracted from figure 11.7a or determined after iteration using equation (13.4). The rms load voltage is given by equation (13.5).

$$V_{rms} = V\left[\tfrac{1}{2\pi}\left\{\left(\beta-\alpha\right) - \tfrac{1}{2}\left(\sin 2\beta - \sin 2\alpha\right)\right\}\right]^{1/2}$$
$$= 240\times\left[\tfrac{1}{2\pi}\left\{\left(3.91 - \tfrac{1}{3}\pi\right) - \tfrac{1}{2}\left(\sin 2\times 3.91 - \sin\tfrac{2}{3}\pi\right)\right\}\right]^{1/2}$$
$$= 240\times\sqrt{2.7\tfrac{1}{2\pi}} = 226.4\text{V}$$

(b) The rms output (and input) current is given by equation (13.6), that is

$$I_{Orms} = \frac{V}{Z}\left[\frac{1}{\pi}\left(\beta - \alpha - \frac{\sin\left(\beta-\alpha\right)}{\cos\phi}\cos\left(\beta+\alpha+\phi\right)\right)\right]^{1/2}$$
$$= \frac{240}{10}\left[\frac{1}{\pi}\left(3.91 - \tfrac{1}{3}\pi - \frac{\sin\left(3.91 - \frac{1}{3}\pi\right)}{\cos\tfrac{1}{4}\pi}\cos\left(3.91 + \tfrac{1}{3}\pi + \tfrac{1}{4}\pi\right)\right)\right]^{1/2} = 18.0\text{A}$$

The output power is given by

$$P_o = I_{rms}^2 R$$
$$= 18.0^2\times 7.1\Omega = 2292\text{W}$$

The load and supply power factors are

$$pf_o = \frac{P_o}{S} = \frac{2292\text{W}}{226.4\text{V}\times 18.0\text{A}} = 0.562 \quad pf = \frac{P_o}{S} = \frac{2292\text{W}}{240\text{V}\times 18.0\text{A}} = 0.531$$

The Fourier coefficients of the fundamental, a_1 and b_1, are given by equation (13.14)

$$a_1 = \frac{\sqrt{2}\,V}{2\pi}\left(\cos 2\alpha - \cos 2\beta\right) = \frac{\sqrt{2}\;240\text{V}}{2\pi}\left(\cos 2\tfrac{\pi}{3} - \cos 2\times 3.91\right) = -28.8\text{V}$$

$$b_1 = \frac{\sqrt{2}\,V}{2\pi}\left(2\left(\beta-\alpha\right) - \sin 2\beta + \sin 2\alpha\right) = \frac{\sqrt{2}\;240\text{V}}{2\pi}\left(2\times\left(3.91 - \tfrac{\pi}{3}\right) - \sin 2\times 3.91 - \sin 2\times\tfrac{\pi}{3}\right) = 302.1\text{V}$$

The fundamental power factor is

$$\cos\phi_1 = \cos\left(\tan^{-1}\left(a_1/b_1\right)\right) = \cos\left(\tan^{-1}\left(-28.8\text{V}/302.1\text{V}\right)\right) = 0.995$$

The current distortion factor is derived from

$$pf = \mu\times\cos\phi_1$$
$$0.531 = \mu\times 0.995$$

That is, the current distortion factor is μ = 0.533.

♣

Example 13.1b: ***Single-phase ac regulator – 2***

If the load of the 50 Hz 240V ac voltage regulator shown in figure 13.1 is Z = 7.1+j7.1 Ω, calculate the minimum controllable delay angle. Using this angle calculate

i. maximum rms output voltage and current, and hence
ii. maximum output power and power factor
iii. thyristor *I-V* and *di/dt* ratings

Solution

As in example 13.1a, from equation (13.3) the load natural power factor angle is

$$\phi = \tan^{-1}\omega L/R = \tan^{-1}7.1/7.1 = \pi/4$$

The load impedance is Z=10Ω. The controllable delay angle range is $¼\pi \le \alpha \le \pi$.

i. The maximum controllable output occurs when α = ¼π.
From equation (13.2) when $\alpha = \phi$ the output voltage is the supply voltage, V, and

$$i(\omega t) = \frac{\sqrt{2}V}{Z}\sin\ (\omega t - ¼\pi) \qquad \text{(A)}$$

The load hence supply rms maximum current, is therefore

$$I_{rms} = 240\text{V}/10\Omega = 24\text{A}$$

ii. Power $= I_{rms}^2 R = 24^2 \times 7.1\Omega = 4090\text{W}$

$$\text{power factor} = \frac{\text{power output}}{\text{apparent power output}}$$

$$= \frac{I_{rms}^2 R}{VI_{rms}} = \frac{24^2 \times 7.1\Omega}{240\text{V} \times 10\text{A}} = 0.71 \quad (=\cos\phi) \quad \mu = 1$$

iii. Each thyristor conducts for π radians, between α and $\pi+\alpha$ for T1 and between $\pi+\alpha$ and $2\pi+\alpha$ for T2.
The thyristor average current is

$$\bar{I}_T = \frac{1}{2\pi}\int_{\alpha=\phi}^{\alpha+\pi=\phi+\pi} \sqrt{2}V\sin(\omega t - \phi)\, d\omega t$$

$$= \frac{\sqrt{2}V}{\pi Z} = \frac{\sqrt{2}\times 240\text{V}}{\pi \times 10\Omega} = 10.8\text{A}$$

The thyristor rms current rating is

$$I_{Trms} = \left[\frac{1}{2\pi}\int_{\alpha=\phi}^{\alpha+\pi=\phi+\pi} \left\{\sqrt{2}V\sin(\omega t - \phi)\right\}^2 d\omega t\right]^{½}$$

$$= \frac{\sqrt{2}V}{2Z} = \frac{\sqrt{2}\times 240\text{V}}{2\times 10\Omega} = 17.0\text{A} = I_{rms}/\sqrt{2}$$

Maximum thyristor *di/dt* is derived from

$$d\,i(\omega t)/dt = d/dt\ \frac{\sqrt{2}V}{Z}\sin\ (\omega t - ¼\pi)$$

$$= \frac{\sqrt{2}V}{Z}\omega\cos\ (\omega t - ¼\pi) \qquad \text{(A/s)}$$

This has a maximum value when ωt-¼π = 0, that is at $\omega t = \alpha = \phi$, then

$$d\hat{i}(\omega t)/dt = \frac{\sqrt{2}V\omega}{Z}$$

$$= \frac{\sqrt{2}\times 240\text{V}\times 2\pi\times 50\text{Hz}}{10\Omega} = 10.7\text{A/ms}$$

Thyristor forward and reverse blocking voltage requirements are $\sqrt{2}V = \sqrt{2}\times 240 = 340$Vdc.

♣

Example 13.1c: ***Single-phase ac regulator – pure inductive load***

If the load of the 50 Hz 240V ac voltage regulator shown in figure 13.1 is Z = jX= j10 Ω, and the delay angle α is first ¾π then second ¼π calculate

i. maximum rms output voltage and current, and hence
ii. thyristor *I-V* ratings

Assume the thyristor gate pulses are of a short duration relative to the 10ms half period.

Solution

For a purely inductive load, the current extinction angle is always $\beta = 2\pi-\alpha$, that is, symmetrical about π and tan$\Phi \to \infty$.

a. If the delay angle π > α > ½π and symmetrical, then the load current is discontinuous alternating polarity current pulses as shown in figure 13.3a.
b. If the delay angle 0 < α < ½π, and a short duration gate pulse is used for each thyristor, then the output comprises discontinuous unidirectional current pulses of duration 2π-2α, as shown in figure 13.3c.

a. α = ¾π: symmetrical gate pulses - discontinuous alternating current pulses.
The average output voltage and current are zero, $\bar{I}_o = \bar{V}_o = 0$. The maximum rms load voltage and current, with bidirectional output current and voltage, are when α = ½π

$$\hat{V}_{rms} = V = 240\text{V}$$

$$\hat{I}_{rms} = \frac{V}{X} = \frac{240\text{V}}{10\Omega} = 24\text{A}$$

i. The rms output current and voltage are given by equations (13.45) and (13.46), respectively, with $\Phi = \pi$ and $\beta = 2\pi-\alpha$, that is

$$I_{rms} = \frac{V}{X}\left[\frac{2}{\pi}\left((\pi-\alpha)(2+\cos 2\alpha) + \frac{3}{2}\sin 2\alpha\right)\right]^{½}$$

$$= \frac{240\text{V}}{10\Omega}\left[\frac{2}{\pi}\left((\pi - ¾\pi)\left(2+\cos\frac{3}{2}\pi\right) + \frac{3}{2}\sin\frac{3}{2}\pi\right)\right]^{½} = 5.1\text{A}$$

$$V_{rms} = V\left[⅔\left\{(\pi-\alpha) + ½\sin 2\alpha)\right\}\right]^{½}$$

$$= 240\text{V}\left[⅔\left\{(\pi - ¾\pi) + ½\sin\frac{3}{2}\pi\right)\right\}\right]^{½} = 54.65\text{V}$$

ii. Each thyristor conducts half the load current hence I_T = 5.1A/√2=3.6A rms. Before start-up, at shut-down or during operation, each thyristor has to block bi-directionally √2 240 = 340V, peak.
The average thyristor current is

$$\bar{I}_T = \frac{\sqrt{2}\ V}{\pi\omega L}\left[(\pi-\alpha)\cos\alpha + \sin\alpha\ \right]$$

$$= \frac{\sqrt{2}\ 240\text{V}}{\pi\times 10\Omega}\left[(\pi - ¾\pi)\cos ¾\pi + \sin ¾\pi\ \right] = 1.64\text{A}$$

b. α = ¼ π: short gate pulses – discontinuous unidirectional current pulses.
The average output voltage and current are not zero, $\bar{I}_o \neq 0$ and $\bar{V}_o \neq 0$.

i. The rms output current and voltage are given by equations (13.45) and (13.46), respectively, with $\Phi = \pi$ and $\beta = 2\pi-\alpha$, that is

$$I_{rms} = \frac{V}{X}\left[\frac{1}{\pi}\left((\pi-\alpha)(2+\cos 2\alpha) + \frac{3}{2}\sin 2\alpha\right)\right]^{½}$$

$$= \frac{240\text{V}}{10\Omega}\left[\frac{1}{\pi}\left((\pi - ¼\pi)(2+\cos ½\pi) + \frac{3}{2}\sin ½\pi\right)\right]^{½} = 37.75\text{A}$$

$$V_{rms} = V\left[⅟\pi\left\{(\pi-\alpha) + ½\sin 2\alpha)\right\}\right]^{½}$$

$$= 240\text{V}\left[⅟\pi\left\{(\pi - ¼\pi) + ½\sin ½\pi)\right\}\right]^{½} = 228.8\text{V}$$

ii. Although only one thyristor conducts, which one that actually conducts may be random, thus both thyristor are rms rated for I_T = 37.75A. Whilst operational, the maximum thyristor voltage is √2 240 sin¼π, that is 240V. But before start-up or at shut-down, each thyristor has to block bi-directionally, √2 240 = 340V, peak.
The average thyristor (and supply and load) current is

$$\bar{I}_T = \frac{\sqrt{2}\ V}{\pi\omega L}\left[(\pi-\alpha)\cos\alpha + \sin\alpha\ \right]$$

$$= \frac{\sqrt{2}\ 240\text{V}}{\pi\times 10\Omega}\left[(\pi - ¼\pi)\cos ¼\pi + \sin ¼\pi\ \right] = 25.6\text{A}$$

♣

Example 13.1d: ***Single-phase ac regulator – 1 with ac back emf composite load***

A 230V 50Hz mains ac thyristor chopper has a load composed of 10Ω resistance in series with a138V 50Hz ac voltage source that leads the mains by 30°. If the thyristor triggering angle is 90° with respect to the ac mains, determine

i. The rms load current and maximum rms load current for any phase delay angle
ii. The power dissipated in the passive part of the load
iii. The thyristor average and rms current ratings and voltage ratings
iv. Power dissipated in the thyristors when modelled by $v_T = v_o + r_o \times i_T = 1.2 + 0.01 \times i_T$

Repeat the calculations if the passive part of the load is a 20mH inductor and the ac back emf lags the 50Hz ac mains by 30°.

Solution

ac back emf with a pure resistive load
From equation (13.48), the voltage across the resistive part of the load is

$$v_R = \sqrt{V^2 + V_{bac}^2 - 2VV_{bac}\cos\psi}$$
$$= \sqrt{230^2 + 138^2 - 2\times 230\times 138\times\cos 30} = 130.3\text{V}$$

with an angle of $\varphi = -32.8°$ with respect to the ac mains, given by $\psi = 30°$ and $V_{bac}/V = 138\text{V}/230\text{V} = 0.6$ in the fourth quadrant of figure 13.5. From the phasor diagram in figure 13.5, the thyristor firing angle with respect to the load resistor voltage is $\alpha_R = \alpha - \varphi = 90° - 32.8° = 57.8°$.

i. The load rms current is given by equation (13.24), that is

$$I_{rms} = \frac{V_R}{R}\sqrt{1 - \frac{2\alpha_R - \sin 2\alpha_R}{2\pi}}$$
$$= \frac{130.3\text{V}}{10\Omega}\sqrt{1 - \frac{57.8°}{180°} + \frac{\sin 2\times 57.8°}{2\pi}} = 13.03\text{A}\times 0.732 = 9.54\text{A}$$

The maximum rms load current is 13A when is $\alpha_R = 0$, that is when $\alpha = -\varphi = 32.8°$.

ii. The 10Ω resistor losses are

$$P_{10\Omega} = I_{rms}^2 \times 10\Omega$$
$$= 9.54^2 \times 10\Omega = 910.1\text{W}$$

iii. The thyristor current ratings are

$$I_{T\,rms} = I_{rms}/\sqrt{2}$$
$$= 16.83/\sqrt{2} = 11.9\text{A}$$

From equation (13.22), the average thyristor current is

$$\bar{I}_T = ½\frac{\sqrt{2}\,V_R}{\pi R}(1 + \cos\alpha_R)$$
$$= ½\frac{\sqrt{2}\times 130.3\text{V}}{\pi\times 10\Omega}(1 + \cos 57.8°) = 4.5\text{A}$$

The thyristors effectively experience a forward and reverse voltage associated with a single ac source of 130.3V ac. Without phase control the maximum thyristor voltage is √2×130.3V=184.3V. If the triggering angle α is less than 90°-φ=122.8 ° (with respect to the ac mains) then the maximum off-state voltage is less, namely

$$\hat{V}_T = \sqrt{2}\times 130.3\times\sin(\alpha - 32.8°) \text{ if } \alpha < 122.3°$$

iv. The power dissipated in each thyristor is

$$P_T = v_o\bar{I}_T + r_o I_{T\,rms}^2$$
$$= 1.2\text{V}\times 4.5\text{A} + 0.01\Omega\times 11.9^2 = 6.8\text{W}$$

ac back emf with a pure reactive load
The voltage across the inductive part of the load is the same as for the resistive case, namely 130.3V. In this case the ac back emf lags the ac mains. The phase angle with respect to the ac mains is $\varphi = 32.8°$, given by $\psi = -30°$ and $V_{bac}/V = 138\text{V}/230\text{V} = 0.6$ in the second quadrant of figure 13.5. Being a purely inductive load across the 130.3V ac voltage, the current lags this voltage by 90°. From the phasor

diagram in figure 13.5, the thyristor firing angle with respect to the load inductor voltage is $\alpha_L = \alpha + \varphi = 90° + 32.8° = 122.8°$. Since the effective delay angle α_L is greater than 90°, symmetrical, bipolar, discontinuous load current flows, as considered in section 13.1ii.

i. With a 20mH load inductor, the load rms current is given by equation (13.45), that is

$$I_{rms} = \frac{V_L}{X}\left[\frac{2}{\pi}\left((\pi - \alpha_L)(2 + \cos 2\alpha_L) + \frac{3}{2}\sin 2\alpha_L\right)\right]^{½}$$
$$= \frac{130.3\text{V}}{2\pi 50\text{Hz}\times 0.02\text{H}}\sqrt{\left(2 - \frac{122.8°}{90°}\right)(2 + \cos 2\times 122.8°) + \frac{3}{\pi}\sin 2\times 122.8°}$$
$$= 20.74\text{A}\times 0.373 = 7.73\text{A}$$

The maximum bipolar rms load current is when $\alpha_R = 90°$, $I_{rms} = 20.74$, and $\alpha = 90 - \varphi = 32.8°$.

ii. The 20mH inductor losses are zero.

iii. The thyristor current ratings are

$$I_{T\,rms} = I_{rms}/\sqrt{2}$$
$$= 7.73/\sqrt{2} = 5.47\text{A}$$

From equation (13.43), the average thyristor current is

$$\bar{I}_T = \frac{\sqrt{2}\,V_L}{\pi\omega L}\left[(\pi - \alpha_L)\cos\alpha_L + \sin\alpha_L\right]$$
$$= 20.74\text{A}\frac{\sqrt{2}}{\pi}\left[\pi\left(1 - \frac{122.8°}{180°}\right)\times\cos 122.8° + \sin 122.8°\right] = 2.80\text{A}$$

The thyristors effectively experience a forward and reverse voltage associated with a single ac source of 130.3V ac. Without phase control the maximum thyristor voltage is √2×130.3V=184.3V. Since $\alpha_R \geq 90°$ is necessary for continuous bipolar load current, 184.3V will always be experienced by the thyristors for any $\alpha_R > 90°$.

iv. The power dissipated in each thyristor is

$$P_T = v_o\bar{I}_T + r_o I_{T\,rms}^2$$
$$= 1.2\text{V}\times 2.8\text{A} + 0.01\Omega\times 5.47^2 = 3.66\text{W}$$

♣

13.1.2 Single-phase ac regulator – integral cycle control - line commutated

In thyristor heating applications, load harmonics are unimportant and integral cycle control, or burst firing, can be employed. Figure 13.6a shows the regulator when a triac is employed and figure 13.6b shows the output voltage indicating the regulator's operating principle. Because of the low frequency sub-harmonic nature of the output voltage, this type of control is not suitable for incandescent lighting loads since flickering would occur and with ac motors, undesirable torque pulsations would result.

In many heating applications the load thermal time constant is long (relative to 20ms, that is 50Hz) and an acceptable control method involves a number of mains cycles on and then off. Because turn-on occurs at zero voltage cross-over and turn-off occurs at zero current, which is near a zero voltage cross-over, supply harmonics and radio frequency interference are low. The lowest order harmonic in the load is $1/T_p$.

For a resistive load, the output voltage (and current) is defined by

$$v_o = i_o R = \sqrt{2}V\sin(\omega t) \quad \text{for } 0 \leq \omega t \leq 2\pi m$$
$$= 0 \quad \text{for } 2\pi m \leq \omega t \leq 2\pi N \qquad (13.51)$$

where $T_p = 2\pi N/\omega$.

The rms output voltage (and current) is

$$V_{rms} = \left(\frac{1}{2\pi}\int_0^{2\pi m/N}\left(\sqrt{2}V\sin N\omega t\right)^2 d\omega t\right)^{½} \qquad (13.52)$$
$$V_{rms} = I_{rms}R = V\sqrt{m/N} = V\sqrt{\delta} \quad \text{where the duty cycle } \delta = m/N$$

The Fourier coefficient and phase angle for each load voltage harmonic (for $n \neq N$) are given by

$$c_n = \sqrt{2}V \frac{2N}{\pi\left(N^2 - n^2\right)} \sin \pi n\delta$$

$$\phi_n = \pi\left(1 - n\delta\right) \quad \text{for} \quad n < N \qquad (13.53)$$

$$\phi_n = \pi\left(n\delta - 1\right) \quad \text{for} \quad n > N$$

When $n > N$ the harmonics are above $1/T_p$, while if $n < N$ subharmonics of $1/T_p$ are produced. For the case when $n = N$, the coefficient and phase angle for the sin πm term ($a_{n=N} = 0$) are

$$b_{n=N} = c_{n=N} = \sqrt{2}V\frac{m}{N} = \sqrt{2}V\delta \quad \text{and} \quad \phi_{n=N} = 0 \qquad (13.54)$$

Note the displacement angle between the ac supply voltage and the load voltage frequency component at the supply frequency, $n = N$, is $\phi_{n=N} = 0$. Therefore the fundamental power factor angle $\cos\phi_{n=N} = \cos 0 = 1$.

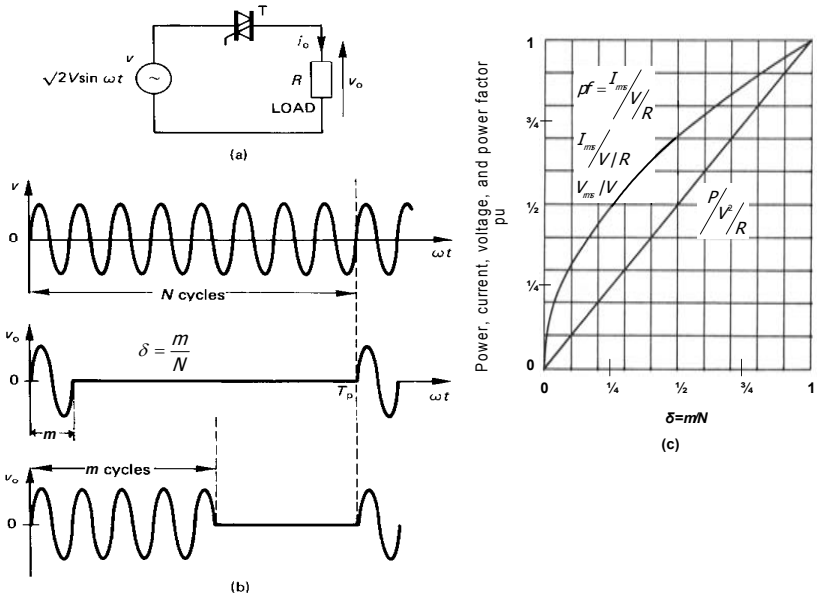


Figure 13.6. *Integral half-cycle single-phase ac control: (a) circuit connection using a triac; (b) output voltage waveforms for one-eighth maximum load power and nine-sixteenths maximum power; and (c) normalised supply power factor and power output.*

The output power is

$$P = \frac{m}{N}\frac{V^2}{R} = \delta \times \frac{V^2}{R} = I_{rms}^2 R \qquad \text{(W)} \qquad (13.55)$$

where n is the number of on cycles and N is the number of cycles in the period T_p.
The average and rms thyristors currents are, respectively,

$$\bar{I}_T = \frac{\sqrt{2}V}{\pi R}\frac{m}{N} = \frac{\sqrt{2}V}{\pi R}\delta \qquad I_{T_{rms}} = \frac{\sqrt{2}V}{2R}\sqrt{\frac{m}{N}} = \frac{\sqrt{2}V}{2R}\sqrt{\delta} \qquad (13.56)$$

From equation (13.54), the supply displacement factor $\cos\psi_{n=N}$ is unity and supply power factor λ is $\sqrt{m/N} = \sqrt{P/\hat{P}} = \sqrt{\delta}$. From $pf = \lambda = \mu\cos\phi_{n=N} = \mu$, the distortion factor μ is $\sqrt{m/N} = \sqrt{\delta}$. The rms voltage at the supply frequency is $V\,m/N = \delta V$ *and* the power transfer ratio is $m/N = \delta$. For a given percentage of maximum output power, the supply power factor is the same for integral cycle control and

phase angle control. The introduction of sub-harmonics tends to restrict this control technique to resistive heating type application. Temperature effects on load resistance R have been neglected, as have semiconductor on-state voltages. Finer resolution output voltage control is achievable if integral half-cycles are used rather than full cycles. The equations remain valid, but the start of multiplies of half cycles are alternately displaced by π so as to avoid a dc component in the supply and load currents. Multiple cycles need not be consecutive within each period.

Example 13.2: ***Integral cycle control***

The power delivered to a 12Ω resistive heating element is derived from an ideal sinusoidal supply $\sqrt{2}$ 240 sin 2π 50 t and is controlled by a series connected triac as shown in figure 13.6. The triac is controlled from its gate so as to deliver integral ac cycle pulses of three (m) consecutive ac cycles from four (N).

Calculate

i. The percentage power transferred compared to continuous ac operation
ii. The supply power factor, distortion factor, and displacement factor
iii. The supply frequency (50Hz) harmonic component voltage of the load voltage
iv. The triac maximum *di/dt* and *dv/dt* stresses
v. The phase angle α, to give the same load power when using phase angle control. Compare the maximum *di/dt* and *dv/dt* stresses with part iv.
vi. The output power steps when m, the number of conducted cycles is varied with respect to N = 4 cycles. Calculate the necessary phase control α equivalent for the same power output. Include the average and rms thyristor currents.
vii. What is the smallest power increment if half cycle control were to be used?
viii. Tabulate the harmonics and rms subharmonic component per unit magnitudes of the load voltage for m = 0, 1, 2, 3, 4; and for harmonics n = 0 to 12. (Hint: use Excel)

Solution

The key data is: $m = 3$ $N = 4$ (δ = ¾) V = 240 rms ac, 50Hz

i. The power transfer, given by equation (13.55), is

$$P = \frac{V^2}{R}\frac{m}{N} = \frac{V^2}{R}\delta = \frac{240^2}{12\Omega} \times ¾ = 4800 \times ¾ = 3.6\text{kW}$$

That is 75% of the maximum power is transferred to the load as heating losses.

ii. The displacement factor, cosψ, is 1. The distortion factor is given by

$$\mu = \sqrt{\frac{m}{N}} = \sqrt{\delta} = \sqrt{\frac{3}{4}} = 0.866$$

Thus the supply power factor, λ, is

$$\lambda = \mu\cos\psi = \sqrt{\frac{m}{N}} = \sqrt{\delta} = 0.866 \times 1 = 0.866$$

iii. The 50Hz rms component of the load voltage is given by

$$V_{50Hz} = V\frac{m}{N} = V\delta = 240 \times ¾ = 180\text{V rms}$$

iv. The maximum *di/dt* and *dv/dt* occur at zero cross over, when t = 0.

$$\left.\frac{dV_s}{dt}\right|_{\max} = \left.\frac{d}{dt}\sqrt{2}\,240\sin 2\pi 50t\right|_{t=0}$$

$$= \sqrt{2}\,240\,(2\pi 50)\cos 2\pi 50t\Big|_{t=0}$$

$$= \sqrt{2}\,240\,(2\pi 50) = 0.107\ \text{V/μs}$$

$$\left.\frac{d}{dt}\frac{V_s}{R}\right|_{\max} = \left.\frac{d}{dt}\frac{\sqrt{2}\,240}{12\Omega}\sin 2\pi 50t\right|_{t=0}$$

$$= \sqrt{2}\,20\,(2\pi 50)\cos 2\pi 50t\Big|_{t=0}$$

$$= \sqrt{2}\,20\,(2\pi 50) = 8.89\ \text{A/ms}$$

v. To develop the same load power, 3600W, with phase angle control, with a purely resistive load, implies that both methods must develop the same rms current and voltage, that is, $V_{rms} = \sqrt{R\,P} = V\sqrt{m/N} = V\sqrt{\delta}$. From equation (13.5), when the extinction angle, β = π, since the load is resistive

$$V_{rms} = \sqrt{R \times P} = V\sqrt{m/N} = V\sqrt{\delta} = V\left[\tfrac{1}{\pi}\left\{(\pi-\alpha)+\tfrac{1}{2}\sin 2\alpha\right\}\right]^{\frac{1}{2}}$$

that is

$$\delta = \frac{m}{N} = \tfrac{1}{\pi}\left\{(\pi-\alpha)+\tfrac{1}{2}\sin 2\alpha\right\} \quad = \tfrac{3}{4} = \tfrac{1}{\pi}\left\{(\pi-\alpha)+\tfrac{1}{2}\sin 2\alpha\right\}$$

Solving $0 = \tfrac{1}{4}\pi - \alpha + \tfrac{1}{2}\sin 2\alpha$ iteratively gives α = 63.9°.

When the triac turns on at α = 63.9°, the voltage across it drops virtually instantaneously from √2 240 sin 63.9 = 305V to zero. Since this is at triac turn-on, this very high *dv/dt* does not represent a turn-on *dv/dt* stress. The maximum triac *dv/dt* stress tending to turn it on is at zero voltage cross over, which is 107 V/ms, as with integral cycle control. Maximum *di/dt* occurs at triac turn on where the current rises from zero amperes to 305V/12Ω = 25.4A quickly. If the triac turns on in approximately 1µs, then this would represent a *di/dt* of 25.4A/µs. The triac initial *di/dt* rating would have to be in excess of 25.4A/µs.

cycles	period	duty	power	$\bar{I}_{Th}$	$I_{Th_{rms}}$	Delay angle	Displacement factor	Distortion factor	Power factor
m	*N*	*δ*	W	A	A	*α*	cos*ψ*	*μ*	*λ*
0	4	0	0	0	0	180°			
1	4	¼	1200	2.25	7.07	114°	1	½	½
2	4	½	2400	4.50	10.0	90°	1	0.707	0.707
3	4	¾	3600	6.75	12.2	63.9°	1	0.866	0.866
4	4	1	4800	9	14.1	0	1	1	1

vi. The output power can be varied using *m* = 0, 1, 2, 3, or 4 cycles of the mains. The output power in each case is calculated as in part 1 and the equivalent phase control angle, α, is calculated as in part v. The appropriate results are summarised in the table.

vii. Finer power step resolution can be attained if half cycle power pulses are used as in figure 13.6b. If one complete ac cycle corresponds to 1200W then by using half cycles, 600W power steps are possible. This results in nine different power levels if *N* = 4, from 0W to 4800W, in 600W steps.

vii. The following table show harmonic components, rms subharmonics, etc., for *N* = 4, (up to the twelfth) which are calculated as follows.

For $n \neq 4$, (that is not 50Hz) the harmonic magnitude is calculated from equation (13.53).

$$c_n = \frac{2\sqrt{2}\,V\,N}{\pi\left(N^2 - n^2\right)}\sin\left(\frac{\pi n m}{N}\right) = \frac{8\sqrt{2}\,V}{\pi\left(16 - n^2\right)} \times \sin\left(\frac{\pi n m}{4}\right) \quad \text{when } N = 4 \text{ and } n \neq 4$$

Equation (13.54) gives the 50Hz load component (n = 4).

$$c_{n=N=4} = \sqrt{2}\,V\frac{m}{N} = \sqrt{2}\,V\frac{m}{4} \quad \text{when } N = 4 \text{ and } n = 4$$

Component magnitudes (but not necessarily phase) are equal about δ=½ when *N* is even.
The rms output voltage is given by equation (13.52) or the square root of the sum of the squares of the harmonics, that is

$$V_{rms} = V\sqrt{m/N} = V\sqrt{\sum_{n=1}^{\infty} c_n^2}$$

The ac subharmonic component (that is components less than 50Hz) is given by

$$V_{ac,sub} = \sqrt{2}\;V\left[c_1^2 + c_2^2 + c_3^2\right]^{\frac{1}{2}}$$

From equations (13.52) and (13.54), the non fundamental (50Hz ac) component is given by

$$V_{ac} = \sqrt{V_{rms}^2 - V_{50Hz}^2} = V\sqrt{\frac{m}{N} - \left(\frac{m}{N}\right)^2} = V\sqrt{\delta(1-\delta)}$$

Normalised components		*δ* and *m* 0	¼	½	¾	1	*m/N*
n	Hz	0	1	2	3	4	
0	0	0	0	0	0	0	
1	12.5	0	0.120	0.170	0.120	0	
2	25	0	0.212	0	-0.212	0	
3	37.5	0	0.257	-0.364	0.257	0	*m/N*
4	50	0	¼	½	¾	1	fundamental
5	62.5	0	0.200	-0.283	0.200	0	
6	75	0	0.127	0	-0.127	0	
7	87.5	0	0.055	0.077	0.055	0	
8	100	0	0	0	0	0	
9	112.5	0	-0.028	-0.039	-0.028	0	
10	125	0	-0.030	0	0.030	0	
11	137.5	0	-0.017	0.024	-0.017	0	
12	150	0	0	0	0	0	
all *n*	sum square	0	0.249	0.499	0.749	1	$\sum_1^{12} c_n^2$
all *n*	rms	0	0.499	0.707	0.866	1	$\sqrt{\sum_1^{12} c_n^2}$
all *n* check	exact rms	0	0.5	0.707	0.866	1	$\sqrt{\frac{m}{4}} = \sqrt{\delta}$
all *n* but not *n* = 4	ac harmonic rms	0	0.432	0.499	0.432	0	$\sqrt{\sum_1^{12} c_n^2 - c_4^2}$
$n \leq 3$	sub harmonics rms	0	0.354	0.401	0.354	0	$\sqrt{\sum_1^{3} c_n^2}$
$n \geq 5$	upper harmonics rms	0	0.247	0.297	0.247	0	$\sqrt{\sum_5^{12} c_n^2}$
$n \geq 5$ check	upper harmonics rms	0	0.249	0.298	0.249	0	$\sqrt{\frac{m}{4} - \sum_1^{3} c_n^2 - \left(\frac{m}{4}\right)^2}$
pf = *λ*	power factor	0	½	0.707	0.866	1	$\sqrt{\frac{m}{4}} = \sqrt{\delta}$
	power pu	0	¼	½	¾	1	¼*m*

♣

13.1.3 The solid-state relay (SSR)

A *Solid-State Relay* (SSR) is a normally-open, electrical switch comprising solid-state semiconductors and/or electronic components that can be used in place of a mechanical relay to switch electricity to a load. An SSR offers enhanced electrical performance and reliability over any electro-mechanical relay alternative.
It is a totally electronic device that depends on the electrical, magnetic, and optical properties of semiconductors to control the flow of current in a circuit. It is normally comprised of a low current control side (equivalent to the coil on a mechanical relay) and a high-current load side (equivalent to the contacts of a conventional mechanical relay). SSRs typically feature electrical isolation up to several kilovolts between the control and load sides. Because of this isolation, the load side electronics of the relay is powered from the switched line, such that both the line voltage and a load (additionally to a control signal) must be present for the relay to operate.

An SSR contains one or more LEDs in the input (drive) section and provides optical coupling to a phototransistor or photodiode array, which in turn connects to driver circuitry that functions as an interface to the switching device or devices at the output. The power switching device is typically a MOSFET or TRIAC (usually two anti-parallel connected silicon controlled rectifiers, SCRs).

The advantages of solid-state relays versus their electro-mechanical counterparts are numerous:

- Higher reliability, reliable operation in harsh environments, longer life
- Elimination of switch bounce since no moving parts, giving a longer lifetime than electromechanical devices
- Smaller size
- Faster switching times, typically 100µs
- Elimination of HV arcing and pitting with reduced electromagnetic interference
- Lower triggering currents, that is, low power consumption, compared to electromechanical devices
- Robust packaging, resistant to shock and vibration, not dependent upon orientation
- No moving parts, thus silent operation

In order to utilize the benefits and flexibility of power MOSFET/IGBT technology, it may be required that sophisticated driver circuitry be used to drive these devices. MOSFET/IGBT based SSRs are more versatile than their TRIAC based counterparts. TRIAC based devices are limited only to ac load applications, while MOSFET/IGBT based devices can cater for ac and dc loads. Triac output SSRs are general-purpose relays typically used only for resistive type loads. SCR output solid-state relays are used to switch resistive or inductive loads, especially loads with high inrush currents. Thyristor technology is slower switching than MOSFET/IGBT SSR technology, but can handle higher voltage and current levels.

Thyristor technology based SSRs are specifically considered in this section, although many of the aspects considered are applicable to MOSFET/IGBT based SSR technology.

AC output solid-state relays are used to control the flow of electrical energy in alternating current power systems. The input control (equivalent to the electro-mechanical relay coil) voltages can be either ac or dc. The majority of solid-state relays require less power than electromechanical types to turn on and are readily interfaced. Another advantage of having no moving parts is that solid-state relays offer a fast response time with no contact bounce. For instantaneous turn-on types, the time between applying a command signal to the control circuit and the output circuit turning on is typically 20µs, although 100µs is usually quoted as a maximum. Alternatively, because of the nature of the electronic control circuitry, it is possible to delay the turn-on of a solid-state relay until the next voltage zero of the ac supply. Thus ac output solid-state relays can have two types of turn-on response: instantaneous (also known as phase control or random turn-on) and zero crossing.
Zero voltage turn on refers to a control circuit which after the presence of a control signal, only permits the relay's output to switch on load current if the ac line voltage is at or near a zero ac supply voltage point. Random turn on refers to a control circuit that energizes the relay's output irrespective of the value of the ac line voltage at the time of the turn-on command. The opto-coupler design and selection determine the zero or random function.

All ac output solid-state relays, which use SCRs or triacs as the output switch, after the removal of the control signal, will turn off at the next ac current zero. The relay may conduct for an additional half cycle of the ac supply frequency if the control signal is removed within 100µs of the next current zero. Because of the response time of solid-state relays, power to a load can be applied accurately and removed precisely. This is especially important when applications involve the switching of highly capacitive loads, and is a major advantage over electromechanical switching.

Zero-crossing relays are used with resistive loads while random turn-on relays are used with inductive loads, for example motors, transformers, coils, etc. Zero-crossing relays may also be used with inductive loads, but consideration must be given to the power factor of the load. If the load is too inductive, then the output of a zero-crossing relay may half-wave (half-wave rectification of the ac source), whence a random-fire SSR should be used in the application.

A random-fire SSR may also be used in resistive applications. Some applications require that the load only be energized for a portion of the ac cycle. This can be accomplished with a random SSR (due to its relatively fast turn-on time), given that an appropriate controller is used. However, the initial surge current will be higher due to the SSR switching power when the line voltage is closer to its peak.

The key basic element of a solid-state relay is the output switch, sometimes a triac but more often (and more reliably) back-to-back SCRs. This output switch is the key part of a solid-state relay, being the component that handles the power.

Circuit description
The basic SSR comprises a number of stages, from the control signal input through to the power output stage and its voltage transient protection, as shown by the functional block diagram in figure 13.7.

Input Circuit Commonly referred to as the 'primary', the input of an SSR may consist of a simple resistor in series with the optical-isolator, or of a more complex circuit with current regulation, reverse polarity protection, EMC filtering, etc. In either case, both serve the same function, which is to sense the application of a control signal commanding SSR turn on.
Optical Isolation The optical isolator in an SSR provides isolation between the input circuitry / control system, and the output circuit connected to the ac mains. The type of optical isolator used may also determine whether it is a zero-crossing or random-fire output.
Trigger Circuit This circuitry processes the input signal and switches the output state of the SSR. The trigger circuit may be internal or external to the optical-isolator.
Switching Circuit This is the portion of the SSR that switches the power to the load. It usually consists of an IGBT or MOSFET in a dc application, and a triac or parallel connected back-to-back SCRs in an ac application.
Protection Circuit Many applications require some form of electrical protection to prevent the SSR from being damaged in the application, or from misfiring due to environmental conditions. The protective devices (RC snubbers and transient voltage suppressors) may be incorporated into the design of an SSR, or mounted externally.

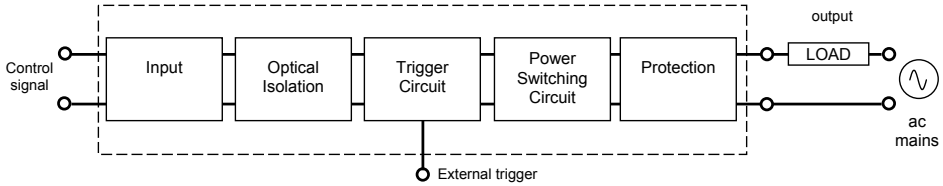


Figure 13.7. *Block diagram of the functional stages of a solid-state relay.*

13.1.3i Principle of operation
AC output solid-state relays are normally powered by the ac line, by connecting the two gates of the output SCRs through a controlled high voltage switch. In Figure 13.8a, when S1 is closed, the gates of SCR1 and SCR2 are connected, and current from the ac supply flows through either R1 or R2 into the gate of whichever SCR is forward biased; turning the SCR on and the relay conducts. While S1 is closed this action continues, reversing each half cycle of the ac supply and SCR1 and SCR2 conduct alternately. When S1 is opened, whichever SCR is conducting continues to conduct until reversed biased, when it turns off, and since the other SCR now has no gate current, the relay opens.

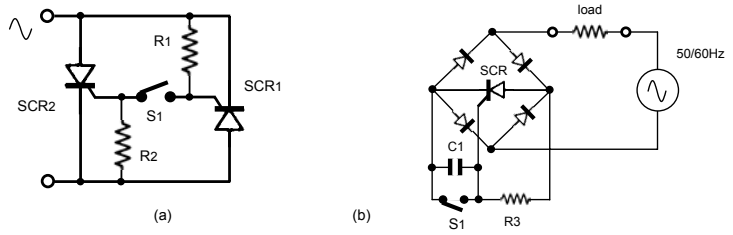


Figure 13.8. *SSR: (a) back-to-back SCR ac, normally-off output stage and (b) normally-on full-wave diode bridge version.*

Either of two circuits can provide the S1 switch function, with both offering optical isolation between the control and output of the solid-state relay. The circuit in Figure 31.9a uses an opto-triac as the isolating element, while the circuit in Figure 13.9b uses an opto-transistor as the isolating element. Each approach has specific features. Optically coupled SCRs are also used, with photovoltaic couplers used in MOSFET output dc relays.
The opto-transistor circuit in Figure 13.9b requires less control current to operate, conserving power, space and money in installations that use many relays. Another advantage of the opto-transistor approach is the flexibility offered to modify and tailor the control circuit characteristics in terms of a zero crossing voltage window, noise suppression, etc. The disadvantage of this approach is that it is more expensive.
The opto-triac circuit in figure 13.9a requires a higher control current for reliable operation, especially with inductive loads. Additionally, the control circuit characteristics are not accessible so is inflexible. With fewer components, this type of circuit is usually less expensive. With a modified version of the opto-transistor circuit of Figure 13.9b, a normally closed solid-state relay can be designed.

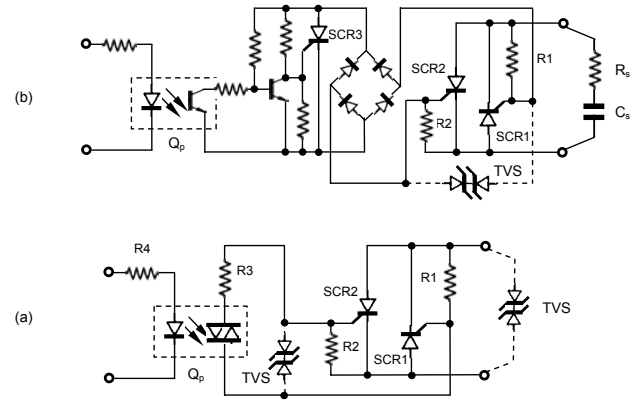


Figure 13.9. *Opto-coupled output stage and optional over-voltage protection, interfaced using an: (a) opto-triac and (b) opto-transistor.*

Solid-state relays can be either dc or ac voltage controlled, as shown in figure 13.10 parts a and b respectively. For ac input control, in figure 13.10b, the ac signal is rectified and filtered with capacitors to provide a dc signal to the opto-transistor or opto-triac LED in figure 13.9. The ac control versions can also be dc controlled by half-wave rectifying. AC input SSRs are slower to switch on due to the time it takes for the ac signal to increase in magnitude, to be rectified, and filtered to a useful dc current level for the opto-coupler input LED.

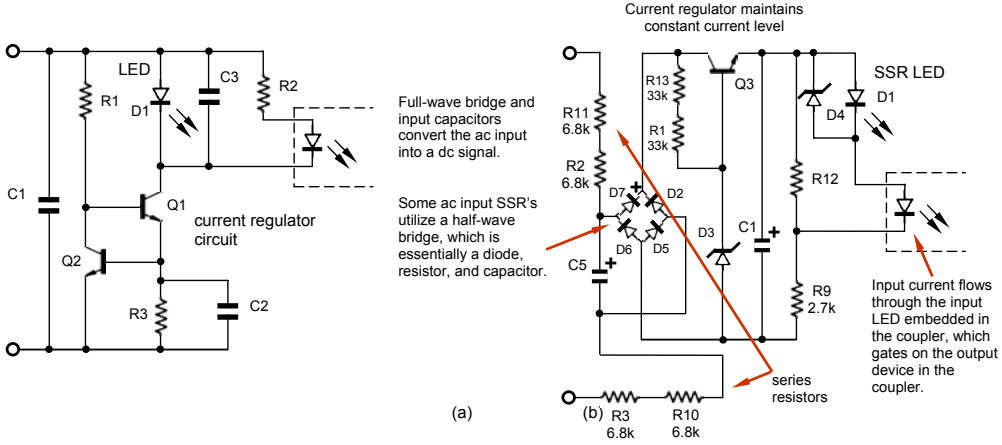


Figure 13.10. *SSR input control stage for: (a) regulated dc input circuit and (b) regulated ac input circuit control.*

13.1.3ii Key power elements in solid-state relays

In most power electronics cases, back-to-back SCRs are used as the output elements of ac output solid-state relays. The back-to-back SCR configuration shown in figure 13.8 has performance advantages when compared with triac outputs, notably higher *dv/dt*. Triacs have a *dv/dt* limitation when turning off: the commutating *dv/dt* of a triac is of the order of 10V/µs. Back-to-back SCRs do not have a commutating *dv/dt* limitation, just a critical *dv/dt* which is greater than 500V/µs. Using two output elements (anti-parallel connected, back-to-back SCRs) offers thermal benefits compared with a single triac as the heat dissipated is spread over a wider area of the SSR package substrate ceramic insulator.

Although the aluminium oxide ceramic substrate used to isolate the solid-state relay from the copper base plate is a good compromise as a thermal conductor, it has its limitations. The ceramic substrate does not efficiently conduct heat laterally, so by separating the heat source into two elements, more of the substrate is used to conduct the heat vertically through the thickness of the ceramic. Additionally, the two SCRs are attached to their own substantial copper bus-bar lead frames, which further help to spread the heat over a larger area of the ceramic substrate. Even with the extensive use of copper lead frames to spread the heat, the ceramic substrate is the dominant source of thermal impedance, contributing approximately 50% to the total thermal impedance of the relay, from SCR chip to relay copper base-plate.

The back-to-back SCR approach is preferred if the relay is subjected to surge currents, because each element is isolated from its partner both thermally and electrically (conduction 180° apart), unlike in a single element triac output. The control circuit determines turn-on and turn-off characteristics, but it is the output silicon switch that is the key to SCR performance.

The smaller the SCR die, the lower the cost, but this also results in poorer performance, where the surge (or overload) current is reduced, plus power dissipation and thermal impedance are increased. Reducing the thickness of the silicon can marginally increase surge current rating, the forward voltage drop and, therefore, power dissipation will also be reduced. However, with thinner die, the SCR chip manufacturer experiences yield penalties owing to increased wafer breakage and lower blocking voltage yield, resulting in higher overall manufacturing cost. The blocking voltage of thinner SCR dies will be lower, making the final solid-state relay significantly more susceptible to transient overvoltage damage. This is especially true if the thickness is reduced such that SCR break-over is not due to avalanche break-over but due to punch-through breakdown. If the silicon is susceptible to punch-through breakdown, which is more likely at sub-zero temperatures, any overvoltage will destroy the SCR die, whereas in avalanche breakdown, the SCR will normally self turn-on, conduct for the remainder of the half cycle and then return to its normal blocking condition, undamaged. Thicker silicon die will generally be more rugged relative to overvoltage transients and have a higher blocking voltage rating. The forward voltage drop will be higher, resulting in higher power dissipation and lower surge current capability. The overall die yield will be higher due to less wafer breakage and a higher useful voltage yield, resulting in lower manufacturing cost.

Optimized SCR design also involves considerations such as gate current and voltage to fire, holding current, latching current and *dv/dt* capability.

13.1.3iii Solid-state relay overvoltage fault modes

AC output solid-state relays operate in a wide variety of electrical environments. Some are benign with well-regulated and controlled ac supply lines, unlike others that are hostile with switching transients from a wide variety of sources. These transients can range from insignificant low-voltage, low-energy levels to high-voltage, high-energy pulses. Provided the amplitude of any line-borne voltage transients are below the rated transient voltage of the solid-state relay, safe reliable operation ensues.

For relays, generally, these transient voltage ratings are 600V peak for 240V rms rated relays and 1,200V peak for 480V rms rated relays. However, if a relay's transient voltage rating is exceeded, the relay may be damaged. Generally the relay will break over into uncontrolled conduction and recover at the next current zero with no damage. In other cases, depending on the frequency of break-over and the voltage capability of the various semiconductor elements that are exposed to the transient, the relay can be permanently damaged.

Two control circuits used in the ac output of solid-state relays are shown in Figure 13.9.

In the circuit in Figure 13.9b, seven elements are exposed to the ac line voltage: two output SCRs, four diodes in the bridge rectifier and the pilot SCR. Of these seven elements, if the lowest break down voltage is the reverse voltage of one of the output SCRs or a bridge diode, any overvoltage transient will permanently damage this element. The result will be a solid-state relay permanently on. If the lowest voltage breakdown voltage is the forward blocking voltage of any of the three SCRs, the SSR will likely break over into conduction without damage. The relay will conduct until the next current zero, then turn off and continue to operate normally. In practice, the lowest voltage breakdown element is normally the pilot SCR (SCR3), which is only forward biased, so it breaks over into conduction and turns on an output SCR (SCR1 or SCR2) through its gate, which is the normal turn-on mechanism for the main SCRs. The relay conducts until the next current zero and then regains control.

In Figure 13.10a, three elements are exposed to the ac line voltage: the output SCRs (SCR1 and SCR2) and the opto-triac Q_p. If the lowest breakdown voltage level is the reverse voltage of one of the output SCRs, then any over voltage transient will damage it and the relay output will be permanently on. If the lowest breakdown voltage is the forward voltage of one of the output SCRs, it will likely break over into conduction without damage. The relay will conduct until the next current zero, turn off and function normally. If the lowest breakdown element is the opto-triac, Q_p, it will likely break over into conduction, turning on an output SCR (SCR1 or SCR2) through its gate, which is the normal turn-on sequence for the main SCRs. The break-over current through Q_p is limited by the series resistor R3. If, however, Q_p is repeatedly broken over into conduction, it will eventually fail, resulting in the relay being permanently on.

13.1.3iv Standard transient voltage protection devices, reviewed in terms of SSR requirements

MOVs and transorbs are commonly used transient voltage protection devices, in both ac and dc circuits. See Chapter 10.4.

- A MOV is a metal oxide varistor which is made from a metal oxide material, typically zinc oxide, and it dissipates energy in the grain structure of the device. It switches into conduction relatively slowly, but can dissipate large levels of energy.
- A Transient Voltage Suppressor, TVS, or transorb is made from traditional semiconductor silicon material and switches very fast. It can handle large amounts of power, but only for short periods (low energy). It is like a Zener diode and it is used to protect electronic devices from transients. A Transorb functions by clamping any excessive voltage to a specific limit. It does this by conducting when excessive voltage is impressed across its two terminals. For a short time, the transorb absorbs high power. The transorb also reacts within a few nanoseconds making it superior to any traditional transient protection solution.

An MOV can be used to protect a SSR from transients, which can cause the SSR to turn-on without the control voltage applied. An MOV is typically placed externally, in parallel with the output of the SSR, as shown in figure 13.9a. After a MOV reaches its life expectancy, it typically fails shorted. By contrast, a transorb typically does not fail shorted, rather fails open, and is normally installed within the SSR.

MOVs are widely used to protect voltage-sensitive elements from overvoltage transients. The MOV has a voltage dependant resistance so that as the voltage increases, its impedance changes from a high resistance to a lower value at some specified voltage. The slope resistance of the characteristic is high, so that in the event of a relatively high current pulse, it can be difficult to coordinate and discriminate the MOV clamping voltage, the protected device voltage rating, and the system operating voltage. However, MOVs are available in a variety of energy absorbing sizes (joule rating) so that a suitable compromise can be made.
A significant MOV feature is its wear-out mechanism: every time it absorbs transient energy, its characteristics are changed, normally by the clamping voltage being reduced. Obviously if the MOV clamping voltage degrades to the point where it overlaps the supply voltage, then it will overheat and create a potentially hazardous over-temperature condition. For this reason, MOVs are usually oversized and used with caution.

Where transient energy levels are known and are of limited magnitude, the better option is to use TVS clamping diodes or break-over diodes. Both are silicon semiconductor devices with no deterioration mechanism but in modest sizes and costs, they cannot absorb the same amount of energy as an equivalent MOV. The TVS diode exhibits high impedance until its clamping voltage is reached and then essentially goes into a controlled avalanche mode with a low slope resistance. The break-over (or crowbar) diode exhibits a high impedance until its break-over voltage is reached and then breaks down to a low impedance (hence low voltage), thus protecting against high voltage transients.

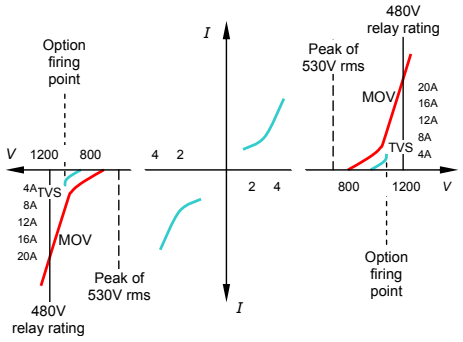


Figure 13.11. *Voltage-Current characteristic curves for MOVs, break-over diodes, and TVS's.*

Figure 13.11 shows the various voltage levels involved in the protection of a 480V ac solid-state relay, which has a maximum rms voltage rating of 530V rms, resulting in a peak voltage of approximately 750V, as shown on Figure 13.11. Solid-state relays rated at 480V rms are normally tested at three times this level, 1,200V peak, shown as the '480V relay rating' in Figure 13.11. A typical MOV clamping curve is shown, starting to clamp at about 900V but with a relatively high resistive component taking the clamping voltage above the relay rating if the voltage transient can supply about 21A peak.
Also shown in Figure 13.11, 1100V is the typical breakdown voltage of the type of TVS that is used to protect a relay. The current in the internal TVS will never reach the 4A shown, indicating the much lower slope resistance of TVS protectors. In most cases, the internal TVS diode will trigger the relay before the TVS current reaches 100mA. Although this energy level may appear high, it is spread over a number of series TVS elements, which yield excellent thermal management and dissipation paths.
Figure 13.9 shows the SSR output stage protected by an external parallel-connected *RC* snubber circuit. An RC snubber is more effective for *dv/dt* suppression than for transient over-voltage protection. The snubber losses, CV^2, are higher than those in MOV type protection, since the RC snubber is active during normal ac operation, as opposed to only being functional during a transient over-voltage situation in the MOV case. The energy associated with the MOV capacitance continually being cycled, limits its upper operating frequency.

13.1.3v Solid-state relay internal protection methods

Because of the problems associated with MOVs, particularly the possibility of over heating and lifetime clamping voltage downward drift, solid-state relays tend not to be supplied with internal MOVs.
Certain applications may employ TVS diodes (as shown by the dotted line connections in the two parts of Figure 13.9), connected to the gates of the output SCRs, which will conduct at voltages above maximum line voltage and below the voltage rating of the solid-state relay. If a transient voltage occurs, the TVS diodes conduct and normal gate current flows into the gates of the output SCRs, and the forward biased SCR turns on. This is the standard turn-on sequence for the solid-state relay output SCRs; they are turned on by the normal injection of gate current and no component is overstressed.

This internal overvoltage protection is available on specific solid-state relays.

- Generally, for resistive loads TVS over-voltage protection is suitable, provided the load can tolerate the transfer of the transient to the load.
- For capacitive loads, it is inadvisable to use internal TVS protection as this could lead to high inrush (surge) currents and possibly latent or catastrophic *di/dt* failures.
- In some motor start and stop applications TVS protection may be suitable depending on the motor load, the effects of a sudden, undesired small movement of the motor, etc. This form of protection cannot be used with motor reversing applications. This poses the significant danger of two relays (for forward and reverse operation) being turned on by a transient, resulting in a line-to-line short that damages the relays and other circuit elements.

13.1.3vi Application considerations

Different applications require different solid-state relay characteristics. Generally, the two turn-on methods, zero-crossing (for minimum EMC) or instantaneous, have specific application areas. However, there are general guidelines governing when either should be used, or not used.
If the load requires proportional control every ac half cycle (such as incandescent lamp dimming or low thermal mass temperature control), the instantaneous turn-on type is used. For high thermal mass loads, a zero-crossing relay with complete cycles of conduction and non-conduction is usually the preferred method of temperature control. A zero-crossing relay is generally used for inductive loads. However, for these types of loads a random turn-on type should always be considered. Under certain low load current and low power factor conditions it is possible for a zero-crossing relay to conduct only on every other half cycle (half-waving). This is caused by the relay terminal voltage rising so rapidly through the zero cross voltage window (at the lagging current zero) that the relay control circuit does not have time to react and so is locked off until the next voltage zero. With a random turn-on SSR type, there is no zero crossing window so no possibility of the relay half-waving. If in doubt, use a random turn-on relay for inductive loads.
Minimum load current is the least conducting current that the SSR will switch on and continue to carry with a nominal output voltage drop. Load currents less than this value, typically 50 to 100mA, may not be switched by the SSR.

The most common failure mode of an SSR is a shorted output, either half-wave or fully shorted, caused by excess load current flow or over temperature. The most common end of life failure is an open circuit as a result of thermal fatigue of internal solder joints and substrate.
In general, the best means to avoid such failures or prolong the life of an SSR is to operate at the lowest possible temperature and avoid large temperature excursions. Applications that have repetitive current surges should employ a higher current rated SSR to accommodate the heating caused by the surges.

Example 13.3: ***Solid-state relay turn-on***

Calculate the expected turn-on (trigger) voltage of an ac output SSR using SCRs and a trigger circuit, as in figure 13.9a, with the following parameters:

I_{gt} = 50mA, for both SCRs (minimum gate current for turn-on at 25°C)
V_{gt} = 0.7V for each SCR (minimum gate voltage for turn-on at 25°C)
Single opto-coupler with 1.0V V_f drop
Trigger circuit impedance of 68Ω.
No R_{gk} resistor.

Solution

With the aid of figure 13.9a:
Trigger circuit drop: 50mA x 68Ω= 3.4V.
SCR gate drops: 0.7V x 2 = 1.4V.
The expected turn-on (trigger) voltage: 3.4V + 1.4V + 1.0V coupler drop = 5.8V.

This turn on or trigger voltage constitutes the lower value of the zero turn-on 'window'.

♣

Example 13.4: ***Solid-state relay heatsink requirements***

A solid-state relay carries 50A in an application with a forward voltage drop V_f of 1.1V pk, resulting in 55W of power being dissipated. The ambient temperature is 35ºC, giving a 45ºC difference between ambient and the maximum recommended base plate temperature of 80ºC.
What is the heat-sinking requirement?

Solution

Division of the 45ºC temperature differential by the 55W of power being dissipated, results in a 0.82ºC/W heat sink requirement for the application. Prudently, 0.1ºC/W is deduced from the result to account for the thermal compound used in the assembly. Therefore, a 0.82ºC/W heat sink less 0.1ºC/W is 0.72ºC/W.
The heat sink needed requires a thermal resistance of no more than 0.7ºC/W.

♣

13.1.3vii DC output solid-state relays

DC output SSRs rated to 400V dc are usually MOSFET output based, while 1000V dc, 25A SSRs have an IGBT output stage. Both dc SSR types are usually dc input controlled. DC output solid-state relays are used for switching dc since, unlike dc electromechanical relays, there are no moving parts, hence no contact arcing or wear-out mechanism. However, there are some precautions, which have to be assessed when using dc output solid-state relays with inductive dc loads.

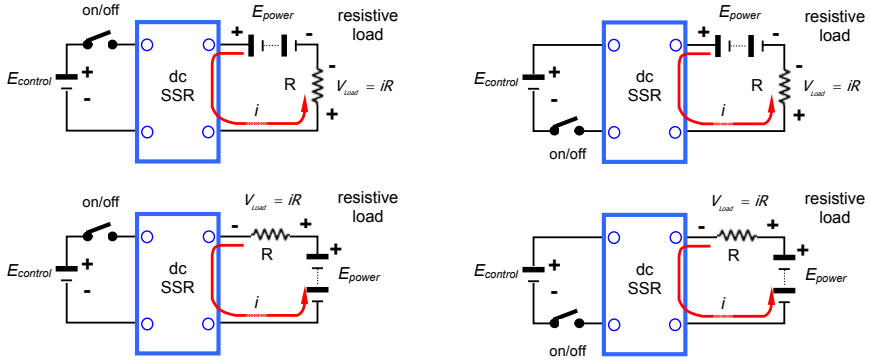


Figure 13.12. *Control and load connection possibilities for dc output solid-state relays.*

Relay Connections

Because the input and output terminals of dc solid-state relays are electrically isolated by up to 3,750V rms, the relative electrical potentials of the input signal and load connections are irrelevant. As shown in Figure 13.12, the input can be supplied either from a source or sink configuration and the load can be connected to either the relay positive or negative output terminals. The solid-state relay can be used as a level shifter because no electrical relationship is required between the input control and power output sections of the relay.

Inductive Load Considerations

When the dc load is inductive, precautions have to be taken to protect the solid-state relay at turn off. Energy is stored in the magnetic flux created by the current flowing through the inductive load. When the solid-state relay is turned off the collapse of the magnetic flux $d\Phi/dt$ creates an electro-magnetic force with a polarity that tries to maintain the pre-existing current flow. This is shown schematically in Figure 13.13.

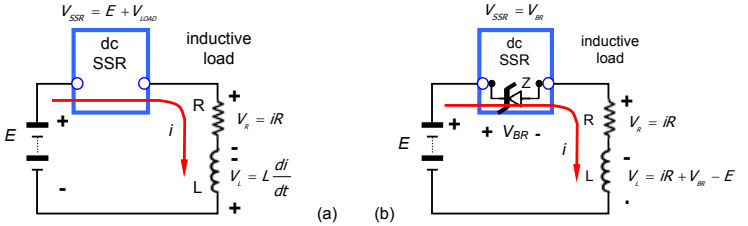


Figure 13.13. *Inductive loads with dc solid-state relays: (a) circuit connection and (b) induced voltages at turn-off.*

If no electrical path is provided for the inductive load current to flow, the rapid collapse of magnetic flux will generate a voltage high enough to break-over any limiting voltage element in the output load circuit. One element is usually the solid-state relay, shown schematically as V_{BR}. In the case of a solid-state relay, either the output power semiconductor device or one of the driving semiconductors will break-over into conduction, which may permanently damage the semiconductor resulting in a relay with a permanently shorted output. In most dc circuits, a circulating path for the inductive current can be created by the addition of a freewheel diode as shown in Figure 13.14a.
Unless the solid-state relay is to be turned on while current is still flowing in the freewheel diode, the diode can be a standard recovery type. If, however, the solid-state relay may need to turn on before the load current has been completely decayed to zero, then a fast recovery diode must be used in the freewheel position. The use of a fast recovery diode reduces the instantaneous reverse recovery inrush current amplitude and duration when the relay is turned on into an existing freewheel diode current.

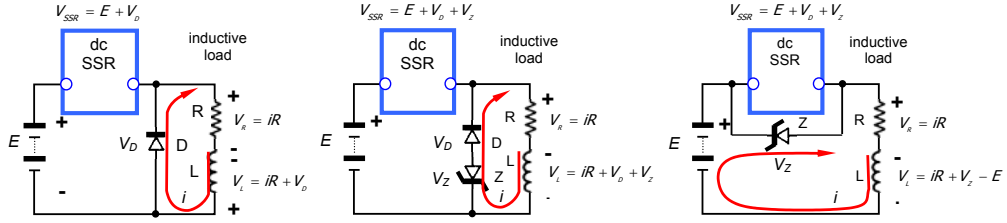


Figure 13.14. *The dc output SSR with an inductive load incorporating: (a) a load freewheel diode; (b) load diode/Zener diode combination; and (c) a Zener diode across the dc SSR output alternative.*

Generally, it is necessary to collapse the inductive current rapidly, for example, to open a solenoid as quickly as possible. If rapid current discharge is a requirement, then, the discharge path must be designed with a high voltage generated. This follows from the fact $E = L di/dt$, where E is the voltage generated by the collapse of magnetic flux, L is the inductance of the load and di/dt is the rate of change of current. The greater the value of E , for a fixed value of load L, then the greater the di/dt and the more rapidly the load current is reduced to zero. Increasing the voltage, which has to be generated to create a freewheel path, can be accomplished by adding to the diode shown in Figure 13.14a, a series Zener or

TVS diode as shown in Figure 13.14b. The voltage appearing across the solid-state relay is the sum of the supply voltage and the Zener or TVS diode voltage. So for a system using a dc supply of E and a solid-state relay voltage rated at V_R, the voltage across the freewheel components cannot be greater than $V_{BR} - E$ at maximum load current.
An alternate method to create the same inductor voltage is to connect the Zener or TVS diode externally across the output terminals of the solid-state relay, as shown in figure 13.14c (and internally in figure 13.13b), in this case the clamping voltage can be V_{BR}, the voltage rating of the dc SSR. As this voltage is in series with the dc supply E, the net result is the same as putting the clamping component across the load, a voltage of $V_{BR} - E$ is generated by the inductive load. The energy loss in the clamping device is greater in this case, since the freewheel path involves energy being drawn from the dc source. However, importantly, the voltage clamping protection is directly across the primary component to be protected, viz., the SSR, and even better if placed inside the SSR module as shown in figure 13.14b.

13.2 Single-phase transformer tap-changer – line commutated

Figure 13.15 shows a single-phase tap changer using two ac output solid-state relays, where the tapped ac voltage supply can be provided by a tapped transformer or autotransformer.
Thyristor T_3 (T_4) is triggered at zero voltage cross-over (or later), subsequently under phase control T_1 (T_2) is turned on. The output voltage (and current) for a resistive load R is defined by

$$v_o(\omega t) = i_o(\omega t) \times R = \sqrt{2}\, V_2 \sin \omega t \quad \text{(V)} \qquad \text{for } 0 \le \omega t \le \alpha \quad \text{(rad)} \tag{13.57}$$

$$v_o(\omega t) = i_o(\omega t) \times R = \sqrt{2}\, V_1 \sin \omega t \quad \text{(V)} \qquad \text{for } \alpha \le \omega t \le \pi \quad \text{(rad)} \tag{13.58}$$

where α is the phase delay angle and $v_2 < v_1$.
If $0 \le \delta = V_2/V_1 \le 1$, then for a resistive load the rms output voltage is

$$V_{rms} = \left[\frac{V_2^2}{\pi}\left(\alpha - \tfrac{1}{2}\sin 2\alpha\right) + \frac{V_1^2}{\pi}\left(\pi - \alpha + \tfrac{1}{2}\sin 2\alpha\right)\right]^{\frac{1}{2}} = V_1\left[\left\{1 - \frac{1}{\pi}\left(1-\delta^2\right)\left(\alpha - \tfrac{1}{2}\sin 2\alpha\right)\right\}\right]^{\frac{1}{2}} \tag{13.59}$$

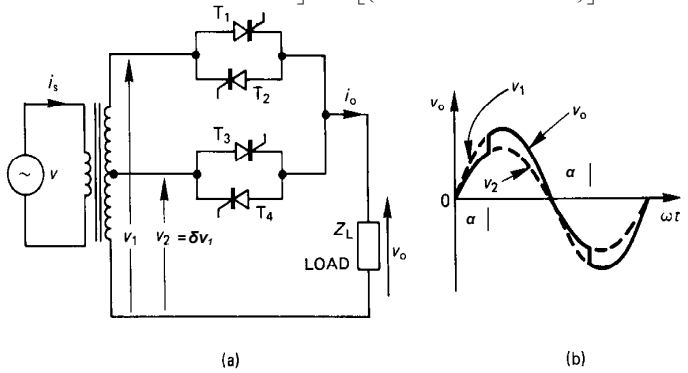


Figure 13.15. *An ac voltage regulator using a tapped transformer: (a) circuit connection and (b) output voltage waveform with a resistive load.*

The Fourier coefficients of the output voltage, which has only odd harmonics, are

$$a_n = V\frac{2}{\pi}\frac{1-\delta}{1-n^2}\left[\cos\alpha\cos n\alpha + n\sin\alpha\sin n\alpha - 1\right]$$
$$b_n = V\frac{2}{\pi}\frac{1-\delta}{1-n^2}\left(\cos\alpha\sin n\alpha - n\sin\alpha\cos n\alpha\right) \tag{13.60}$$

The amplitude of the fundamental quadrature components, $n = 1$, are

$$a_1 = V\frac{1}{\pi}(1-\delta)\sin^2\alpha$$
$$b_1 = V\frac{1}{\pi}(1-\delta)\left(\alpha\sin^2\alpha - \cos^2\alpha - \sin\alpha\cos\alpha\right) \tag{13.61}$$

Initially v_2 is impressed across the load, via T_3 (T_4). Turning on T_1 (T_2) reverse-biases T_3 (T_4), hence T_3 (T_4) turns off and the load voltage jumps to v_1. It is possible to vary the rms load voltage between v_2 and v_1. It is important that T_1 (T_2) and T_4 (T_3) do not conduct simultaneously, since such conduction short-circuits the transformer secondary.

Both load current and voltage information (specifically zero crossing) is necessary with inductive and capacitive loads, if winding short circuiting is to be avoided.
With an inductive load circuit, when only T_1 and T_2 conduct, the output current is

$$i_o = \frac{\sqrt{2}\,V}{Z}\sin(\omega t - \phi) \quad \text{(A)} \tag{13.62}$$

where $Z = \sqrt{R^2 + (\omega L)^2}$ (ohms) $\qquad \phi = \tan^{-1}\omega L/R$ (rad)

It is important that T_3 and T_4 are not fired until $\alpha \ge \phi$, when the load current must have reached zero. Otherwise a transformer secondary short circuit occurs through T_1 (T_2) and T_4 (T_3).
For a resistive load, the thyristor rms currents for T_3, T_4 and T_1, T_2 respectively are

$$I_{Trms} = \frac{V_2}{\sqrt{2}\,R}\sqrt{\frac{1}{\pi}\left(2\alpha - \sin 2\alpha\right)}$$
$$I_{Trms} = \frac{V_1}{\sqrt{2}\,R}\sqrt{\frac{1}{\pi}\left(\sin 2\alpha - 2\alpha\right) + 2\pi} \tag{13.63}$$

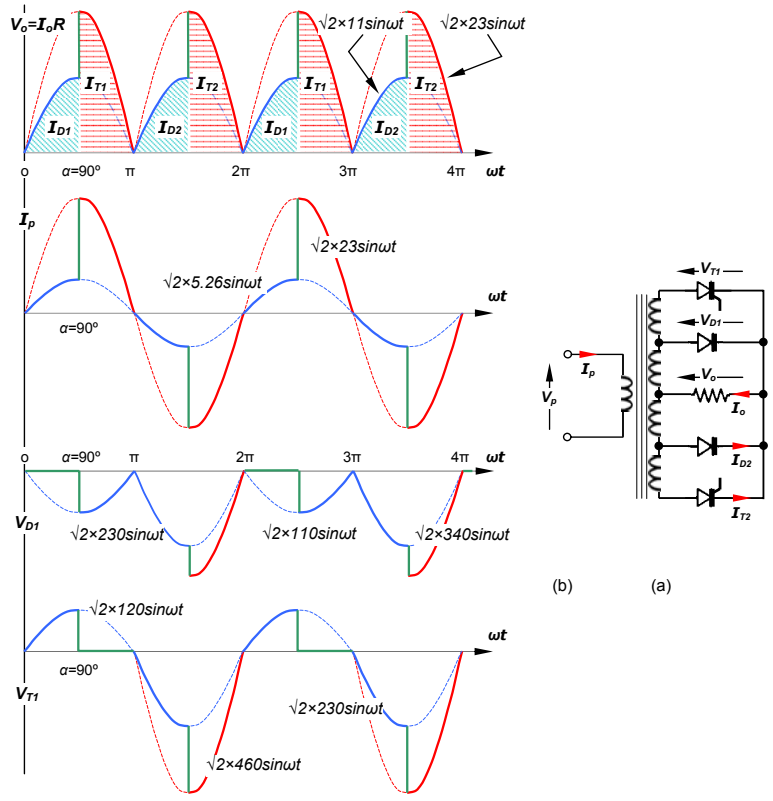


Figure 13.16. *An ac voltage regulator using a tapped transformer connected as a rectifier with a resistive load: (a) circuit diagram and symbols and (b) circuit waveforms, viz., output voltage and current, transformer primary current, diode reverse blocking voltage, and thyristor blocking voltages.*

The thyristor voltages ratings are both v_1 - v_2, provided a thyristor is always conducting at any instant.
An extension of the basic operating principle is to use phase control on thyristors T_3 and T_4 as well as T_1 and T_2. It is also possible to use tap-changing in the primary circuit. The basic principle can also be extended from a single tap secondary to a multi-tap transformer.

The basic operating principle of any multi-output tap changer, in order to avoid short circuits, independent of the load power factor is

- switch up in voltage when the load V and I have the same direction, delivering power
- switch down when V and I have the opposite direction, returning power.

Example 13.5: ***Tap changing converter***

The converter circuit shown in figure 13.16 is a form of ac to dc tap changer, with a 230V ac primary. The inner voltage taps can deliver 110V ac while the outer tap develops 230V ac across the 10Ω resistive load. If the thyristor phase delay angle is 90º determine

i. The mean load voltage hence mean load current
ii. The average diode and thyristor current
iii. The primary rms current
iv. The peak thyristor and diode voltage, for any phase angle

Solution

The output voltage is similar to that shown in figure 13.15b, except rectified, and α = 90º.

i. The mean load voltage can be determine from equation (13.22)

$$V_o = \frac{\alpha}{\pi}\frac{\sqrt{2}V_{110}}{\pi}(1+\cos 0°) + \frac{\sqrt{2}V_{230}}{\pi}(1+\cos\alpha)$$
$$= \frac{½\pi}{\pi}\frac{\sqrt{2}\,110V}{\pi}(1+\cos 0°) + \frac{\sqrt{2}\,230V}{\pi}(1+\cos 90°)$$
$$= 49.5V + 103.5V = 153V$$

whence $\bar{I}_o = \frac{V_o}{R} = \frac{153V}{10\Omega} = 15.3A$

ii. The diode current is associated with the 49.5V component of the average load voltage, while the thyristor component is 103.5V. Taking into account that each semiconductor has a maximum duty cycle of 50%:

The average diode current is

$$\bar{I}_D = 50\% \times \frac{49.5V}{10\Omega} = 2.475A$$

The average thyristor current is

$$\bar{I}_T = 50\% \times \frac{103.5V}{10\Omega} = 5.175A$$

iii. The primary rms current has two components.

When the diode conducts the primary current is

$$i_{p1} = \sqrt{2}\frac{110V}{230V} \times \frac{110V}{10\Omega}\sin\omega t = \sqrt{2}\times 5.26\sin\omega t \qquad 0 \le \omega t \le \alpha$$

$$i_{p2} = \sqrt{2}\frac{230V}{230V} \times \frac{230V}{10\Omega}\sin\omega t = \sqrt{2}\times 23\sin\omega t \qquad \alpha \le \omega t \le \pi$$

The rms of each component, on the primary side, is

$$I_{rmsD} = \left[½\pi\int_0^{\alpha}\left(\sqrt{2}\times 5.26\right)^2 \sin^2\omega t\, d\omega t\right]^{½}$$
$$= \sqrt{2}\times 5.26\left[¼\pi\{\alpha - ½\sin 2\alpha)\}\right]^{½} = \sqrt{2}\times 5.26\left[\frac{1}{2\sqrt{2}}\right] = 2.63A$$

$$I_{rmsT} = \left[½\pi\int_{\alpha}^{\pi}\left(\sqrt{2}\times 23\right)^2 \sin^2\omega t\, d\omega t\right]^{½}$$
$$= \sqrt{2}\times 23\left[¼\pi\{(\pi-\alpha) + ½\sin 2\alpha)\}\right]^{½} = \sqrt{2}\times 23\left[\frac{1}{2\sqrt{2}}\right] = 11.5A$$

The total supply side rms current comprised the contribution of two diodes and two thyristors

$$I_p = \sqrt{I_{rmsD}^2 + I_{rmsD}^2 + I_{rmsT}^2 + I_{rmsT}^2}$$
$$= \sqrt{2.63^2 + 2.63^2 + 11.5^2 + 11.5^2} = 16.68A$$

iv. The peak diode voltage is associated with the turn-on of the thyristor associated with the other half cycle of the supply and worst case is when α<½π.

$$\hat{V}_D = \sqrt{2}\times 230V + \sqrt{2}\times 110V = 466.7V$$

and 466.7×sinα for α>½π.

The thyristor peak forward and reverse voltages are experienced at α=½π:

$$\hat{V}_T^F = \sqrt{2}\times(230V - 110V) = 169.7V$$
$$\hat{V}_T^R = 2\times\sqrt{2}\times 230V = 650.5V$$

The thyristor forward voltage is controlled by its associated diode and is less than 169.7V if α<½π, viz. 169.7×sinα. The peak reverse voltage of 650.5V is experienced if the complementary thyristor is turned on before α=½π, otherwise the maximum is 650.5×sinα for α>½π.

♣

13.3 Single-phase ac chopper regulator – commutable switches

An ac step-down chopper is shown in figure 13.17a. The switches T_1 and T_3 (shown as reverse blocking IGBTs) impress the ac supply across the load while T_2 and T_4 provide load current freewheel paths when the main switches T_1 and T_3 are turned off. In order to prevent the supply being shorted, switches T_1 and T_4 can not be on simultaneously when the ac supply is in a positive half cycle, while T_2 and T_3 can not both be on during a negative half cycle of the ac supply. Zero voltage information is necessary.
If the rms supply voltage is V and the on-state duty cycle of T_1 and T_3 is δ, then rms output voltage V_o is

$$V_o = \delta V \qquad (13.64)$$

When the sinusoidal supply is modulated by a high frequency rectangular-wave carrier ω_s ($2\pi f_s$), which is the switching frequency, the ac output is at the same frequency as the supply f_o but the fundamental magnitude is proportional to the rectangular wave duty cycle δ, as shown in figure 13.17b. Being based on a modulation technique, the output harmonics involve the fundamental at the supply frequency f_o and components related to the high frequency rectangular carrier waveform f_s. The output voltage is given by

$$V_o = \sqrt{2}\delta V\sin\omega_o t + \sum_{\forall n>1}\left[\frac{\sqrt{2}V}{n}\sin n\delta\left\{\sin(\omega_o + n\omega_c)t - \sin(\omega_o - n\omega_c)t\right\}\right] \qquad (13.65)$$

The carrier (switching frequency) components can be filtered by using an output *L*-*C* filter, as shown in figure 13.17a, which has a cut-off frequency of $f_{½}$ complying with $f_o < f_{½} < f_s$.

Rather than using a variable duty cycle to control the output magnitude, selective harmonic elimination, SHE, can be used, where the switches are commutated at pre-calculated angles so as to eliminate specific harmonics, and control the fundamental magnitude of the output voltage. T_1/T_3 are turned on at switching angles α_1, α_3, ….α_{M-1} and turned off at α_2, α_4, …α_M per quarter cycle.
The quarter-wave symmetry results in null even harmonics, including any dc component. By the proper choice of PWM switching angles, α_i, the fundamental component can be controlled and selected low order harmonics can be eliminated. The Fourier series for the output voltage, which is expressed in terms of the M switching point variables per quarter cycle, is:

$$v_o = \sqrt{2}V\sum_{n=1,3,5....}^{\infty} a_n\sin\omega t$$

where the value of a_n is:

$$a_n = \frac{2}{\pi}\sum_{j=1}^{M}(-1)^j\left[\frac{\sin(n-1)\alpha_j}{(n-1)} - \frac{\sin(n+1)\alpha_j}{(n+1)}\right]$$

where n = 3, 5,..., $2M-1$, M is the number of switching angles per quarter cycle, α_l is the l^{th} switching angles and $\sqrt{2}V$ is the maximum value of the input voltage. The rms fundamental component is given by:

$$V_{a_1} = V\left(1+\frac{2}{\pi}\right)\sum_{n=3,5,7..}^{\infty}(-1)^l\left[\alpha_l - \tfrac{1}{2}\sin 2\alpha_l\right]$$

The commutation angle selection to eliminate certain harmonics is addressed in Chapter 15.3.4viii.

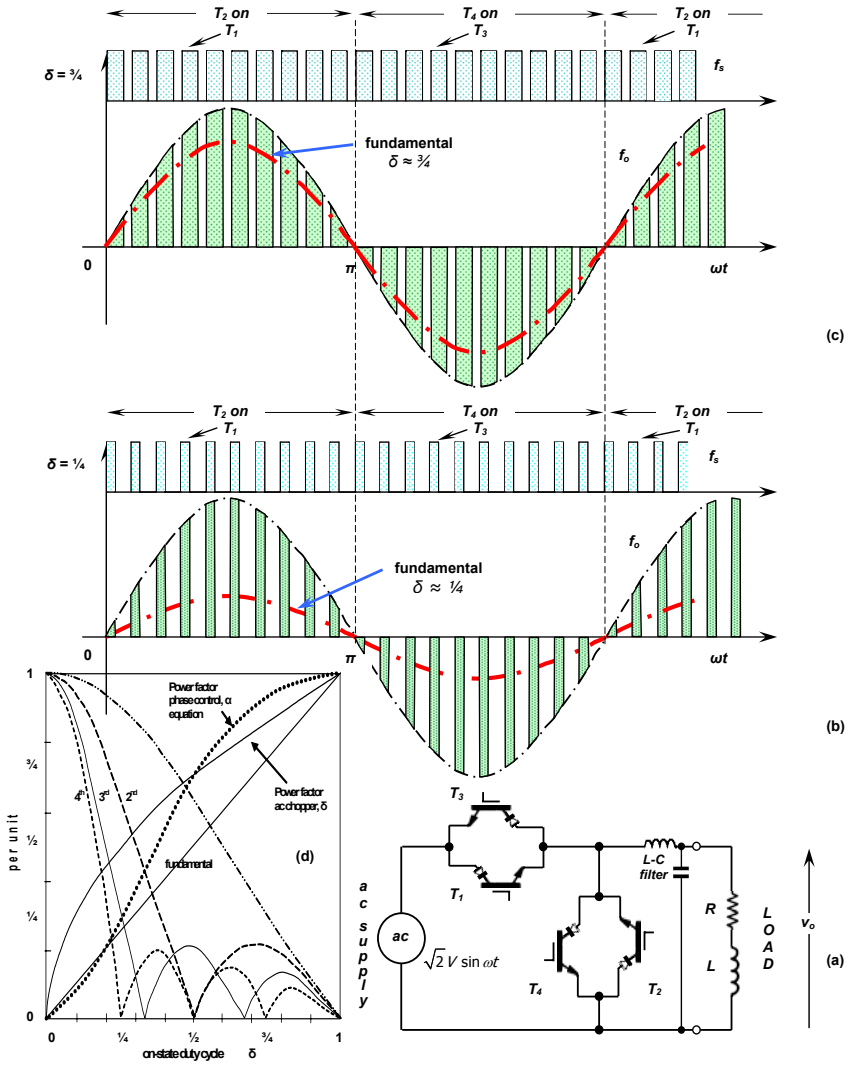


Figure 13.17. *An ac voltage regulator using a chopper, with commutable switches: (a) circuit configuration and output voltage waveform with a resistive load at (b) low modulation, δ≈¼; (c) high modulation, δ≈¾; and (d) harmonic characteristics.*

13.4 Three-phase ac regulator

13.4.1 Fully-controlled three-phase ac regulator with wye load and isolated neutral

The power to a three-phase star or delta-connected load may be controlled by the ac regulator shown in figure 13.18a with a star-connected load shown. The circuit is commonly used to soft start three-phase induction motors. If a neutral connection is made, load current can flow provided at least one thyristor is conducting. At high power levels, neutral connection is to be avoided, because of load triplen currents that may flow through the phase inputs and the neutral. With a balanced delta connected load, no triplen or even harmonic currents occur.
If the regulator devices in figure 13.18a, without the neutral connected, were diodes, each would conduct for ½π in the order T_1 to T_6 at ⅓π radians apart. As thyristors, conduction is from α to ½π.

Purely resistive load

In the fully controlled ac regulator of figure 13.18a without a neutral connection, at least two devices must conduct for power to be delivered to the load. The thyristor trigger sequence is as follows. If thyristor T_1 is triggered at α, then for a symmetrical three-phase load voltage, the other trigger angles are T_3 at α + ⅔π and T_5 at α + 4π/3. For the antiparallel devices, T_4 (which is in antiparallel with T_1) is triggered at α + π, T_6 at α + 5π/3, and finally T_2 at α + 7π/3.

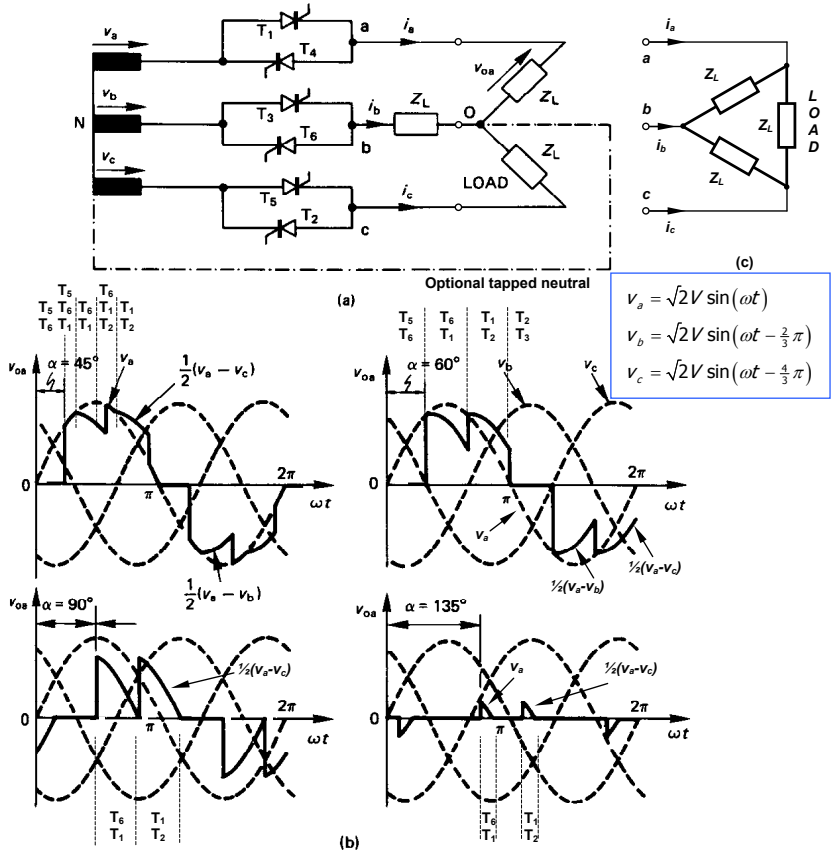


Figure 13.18. *Three-phase ac full-wave voltage controller: (a) circuit connection with a star load; (b) phase a, line-to-load neutral voltage waveforms for four firing delay angles; and (c) delta load.*

Figure 13.18b shows resistive load, line-to-neutral voltage waveforms (which are symmetrical about zero volts) for four different phase delay angles, α. Three distinctive conduction periods, plus a non-conduction period, exist. The waveforms in figure 13.18b are useful in determining the required bounds of integration. When three regulator thyristors conduct, the voltage (and the current) is of the form $\hat{V}/\sqrt{3}\sin\phi$, while when two devices conduct, the voltage (and the current) is of the form $\hat{V}/2\sin(\phi-\pi/6)$. $\hat{V}$ is the maximum line voltage, $\sqrt{3}\sqrt{2}V$.

i. $0 \le \alpha \le \pi/3$ [mode 3/2] – alternating every $\frac{1}{6}\pi$ between 2 and 3 conducting thyristors,
Full output occurs when $\alpha = 0$, when the load voltage is the supply voltage and each thyristor conducts for π. For $\alpha \le \pi/3$, in each half cycle, three alternating devices conduct and one will be turned off by natural commutation. The output voltage is continuous. Only for $\omega t \le \pi/3$ can three sequential devices be on simultaneously.
Examination of the $\alpha = \pi/4$ waveform in figure 13.18b shows the voltage waveform is made from five sinusoidal segments. The rms load voltage per phase (line to neutral), for a resistive load, is

$$V_{rms} = \hat{V}\left[\frac{1}{\pi}\left\{\int_{\alpha}^{\pi/4}\tfrac{1}{3}\sin^2\phi\ d\phi + \int_{\pi/4}^{\pi/3+\alpha}\tfrac{1}{4}\sin^2(\phi+\pi/6)\ d\phi + \int_{\pi/3}^{\pi/2+\alpha}\tfrac{1}{3}\sin^2\phi\ d\phi + \int_{\pi/3}^{\pi/2+\alpha}\tfrac{1}{4}\sin^2(\phi-\pi/6)\ d\phi + \int_{\pi/2+\alpha}^{\pi}\tfrac{1}{3}\sin^2\phi\ d\phi\right\}\right]^{1/2}$$

$$V_{rms} = I_{rms}R = V\left[1-\tfrac{3}{2\pi}\alpha+\tfrac{3}{4\pi}\sin 2\alpha\right]^{1/2} \qquad (13.66)$$

The Fourier coefficients of the fundamental frequency are

$$a_1 = \frac{3}{4\pi}V(\cos 2\alpha - 1) \qquad b_1 = \frac{3}{4\pi}V\left(\sin 2\alpha + \tfrac{4}{3}\pi - 2\alpha\right) \qquad (13.67)$$

Using the five integration terms as in equation (13.66), not squared, gives the average half-wave (half-cycle) load voltage, hence specifies the average thyristor current requirement with a resistive load. That is

$$\begin{aligned}\overline{V}_o^{\text{½cycle}} &= 2\times\overline{I}_T R = \sqrt{2}V\frac{1}{2\pi}\int_{\alpha}^{\pi}\sin\omega t\ d\omega t \\ \overline{V}_o^{\text{½cycle}} &= 2\times\overline{I}_T R = \frac{\sqrt{2}V}{\pi}(1+\cos\alpha)\end{aligned} \qquad (13.68)$$

The thyristor maximum average current is when $\alpha = 0$, that is $\hat{\overline{I}}_T = \frac{\sqrt{2}V}{\pi R}$.

ii. $\pi/3 \le \alpha \le \pi/2$ [mode 2/2] – two conducting thyristors
The turning on of one device naturally commutates another conducting device and only two phases can be conducting, that is, only two thyristors conduct at any time. Two phases experience half the difference of their input phase voltages, while the off thyristor is reverse biased by 3/2 its phase voltage, (off with zero current). The line-to-neutral load voltage waveforms for $\alpha = \pi/3$ and $\pi/2$, which are continuous, are shown in figures 13.18b.

Examination of the $\alpha = \pi/3$ or $\alpha = \pi/2$ waveforms in figure 13.18b show the voltage waveform is comprised from two segments. The rms load voltage per phase, for a resistive load, is

$$V_{rms} = \hat{V}\left[\frac{1}{\pi}\left\{\int_{\alpha}^{\frac{1}{2}\pi+\alpha}\tfrac{1}{4}\sin^2(\phi+\pi/6)\ d\phi + \int_{\frac{1}{2}\pi+\alpha}^{\frac{2}{3}\pi+\alpha}\tfrac{1}{4}\sin^2(\phi-\pi/6)\ d\phi\right\}\right]^{1/2}$$

$$V_{rms} = I_{rms}R = V\left[\tfrac{1}{2}+\tfrac{9}{8\pi}\sin 2\alpha+\tfrac{3\sqrt{3}}{8\pi}\cos 2\alpha\right]^{1/2} = V\left[\tfrac{1}{2}+\tfrac{3\sqrt{3}}{4\pi}\sin\left(2\alpha+\tfrac{\pi}{6}\right)\right]^{1/2} \qquad (13.69)$$

The Fourier co-efficients of the fundamental frequency are

$$a_1 = \frac{3}{4\pi}V\left(\cos 2\alpha - \cos 2\left(\alpha-\tfrac{\pi}{3}\right)\right) \qquad b_1 = \frac{3}{4\pi}V\left(\tfrac{2\pi}{3}+\sin 2\alpha - \sin 2\left(\alpha-\tfrac{\pi}{3}\right)\right) \qquad (13.70)$$

The non-fundamental harmonic magnitudes are independent of α, and are given by

$$V_h = \frac{3}{\pi(h\pm 1)}\times V\times\sin(h\pm 1)\tfrac{\pi}{6} \qquad \text{for} \quad h = 6k\pm 1 \quad k = 1,2,3,.. \qquad (13.71)$$

Using the same two integration terms, not squared, gives the average half-wave (half-cycle) load voltage, hence specifies the average thyristor current with a resistive load. That is

$$\begin{aligned}\overline{V}_o^{\text{½cycle}} &= 2\times\overline{I}_T R = \sqrt{3}\sqrt{2}V\frac{1}{2\pi}\int_{\alpha}^{\frac{1}{3}\pi+q}\sin(\omega t+\pi/6)\ d\omega t + \int_{\frac{1}{3}\pi+\alpha}^{\frac{2}{3}\pi+q}\sin(\omega t-\pi/6)\ d\omega t \\ \overline{V}_o^{\text{½cycle}} &= 2\times\overline{I}_T R = \frac{\sqrt{3}\sqrt{2}V}{\pi}\sin\left(\alpha+\frac{\pi}{3}\right)\end{aligned} \qquad (13.72)$$

iii. $\pi/2 \le \alpha \le \frac{5}{6}\pi$ [mode 2/0] – either 2 or no conducting thyristors
Two devices must be triggered in order to establish load current and only two devices conduct at anytime. Line-to-neutral zero voltage periods occur and each device must be retriggered $\pi/3$ after the initial trigger pulse. These zero output periods (discontinuous load voltage) which develop for $\alpha \ge \pi/2$ can be seen in figure 13.18b and are due to a previously on device commutating at $\omega t = \frac{5}{6}\pi$ then re-conducting at $\alpha + \pi/3$. Except for regulator start up, the second firing pulse is not necessary if $\alpha \le \pi/2$.
Examination of the $\alpha = 3\pi/4$ waveform in figure 13.18b shows the voltage waveform is made from two discontinuous voltage segments. The rms load voltage per phase, for a resistive load, is

$$V_{rms} = \hat{V}\left[\frac{1}{\pi}\left\{\int_{\alpha}^{\frac{5}{6}\pi}\tfrac{1}{4}\sin^2(\phi+\pi/6)\ d\theta + \int_{\frac{1}{3}\pi+\alpha}^{\frac{7}{6}\pi}\tfrac{1}{4}\sin^2(\phi-\pi/6)\ d\phi\right\}\right]^{1/2}$$

$$V_{rms} = I_{rms}R = V\left[\tfrac{5}{4}-\tfrac{3}{2\pi}\alpha+\tfrac{3}{8\pi}\sin 2\alpha+\tfrac{3\sqrt{3}}{8\pi}\cos 2\alpha\right]^{1/2} = V\left[\tfrac{5}{4}-\tfrac{3}{2\pi}\alpha+\tfrac{3}{4\pi}\sin\left(2\alpha+\tfrac{\pi}{3}\right)\right]^{1/2} \qquad (13.73)$$

The Fourier co-efficients of the fundamental frequency are

$$a_1 = -\frac{3}{4\pi}V\left(1+\cos 2\left(\alpha-\tfrac{1}{3}\pi\right)\right) \qquad b_1 = \frac{3}{4\pi}V\left(\tfrac{5}{3}\pi - 2\alpha - \sin 2\left(\alpha-\tfrac{1}{3}\pi\right)\right) \qquad (13.74)$$

Using the same two integration terms, not squared, gives the average half-wave (half-cycle) load voltage, hence specifies the average thyristor current with a resistive load. That is

$$\overline{V}_o^{\text{½cycle}} = 2\times\overline{I}_T R = \frac{\sqrt{3}\sqrt{2}V}{\pi}\left(1+\cos\left(\alpha+\pi/6\right)\right) \qquad (13.75)$$

iv. $\frac{5}{6}\pi \le \alpha \le \pi$ [mode 0] – no conducting thyristors
The interphase voltage falls to zero at $\alpha = \frac{5}{6}\pi$, hence for $\alpha \ge \frac{5}{6}\pi$ the output becomes zero.

In each case the phase current and line to line voltage are related by $V_{Lrms} = \sqrt{3}I_{rms}R$ and the peak voltage is $\hat{V} = \sqrt{2}\,V_L = \sqrt{6}\,V$. For a resistive load, load power $3I_{rms}^2R$ for all load types, and $V_{rms} = I_{rms}R$. Both the line input and load current harmonics occur at $6n\pm1$ times the fundamental.

Inductive-resistive load

Once inductance is incorporated into the load, current can only flow if the phase angle is at least equal to the load phase angle, given by $\phi = \tan^{-1}\omega L/R$. Due to the possibility of continuation of the load current because of the stored inductive load energy, only two thyristor operational modes occur. The initial mode at $\phi \le \alpha$ operates with three then two conducting thyristors mode [3/2], then as the control angle increases, operation in a mode [2/0] occurs with either two devices conducting or all three off, until $\alpha = \frac{5}{6}\pi$. The transitions between 3 and 2 thyristors conducting and between the two modes involves solutions to transcendental equations, and the rms output voltage, whence currents, depend on the solution to these equations.

Purely inductive load

For a purely inductive load the natural ac power factor angle is $\pi/2$, where the current lags the voltage by $\pi/2$. Therefore control for such a load starts from $\alpha = \pi/2$, and since the average inductor voltage must be zero, conduction is symmetrical about π and ceases at $2\pi - \alpha$. The conduction period is $2(\pi - \alpha)$. Two distinct conduction periods exist.

i. $\pi/2 \le \alpha \le 2\pi/3$ [mode 3/2] – either 2 or 3 conducting thyristors
Either two or three phases conduct and five integration terms give the load half cycle average voltage, whence average thyristor current, as

$$\overline{V}_o^{\text{½cycle}} = \frac{2\sqrt{2}V}{\pi}\left(2\cos\alpha - \sqrt{3}\sin\alpha + 1 + \sqrt{3}\right) \qquad (13.76)$$

The thyristor maximum average current is when $\alpha = \pi/2$.
When only two thyristors conduct, the phase current during the conduction period is given by

$$i(\omega t) = \frac{\sqrt{2}V}{\omega L}\left(\frac{3}{2}\cos\alpha - \frac{\sqrt{3}}{2}\cos\left(\omega t+\frac{\pi}{6}\right)\right) \qquad (13.77)$$

The load phase rms voltage and current are

$$\begin{aligned}V_{rms} &= V\left(\tfrac{5}{2}-\tfrac{3}{\pi}\alpha+\tfrac{3}{2\pi}\sin 2\alpha\right)^{1/2} \\ I_{rms} &= \frac{V}{\omega L}\left(\tfrac{5}{2}-\tfrac{3}{\pi}\alpha+\left(7-\tfrac{6}{\pi}\alpha\right)\cos^2\alpha+\tfrac{9}{2\pi}\sin 2\alpha\right)^{1/2}\end{aligned} \qquad (13.78)$$

The magnitude of the sin term fundamental ($a_1 = 0$) is

$$V_1 = b_1 = \frac{3}{2\pi}V\left(\tfrac{5}{3}\pi - 2\alpha + \sin 2\alpha\right) = I_1\omega L \qquad (13.79)$$

while the remaining harmonics ($a_h = 0$) are given by

$$V_h = b_h = \frac{3}{\pi} V \left(\frac{\sin(h+1)\alpha}{h+1} + \frac{\sin(h-1)\alpha}{h-1} \right) \tag{13.80}$$

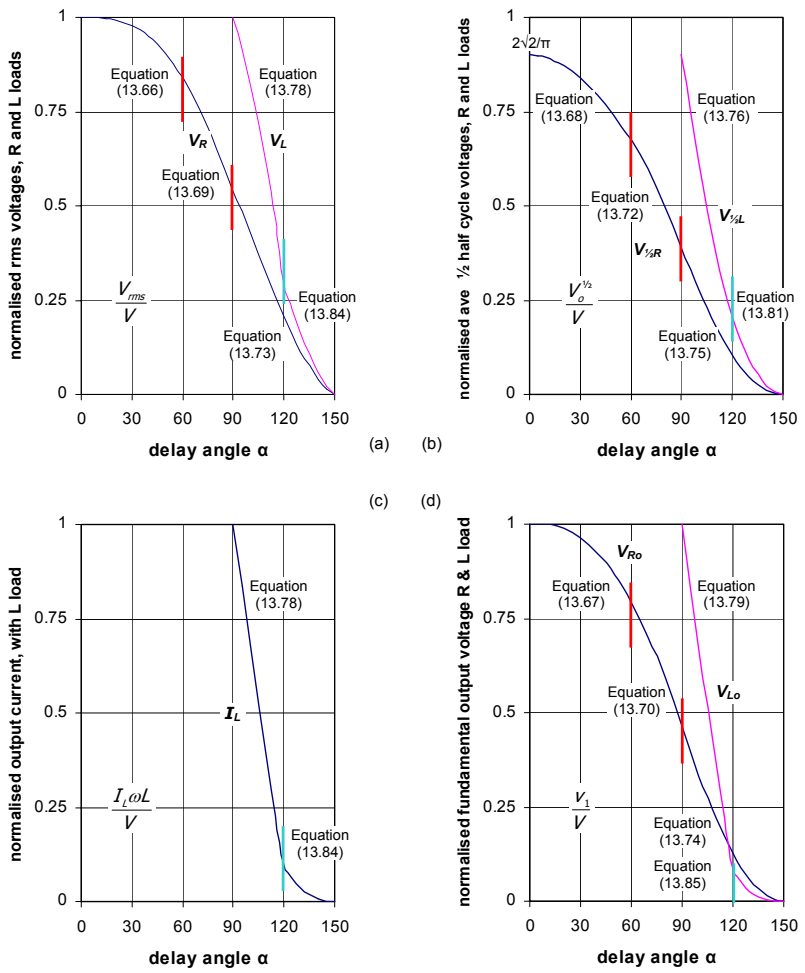


Figure 13.19. *Three-phase ac full-wave voltage controller characteristics for purely resistive and inductive loads: (a) normalised rms output voltages; (b) normalised half-cycle average voltages; (c) normalised output current for a purely inductive load; and (d) fundamental ac output voltage.*

ii. ⅔π ≤ α ≤ ⅚π [mode 2/0] – either 2 or no conducting thyristors
Discontinuous current flows in two phases, in two periods per half cycle and two integration terms (reduced to one after time shifting) give the load half cycle average voltage, whence average thyristor current, as

$$\overline{V}_o^{½\text{cycle}} = \frac{2\sqrt{2}V}{\pi} \sqrt{3} \left(1 + \cos\left(\alpha + \pi/6 \right) \right) \tag{13.81}$$

which reduces to zero volts at $\alpha = \frac{5}{6}\pi$.

The average thyristor current is given by

$$\begin{aligned} \overline{I}_T &= 2 \times \frac{1}{2\pi} \int_{\alpha}^{\frac{5}{3}\pi - \alpha} \frac{\sqrt{3}\sqrt{2}V}{2\omega L} \left[\cos\left(\alpha + \frac{\pi}{6} \right) - \cos\left(\omega t + \frac{\pi}{6} \right) \right] d\omega t \\ &= \frac{\sqrt{3}\sqrt{2}V}{2\pi\omega L} \left[\left(\frac{5}{3}\pi - 2\alpha \right) \cos\left(\alpha + \frac{\pi}{6} \right) - 2\sin\left(\alpha + \frac{\pi}{6} \right) \right] \end{aligned} \tag{13.82}$$

When two thyristors conduct, the phase current during the conduction period is given by

$$i(\omega t) = \frac{\sqrt{2}V}{\omega L} \frac{\sqrt{3}}{2} \left(\cos\left(\alpha + \frac{\pi}{6} \right) - \cos\left(\omega t + \frac{\pi}{6} \right) \right) \tag{13.83}$$

The load phase rms voltage and current are

$$\begin{aligned} V_{rms} &= V \left(\tfrac{5}{2} - \tfrac{3}{\pi}\alpha + \tfrac{3}{2\pi}\sin\left(2\alpha + \tfrac{1}{3}\pi \right) \right)^{½} \\ I_{rms} &= \frac{V}{\omega L} \left[\frac{5}{2} - \frac{3\alpha}{\pi} + \left(5 - \frac{6\alpha}{\pi} \right) \cos^2\left(\alpha + \tfrac{1}{6}\pi \right) + \frac{9}{2\pi}\sin\left(2\alpha + \tfrac{1}{6}\pi \right) \right]^{½} \end{aligned} \tag{13.84}$$

The magnitude of the sin term fundamental ($a_1 = 0$) is

$$V_1 = b_1 = \frac{3}{2\pi} V \left(\tfrac{5}{3}\pi - 2\alpha - \sin 2\left(\alpha - \tfrac{1}{3}\pi \right) \right) = I_1 \omega L \tag{13.85}$$

while the remaining harmonics ($a_h = 0$) are given by

$$V_h = b_h = \pm \frac{3}{\pi} V \left(\frac{\sin(h \pm 1)\alpha}{h \pm 1} + \frac{\sin(h \mp 1)(\alpha - \frac{1}{3}\pi)}{h \mp 1} \right) \quad \text{for } h = 6k \mp 1 \tag{13.86}$$

Various normalised voltage and current characteristics for resistive and inductive equations derived are shown in figure 13.19.

13.4.2 Fully-controlled three-phase ac regulator with wye load and neutral connected

If the load and supply neutral is connected in the three-phase thyristor controller with a wye load as shown in figure 13.20 and dashed in figure 13.18a, then (possibly undesirably) neutral current can flow and each of the three loads can be controlled independently. Undesirably, the third harmonic and its odd multiples are algebraically summed and returned to the supply via the neutral connection. At any instant $i_N = i_a + i_b + i_c$.

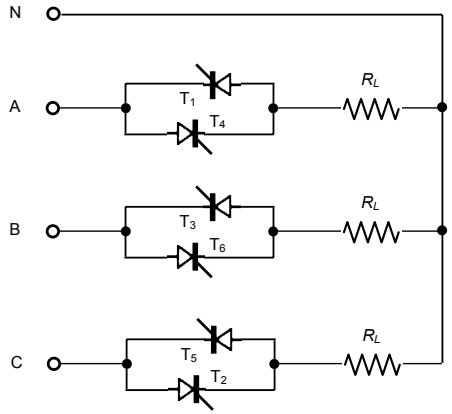


Figure 13.20. *Three-phase ac full-wave 4-wire star-load ac controller.*

For a resistive balanced load there are three modes of thyristor conduction.
When 3 thyristors conduct $i_a + i_b + i_c = I_N = 0$, two thyristor conduct $I_N = -\sqrt{2}V/R \sin\left(\omega t - \frac{4}{3}\pi \right)$, and for one thyristor $I_N = I_T = \sqrt{2}V/R \sin \omega t$.

Mode [3/2] $0 \le \alpha \le \frac{1}{3}\pi$

Periods of zero neutral current occur when three thyristors conduct and the rms of the discontinuous neutral current is given by

$$I_N^2 = \frac{3}{\pi}\int_{\frac{2}{3}\pi}^{\alpha+\frac{2}{3}\pi}\left(-\sqrt{2}V/R\,\sin\left(\omega t - \tfrac{4}{3}\pi\right)\right)^2 dt$$

$$I_N = \frac{V}{R}\times\left[\frac{3}{\pi}\left(\alpha - \tfrac{1}{2}\sin 2\alpha\right)\right]^{\frac{1}{2}} \qquad (13.87)$$

The average neutral current is

$$\bar{I}_N = \frac{3\sqrt{2}V}{\pi R}\left(1-\cos\alpha\right) \qquad (13.88)$$

At $\alpha = 0°$, no neutral current flows since the load is seen as a balance load supplied by the three-phase ac supply, without an interposing controller.

Mode [2/1] $\frac{1}{3}\pi \le \alpha \le \frac{2}{3}\pi$

From α to $\frac{2}{3}\pi$ two phase conduct and after $\frac{2}{3}\pi$ the neutral current is due to one thyristor conducting. The rms neutral current is given by

$$I_N^2 = \frac{3}{\pi}\int_{\alpha}^{\frac{2}{3}\pi}\left(-\sqrt{2}V/R\,\sin\left(\omega t - \tfrac{4}{3}\pi\right)\right)^2 dt + \frac{3}{\pi}\int_{\frac{2}{3}\pi}^{\alpha+\frac{2}{3}\pi}\left(\sqrt{2}V/R\,\sin\omega t\right)^2 dt$$

$$I_N = \frac{V}{R}\times\left[1 - \frac{3\sqrt{3}}{\pi}\cos^2\alpha\right]^{\frac{1}{2}} \qquad (13.89)$$

Maximum rms neutral current occurs at $\alpha = \frac{1}{2}\pi$, when $I_N = V/R$.

The average neutral current is

$$\bar{I}_N = \frac{3\sqrt{2}V}{\pi R}\left(\sqrt{3}\sin\alpha - 1\right) \qquad (13.90)$$

The maximum average neutral current, at $\alpha = \frac{1}{2}\pi$, is

$$\hat{\bar{I}}_N = \frac{3\sqrt{2}V}{\pi R}\left(\sqrt{3}-1\right) = 0.9886\frac{V}{R} \qquad (13.91)$$

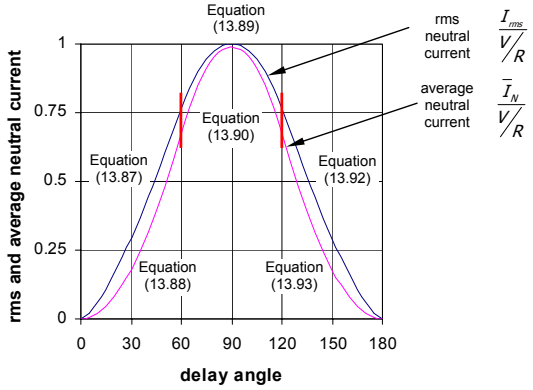


Figure 13.21. *Three-phase ac full-wave voltage neutral-connected controller with resistive load, normalised rms neutral current and normalised average neutral current.*

Mode [1/0] $\frac{2}{3}\pi \le \alpha \le \pi$

The neutral current is due to only one thyristor conducting. The rms neutral current is given by

$$I_N^2 = \frac{3}{\pi}\int_{\alpha}^{\pi}\left(\sqrt{2}V/R\,\sin\omega t\right)^2 dt$$

$$I_N = \frac{V}{R}\times\left[\frac{3}{\pi}\left(\pi - \alpha + \tfrac{1}{2}\sin 2\alpha\right)\right]^{\frac{1}{2}} \qquad (13.92)$$

The average neutral current is

$$\bar{I}_N = \frac{3\sqrt{2}V}{\pi R}\left(1+\cos\alpha\right) \qquad (13.93)$$

The neutral current is greater than the line current until the phase delay angle $\alpha > 67°$. The neutral current reduces to zero when $\alpha = \pi$, since no thyristors conduct.
The normalised neutral current characteristics are shown plotted in figure 13.21.

13.4.3 Fully-controlled three-phase ac regulator with delta load

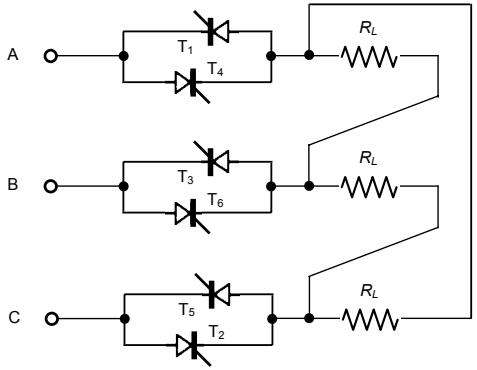


Figure 13.22. *Three-phase ac full-wave 3-wire delta-load ac controller.*

The load in figure 13.18a can be replaced with the start delta in figure 13.18c or figure 13.22. Star and delta load equivalence applies in terms of the same line voltage, line current, and thyristor voltages, provided the load is linear. A delta connected load can be considered to be three independent single phase ac regulators, where the total power (for a balanced load) is three times that of one regulator, that is

$$Power = 3\times VI_1\cos\phi_1 = \sqrt{3}VI_{L1}\cos\phi_1 \qquad (13.94)$$

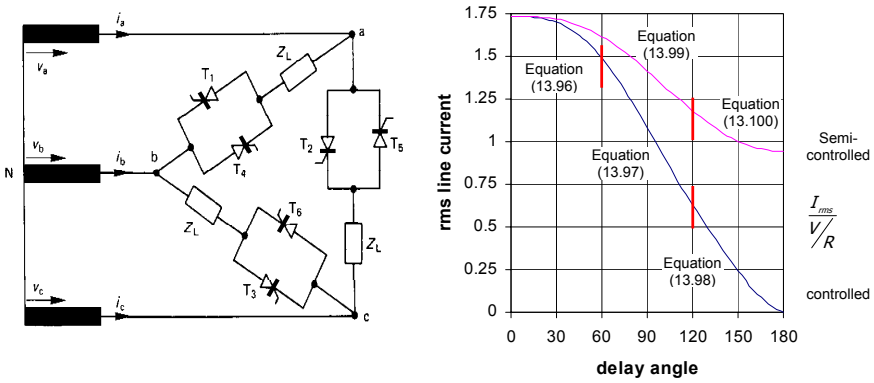
In load delta connection
For delta-connected loads where each phase end is accessible, the regulator shown in figure 13.23 can be employed in order to reduce thyristor current ratings. Each phase forms a separate single-phase ac controller as considered in section 13.1 but the phase voltage is the line-to-line voltage, $\sqrt{3}V$.

For a resistive load, the phase rms voltage, hence current, given by equations (13.23) and (13.24) are increased by √3, viz.:

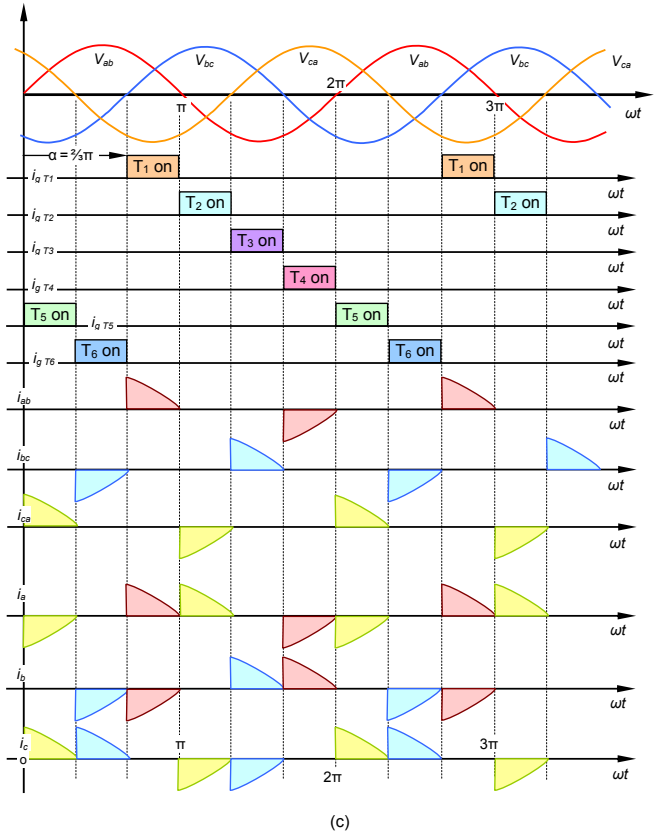
$$V_{rms} = \left[\frac{1}{2\pi}2\int_{\alpha}^{\pi}\left(\sqrt{2}V\sin\omega t\right)^2 dt\right]^{\frac{1}{2}}$$

$$V_{rms} = \sqrt{3}\,V\left[1 - \frac{\alpha}{\pi} + \frac{\sin 2\alpha}{2\pi}\right]^{\frac{1}{2}} = \sqrt{3}\,I_{rms}R \qquad 0 \le \alpha \le \pi \qquad (13.95)$$

The line current is related to the sum of two phase currents, each phase shifted by 120°. For a resistive delta load, three modes of phase angle dependent modes of operation can occur.



(a) (b)



(c)

Figure 13.23. *An in-delta connected three-phase ac regulator: (a) circuit configuration; (b) normalised line rms current for controlled and semi-controlled resistive loads; and (c) waveforms for an in-circuit resistive load with a 120° delay angle.*

Mode [3/2] $0 \le \alpha \le \frac{1}{3}\pi$

The line current is given by

$$I_L = \frac{\sqrt{3}}{R} V \left[1 - \frac{4}{3\pi}\alpha + \frac{2}{3\pi}\sin 2\alpha\right]^{\frac{1}{2}} \qquad (13.96)$$

Mode [2/1] $\frac{1}{3}\pi \le \alpha \le \frac{2}{3}\pi$

The line current is given by

$$I_L = \frac{\sqrt{3}}{R} V \left[\frac{8}{9} - \frac{1}{\pi}\alpha + \frac{\sqrt{3}}{6\pi}\left(1 + \sqrt{3}\sin 2\alpha + \cos 2\alpha\right)\right]^{\frac{1}{2}}$$
$$= \frac{\sqrt{3}}{R} V \left[\frac{8}{9} - \frac{1}{\pi}\alpha + \frac{\sqrt{3}}{6\pi}\left(1 + 2\sin\left(2\alpha + \frac{1}{6}\pi\right)\right)\right]^{\frac{1}{2}} \qquad (13.97)$$

Mode [1/0] $\frac{2}{3}\pi \le \alpha \le \pi$

The line current is given by

$$I_L = \frac{\sqrt{3}}{R} V \left[\frac{2}{3} - \frac{2}{3\pi}\alpha + \frac{1}{3\pi}\sin 2\alpha\right] \qquad (13.98)$$

The thyristors must be retriggered to ensure the current picks up after *α*.

Half-controlled

When the delta thyristor arrangement in figure 13.23 is half controlled (T_2, T_4, T_6 replaced by diodes) there are two mode of thyristor operation, with a resistive load.

Mode [3/2] $0 \le \alpha \le \frac{2}{3}\pi$

The line current is given by

$$I_L = \frac{\sqrt{3}}{R} V \left[1 - \frac{2}{3\pi}\alpha + \frac{1}{3\pi}\sin 2\alpha\right]^{\frac{1}{2}} \qquad (13.99)$$

Mode [2/1] $\frac{2}{3}\pi \le \alpha \le \pi$

The line current is given by

$$I_L = \frac{\sqrt{3}}{R} V \left[\frac{8}{9} - \frac{1}{2\pi}\alpha - \frac{\sqrt{3}}{12\pi}\left(1 - 2\sin\left(2\alpha - \frac{1}{6}\pi\right)\right)\right]^{\frac{1}{2}} \qquad (13.100)$$

13.4.4 Half-controlled three-phase ac regulator

The half-controlled three-phase regulator shown in figure 13.24a requires only a single trigger pulse per thyristor and the return path is via a diode. Compared with the fully controlled regulator, the half-controlled regulator is simpler and does not give rise to dc components but does produce more line harmonics.

Figure 13.24b shows resistive symmetrical load, line-to-neutral voltage waveforms for four different phase delay angles, *α*.

Resistive load

Three distinctive conduction periods exist.

i. **$0 \le \alpha \le \frac{1}{2}\pi$ – [mode3/2]**

Before turn-on, one diode and one thyristor conduct in the other two phases. After turn-on two thyristors and one diode conduct, and the three-phase ac supply is impressed across the load. The output phase voltage is asymmetrical about zero volts, but with an average voltage of zero. Examination of the $\alpha = \frac{1}{4}\pi$ waveform in figure 13.24b shows the voltage waveform is made from three segments. The rms load voltage per phase (line to neutral) is

$$V_{rms} = I_{rms} R = V\left[1 - \frac{3}{4\pi}\alpha + \frac{3}{8\pi}\sin 2\alpha\right]^{\frac{1}{2}} \qquad 0 \le \alpha \le \tfrac{1}{2}\pi \qquad (13.101)$$

The Fourier co-efficients for the fundamental voltage, for a resistive load are

$$a_1 = \frac{3}{8\pi} V\left(\cos 2\alpha - 1\right) \qquad b_1 = \frac{3}{8\pi} V\left(\frac{8\pi}{3} - 2\alpha + \sin\alpha\right) \qquad (13.102)$$

Using three integration terms, the average half-wave (half-cycle) load voltage, for both halves, specifies the average thyristor and diode current requirement with a resistive load. That is

$$\overline{V}_o^{\frac{1}{2}\text{cycle}} = 2 \times \overline{I}_T R = 2 \times \overline{I}_{Diode} R = \frac{\sqrt{2}V}{2\pi}\left(3 + \cos\alpha\right) \qquad 0 < \alpha < \tfrac{1}{3}\pi \qquad (13.103)$$

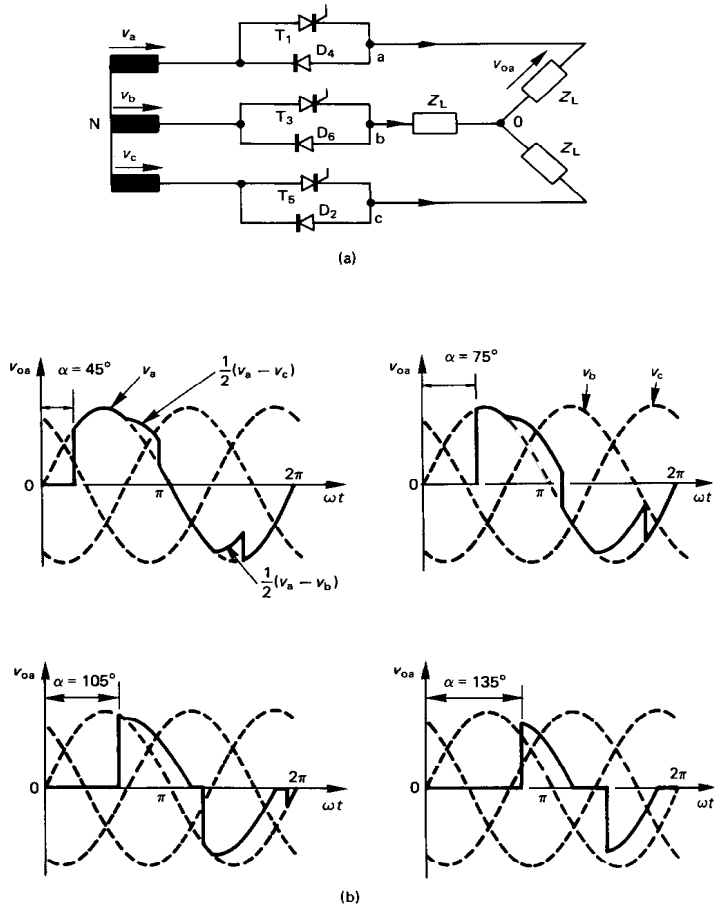


Figure 13.24. *Three-phase half-wave ac voltage regulator: (a) circuit connection with a star load and (b) phase a, line-to-load neutral voltage waveforms for four firing delay angles.*

The diode and thyristor maximum average current is when $\alpha = 0$, that is $\hat{\bar{I}}_T = \hat{\bar{I}}_{Diode} = \dfrac{\sqrt{2}V}{\pi R}$.

After α = ⅓π, only one thyristor conducts at one instant and the return current is a diode. Examination of the $\alpha = \frac{3}{8}\pi$ and $\alpha = \frac{5}{8}\pi$ waveforms in figure 13.24b show the voltage waveform is made from three segments, although different segments of the supply around ωt=π.
Using three integration terms, the average half-wave (half-cycle) load voltage, for both halves, specifies the average thyristor and diode current requirement with a resistive load. That is

$$\bar{V}_o^{½cycle} = 2 \times \bar{I}_T R = 2 \times \bar{I}_{Diode} R = \frac{\sqrt{2}V}{2\pi}\left(1 + 2\cos\alpha + \sqrt{3}\sin\alpha\right) \qquad \alpha > \tfrac{1}{3}\pi \qquad (13.104)$$

ii. **½π ≤ α ≤ ⅔π – [mode3/2/0]**

Only one thyristor conducts at one instant and the return current is shared at different intervals by one (⅓π ≤ α ≤ ½π) or two (½π ≤ α ≤ ⅔π) diodes. Examination of the $\alpha = \frac{3}{8}\pi$ and $\alpha = \frac{5}{8}\pi$ waveforms in figure 13.24b show the voltage waveform comprises two segments, although different segments of the supply around $\omega t = \pi$. The rms load voltage per phase (line to neutral) is

$$V_{rms} = I_{rms} R = V\left[\left\{\frac{11}{8} - \frac{3}{2\pi}\alpha\right\}\right]^{½} \qquad ½\pi \le \alpha \le ⅔\pi \qquad (13.105)$$

The resistive load fundamental is

$$a_1 = -\frac{3}{4\pi}V \quad b_1 = \frac{3}{4\pi}V\left(\tfrac{11}{6}\pi - 2\alpha\right) \quad \rightarrow \quad V_1 = \frac{3}{4\pi}V\sqrt{\left(1 + \left(\tfrac{11}{6}\pi - 2\alpha\right)^2\right)} = I_1 R \qquad (13.106)$$

Using two integration terms, the average half-wave (half-cycle) load voltage, for both halves, specifies the average thyristor and diode current requirement with a resistive load. That is

$$\bar{V}_o^{½cycle} = 2 \times \bar{I}_T R = 2 \times \bar{I}_{Diode} R = \frac{\sqrt{2}V}{2\pi}\left(1 + \sqrt{3} + 2\cos\alpha\right) \qquad (13.107)$$

iii. **⅔π ≤ α ≤ 7π/6 – [mode2/0]**

Current flows in only one thyristor and one diode and at 7π/6 zero power is delivered to the load. The output is symmetrical about zero. The output voltage waveform shown for $\alpha = \frac{3}{4}\pi$ in figure 13.24b has one component.

$$V_{rms} = I_{rms} R = V\left[\tfrac{7}{8} - \tfrac{3}{4\pi}\alpha + \tfrac{3}{16\pi}\sin 2\alpha - \tfrac{3\sqrt{3}}{16\pi}\cos 2\alpha\right]^{½} = V\left[\tfrac{7}{8} - \tfrac{3}{4\pi}\alpha + \tfrac{3}{8\pi}\sin\left(2\alpha - \tfrac{\pi}{3}\right)\right]^{½} \qquad ⅔\pi \le \alpha \le \tfrac{7}{6}\pi \qquad (13.108)$$

with a fundamental given by

$$a_1 = -\frac{3}{4\pi}V\cos^2\left(\alpha - \tfrac{2}{3}\pi\right) \qquad b_1 = \frac{3}{4\pi}V\left[\tfrac{7}{6}\pi - \alpha - ½\sin 2\left(\alpha - \tfrac{2\pi}{3}\right)\right] \qquad (13.109)$$

Using one integration term, the average half-wave (half-cycle) load voltage, for both halves, specifies the average thyristor and diode current requirement with a resistive load. That is

$$\bar{V}_o^{½cycle} = 2 \times \bar{I}_T R = 2 \times \bar{I}_{Diode} R = \frac{\sqrt{2}V}{2\pi}\sqrt{3}\left(1 + \cos\left(\alpha - \tfrac{\pi}{6}\right)\right) \qquad (13.110)$$

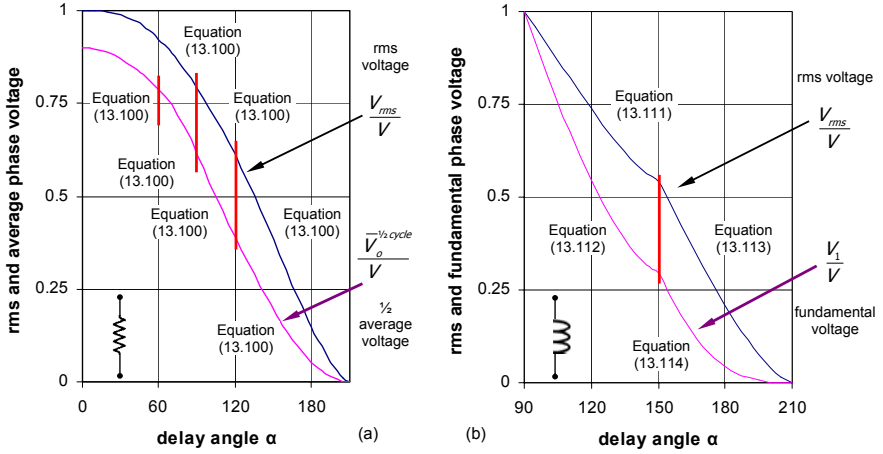


Figure 13.25. *Three-phase half-wave ac voltage regulator characteristics: (a) rms phase and average half cycle voltages for a resistive load and (b) rms and fundamental voltages for an inductive load.*

Purely inductive load

Two distinctive conduction periods exist.

i. **½π ≤ α ≤ ⅚π – [mode3/2]**
For a purely inductive load (cycle starts at α =½π)

$$V_{rms} = I_{rms}\omega L = V\sqrt{\tfrac{7}{4} - \tfrac{3}{2\pi}\alpha + \tfrac{3}{4\pi}\sin 2\alpha} \qquad ½\pi \le \alpha \le \tfrac{5}{6}\pi \qquad (13.111)$$

while for a purely inductive load the fundamental voltage is (a_1 = 0)

$$b_1 = V_1 = \frac{3}{4\pi}V\left(\tfrac{7\pi}{3} - 2\alpha + \sin 2\alpha\right) = I_1\omega L \qquad (13.112)$$

ii. $\frac{5}{6}\pi \le \alpha \le \frac{7}{6}\pi$ **– [mode2/0]**

For a purely inductive load, no mode 3/2/0 exist and rms load voltage for mode2/0 is

$$V_{rms} = I_{rms}\omega L = V\sqrt{\left(\frac{7}{4} - \frac{3}{2\pi}\alpha + \frac{3}{4\pi}\sin\left(2\alpha - \frac{\pi}{3}\right)\right)} \qquad (13.113)$$

with a fundamental given by ($a_1 = 0$)

$$b_1 = V_1 = \frac{3}{4\pi}V\left[\frac{7\pi}{3} - 2\alpha - \sin 2\left(\alpha - \frac{2\pi}{3}\right)\right] = I_1\omega L \qquad (13.114)$$

When $\alpha > \pi$, the load current is dominated by harmonic currents.
Normalised semi-controlled inductive and resistive load characteristics are shown in figure 13.25.

13.4.5 Other thyristor three-phase ac regulators

i. Delta connected fully controlled regulator

For star-connected loads where access exists to a neutral that can be opened, the regulator in figure 13.26a can be used. This circuit produces identical load waveforms to those for the regulator in figure 13.18 regardless of the type of load, except that mean device current ratings are halved (but the line currents are the same). Only one thyristor needs to be conducting for load current, compared with the circuit of figure 13.18 where two devices must be triggered. The triggering control is simplified but the maximum thyristor blocking voltage is increased by 2/√3, from 3V/√2 to √6V.

Three output voltage modes can be shown to occur, depending of the delay control angle.

Mode [2/1] $0 \le \alpha \le \frac{1}{3}\pi$
Mode [1] $\frac{1}{3}\pi \le \alpha \le \frac{1}{2}\pi$
Mode [1/0] $\frac{1}{2}\pi \le \alpha \le \frac{5}{6}\pi$

In figure 13.26a, at $\alpha = 0$, each thyristor conducts for $\frac{4}{3}\pi$, which for a resistive line load, results in a maximum thyristor average current rating of

$$\bar{I}_T = \frac{3}{2\pi}\frac{\sqrt{2}V_{L-L}}{R} = \frac{3}{2\pi}\frac{\sqrt{3}\sqrt{2}V}{R} \qquad (13.115)$$

A half-controlled version is not viable.

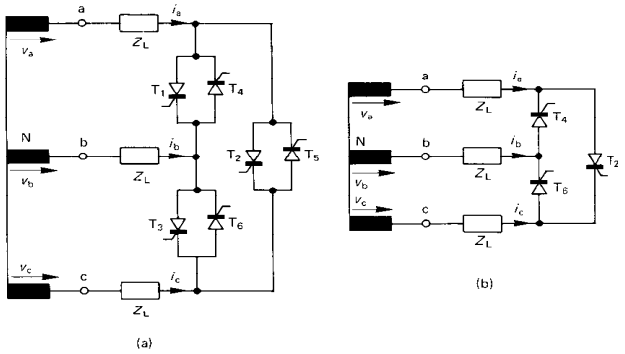


Figure 13.26. *Open-star three-phase ac regulators: (a) with six thyristors and (b) with three thyristors.*

ii. Three-thyristor delta connected regulator

The number of devices and control requirements for the regulator of figure 13.26a can be simplified by employing the regulator in figure 13.26b. In figure 13.26b, because of the half-wave configuration, at α = -⅓π, each thyristor conducts for ⅔π, which for a resistive line load, results in a maximum thyristor average current rating of

$$\bar{I}_T = \frac{3}{2\pi}\frac{\sqrt{2}V_{L-L}}{\sqrt{3}R} = \frac{3\sqrt{2}V}{2\pi R} \qquad (13.116)$$

Two thyristors conduct at any time as shown by the six sequential conduction possibilities that complete one mains ac cycle in figure 13.27.

Three output voltage modes can be shown to occur, depending of the delay control angle.

Mode [2/1] $-\frac{1}{3}\pi \le \alpha \le \frac{1}{6}\pi$
Mode [2/1/0] $\frac{1}{6}\pi \le \alpha \le \frac{1}{3}\pi$
Mode [1/0] $\frac{1}{3}\pi \le \alpha \le \frac{5}{6}\pi$

The control angle reference has been moved to the phase voltage crossover, the first instant the device becomes forward biased, hence able to conduct. This is ⅓π earlier than conventional three-phase fully controlled type circuits.

Another simplification, at the expense of harmonics, is to connect one phase of the load in figure 13.18a directly to the supply, thereby eliminating a pair of line thyristors.

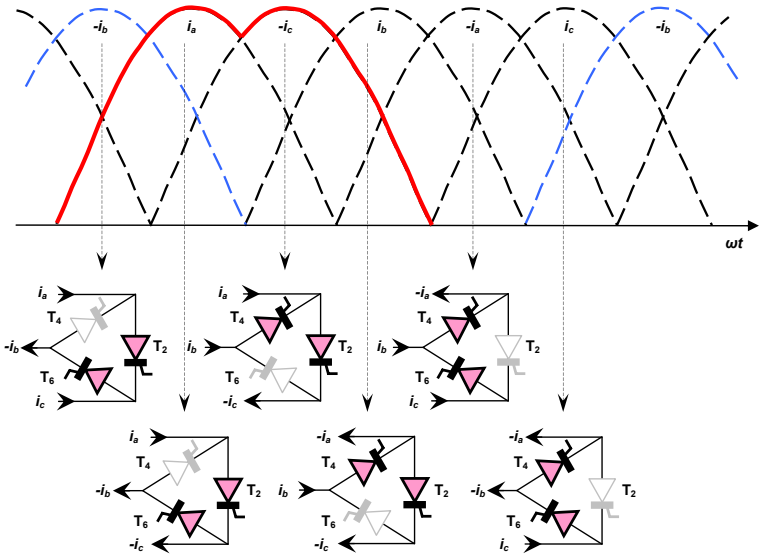


Figure 13.27. *Open-star three-phase ac regulators with three thyristors (figure 13.26b): (a) thyristors currents and (b) six line current possibilities during consecutive 60° segments.*

Table 13.1: Thyristor electrical ratings for four ac controllers

Circuit figure	Max input line rms current, I_{ac}	Max load power	Thyristor				Control delay angle range	
			Voltage	rms current	Peak current	Mean current	Resistive load	Inductive load
13.18	$V/\sqrt{3}Z$	$3I_{ac}^2R$	$\times\sqrt{2}V$	$\sqrt{3/2}$	$1/\sqrt{2}$	$\sqrt{2}$	$\sqrt{2}/\pi$	$0 \le \alpha \le \frac{5}{6}\pi$	$\frac{1}{2}\pi \le \alpha \le \frac{5}{6}\pi$
13.20	$V/\sqrt{3}Z$	$3I_{ac}^2R$	$\sqrt{2/3}$	$1/\sqrt{2}$	$\sqrt{2}$	$\sqrt{2}/\pi$	$0 \le \alpha \le \pi$	
13.22	$\sqrt{3}V/Z$	I_{ac}^2R	1.225	$1/\sqrt{2}$	$\sqrt{2}$	$\sqrt{2}/\pi$	$0 \le \alpha \le \pi$	
13.23	$\sqrt{3}V/Z$	I_{ac}^2R	$\sqrt{2}$	$\frac{1}{2}\sqrt{2/3}$	0.816	$\sqrt{2}/\pi/\sqrt{3}$	$0 \le \alpha \le \pi$	$\frac{1}{2}\pi \le \alpha \le \pi$
13.24	$V/\sqrt{3}Z$	$3I_{ac}^2R$	$\sqrt{3/2}$	$1/\sqrt{2}$	$\sqrt{2}$	$\sqrt{2}/\pi$	$0 \le \alpha \le \frac{1}{6}\pi$	$\frac{1}{2}\pi \le \alpha \le \frac{1}{6}\pi$
13.26a	$V/\sqrt{3}Z$	$3I_{ac}^2R$	$\sqrt{2}$	$\frac{1}{2}\sqrt{2/3}$	0.816	$\sqrt{2}/\pi/\sqrt{3}$	$0 \le \alpha \le \frac{5}{6}\pi$	
13.26b			$\sqrt{2}$	0.766			$-\frac{1}{3}\pi \le \alpha \le \frac{5}{6}\pi$	$\frac{1}{2}\pi \le \alpha \le \frac{1}{6}\pi$

Example 13.6: ***Star-load three-phase ac regulator – untapped neutral***

A 230V (line to neutral) 50Hz three-phase mains ac thyristor chopper has a symmetrical star load composed of 10Ω resistances. If the thyristor triggering delay angle is α = 90° determine

i. The rms load current and voltage, and maximum rms load current for any phase delay angle
ii. The power dissipated in the load
iii. The thyristor average and rms current ratings and voltage ratings
iv. Power dissipated in the thyristors when modelled by $v_T = v_o + r_o \times i_T = 1.2 + 0.01 \times i_T$

Repeat the calculations if each phase load is a 20mH.

Solution

(a) ***10Ω Resistive load -*** α = 90°

i. rms voltage from equation (13.69)

$$V_{rms} = I_{rms} R = V\left[\tfrac{1}{2} + \frac{3\sqrt{3}}{4\pi}\cos\left(2\alpha + \tfrac{\pi}{6}\right)\right]^{\frac{1}{2}}$$

$$= 230\text{V}\left[\tfrac{1}{2} + \frac{3\sqrt{3}}{4\pi}\cos\left(2\times 90^\circ + 30^\circ\right)\right]^{\frac{1}{2}} = 230\text{V}\times 0.377 = 86.6\text{V}$$

Whence the rms current

$$I_{rms} = \frac{V_{rms}}{R} = \frac{86.6\text{V}}{10\Omega} = 8.66\text{A}$$

ii. The load power is

$$P_{10\Omega} = I_{rms}^2 R = 8.66^2 \times 10\Omega = 750.7\text{W}$$

iii. Thyristor average current from equation (13.72)

$$\bar{I}_T = \frac{\sqrt{3}\sqrt{2}V}{2\pi R}\left(\sin\alpha - \tfrac{1}{2}\right)$$

$$= \frac{\sqrt{3}\sqrt{2}\,230\text{V}}{2\pi 10\Omega}\left(\sin\tfrac{\pi}{2} - \tfrac{1}{2}\right) = 4.48\text{A}$$

Thyristor rms current

$$I_{T\,rms} = \frac{I_{rms}}{\sqrt{2}} = \frac{8.66\text{A}}{\sqrt{2}} = 6.12\text{A}$$

iv. Thyristor loss

$$P_T = v_o \bar{I}_T + r_o i_{T\,rms}^2 = 1.2\times\bar{I}_T + 0.01\times i_{T\,rms}^2$$

$$= 1.2\times 4.48\text{A} + 0.01\times 6.12^2 = 5.75\text{W}$$

(b) ***20mH Inductive load -*** α = 90°

i. rms voltage and current from equation (13.78)

$$V_{rms} = V\left(\tfrac{5}{2} - \tfrac{3}{\pi}\alpha + \tfrac{3}{2\pi}\sin 2\alpha\right)^{\frac{1}{2}}$$

$$= 230\text{V}\left(\tfrac{5}{2} - \tfrac{3}{\pi}\tfrac{1}{2}\pi\right)^{\frac{1}{2}} = 230\text{V}$$

$$I_{rms} = \frac{V}{\omega L}\left(\tfrac{5}{2} - \tfrac{3}{\pi}\alpha + \left(7 - \tfrac{6}{\pi}\alpha\right)\cos^2\alpha + \tfrac{9}{2\pi}\sin 2\alpha\right)^{\frac{1}{2}}$$

$$= \frac{230\text{V}}{2\pi 50\text{Hz}\times 0.02\text{H}}\left(\tfrac{5}{2} - \tfrac{3}{\pi}\tfrac{1}{2}\pi\right)^{\frac{1}{2}} = \frac{230\text{V}}{2\pi 50\text{Hz}\times 0.02\text{H}} = 36.6\text{A}$$

ii. The load power is zero.

iii. Since the delay angle is 90°, the natural power factor angle, continuous sinusoidal current flows and the thyristor average current is

$$\bar{I}_T = \frac{1}{\sqrt{2}}\frac{2\sqrt{2}}{\pi} I_{rms} = \frac{1}{\sqrt{2}}\frac{2\sqrt{2}}{\pi} 36.6\text{A} = 23.3\text{A}$$

Thyristor rms current

$$I_{T\,rms} = \frac{I_{rms}}{\sqrt{2}} = \frac{36.6\text{A}}{\sqrt{2}} = 25.88\text{A}$$

iv. Thyristor loss

$$P_T = v_o \bar{I}_T + r_o i_{T\,rms}^2 = 1.2\times\bar{I}_T + 0.01\times i_{T\,rms}^2$$

$$= 1.2\times 25.88\text{A} + 0.01\times 36.6^2 = 44.45\text{W}$$

♣

13.4.6 Solid-state soft starters

An electric motor soft starter is a device used to temporarily reduce the load and torque in the power-train of any electric motor during start-up. This reduces the mechanical stress on the motor and shaft, as well as the electrodynamic stresses on the interconnecting power cables and electrical distribution network, thereby extending system lifetime.
Motor soft starters can consist of mechanical or electrical devices, or a combination of both. Mechanical soft starters include clutches and several types of couplings using a fluid or magnetic forces. Electrical soft starters can be any control system that reduces the torque by temporarily reducing the voltage or current input, or a device that temporarily alters how the motor is connected in the electric supply circuit.
In the case of the three-phase induction motor, electrical soft starters can utilize solid-state devices to control the current flow and therefore the voltage applied to the motor. The starter can be connected in series with the line voltage applied to the motor, or can be connected inside the delta loop of a delta-connected motor, thus is able to control the voltage applied to each winding. Solid-state soft starters can control one or more phases of the voltage applied to the induction motor with the best results achieved by three-phase control. Typically, the voltage is controlled by inverse-parallel-connected silicon-controlled rectifiers (SCR), but in some circumstances with three-phase control, the control elements can be a inverse-parallel-connected SCR and diode combination.
A solid-state soft starter is basically a three-phase ac to ac fully controlled regulating converter as shown in figure 13.18, used to soft-start three-phase ac caged induction motors. A soft-starter is functionally two/three ac instantaneous controlled solid-state relays, as in section 13.1.3, with the two/three isolated dc control inputs connected together.

13.4.6i The induction motor

The induction motor is the simplest and most rugged of all electric motors. They consist of two basic electrical assemblies: the wound stator and the rotor assembly.
Three-phase voltage supplies the stator windings which produce a three-phase rotating magnetic field.
The electrically isolated rotor consists of laminated, cylindrical iron cores with slots for receiving the conductors. On early motors, the conductors were copper bars with ends welded to copper rings known as end rings. Viewed from the end, the rotor assembly resembles a squirrel cage, hence the name squirrel-cage motor is used to refer to induction motors. In modern induction motors, the most common type of rotor has cast-aluminium conductors and short-circuiting end rings. The rotor turns when the stator rotating magnetic field induces a current in the rotor shorted conductors. This rotor current produces a rotor magnetic field which interacts with the stator field, producing a rotating torque. The speed at which the stator magnetic field rotates is the synchronous speed of the motor and is determined by the number of poles in the stator and the frequency of the ac power supply voltage.

$$n_s = \frac{60\times f}{P} \qquad (13.117)$$

where n_s = synchronous speed, rpm
f = frequency, Hz
P = number of pole pairs

Synchronous speed is the absolute upper limit of motor speed. At synchronous speed, there is no difference between rotor speed and the rotating field speed, so no voltage is induced in the rotor bars, hence no torque is developed. Therefore, when running, the rotor must rotate slower than the magnetic field. The rotor speed is just slow enough to cause the proper amount of rotor current to flow, so that the resulting torque is sufficient to overcome windage and friction losses, and drive the load. This speed difference between the rotor and stator magnetic field, called slip, is normally referred to as a percentage of synchronous speed:

$$s = \frac{n_s - n_R}{n_s} \qquad (13.118)$$

where s = slip
n_R = actual rotor speed, rpm

In order to appreciate the attributes of using an electronic motor controller, it is necessary to have an understanding of the characteristics and limitations of the three-phase ac caged induction (asynchronous) motor and the traditional electromechanical systems used to control it.

The standard, fixed-speed induction motor fulfils two basic mechanical requirements:

- accelerates itself and its mechanically connected rotational load to full speed and
- maintains the load at full speed efficiently and effectively over the full range of loadings.

Due to the constraints of machine materials and design, it is difficult to achieve both mechanical objectives effectively and economically in one machine.
Electromechanical motors convert electrical energy drawn from the ac power supply into a mechanical rotating form, usually as a shaft rotating at a speed related to the number of machine pole pairs, *P*, and the frequency of the ac supply, *f*. The mechanical power P_m (W) available from the shaft is equal to the mechanical torque T_m (N) multiplied by the shaft speed, n_R (rad/s), $P_m = T_m \times n_R$. From an initial value at standstill, the torque varies as the machine accelerates, reaching a peak at about 80% full speed, finally reducing to zero at synchronous speed, $n_s = 60f/P$ (rpm). This characteristic means that induction motors always operates at slightly less than synchronous speed, the 'slip speed', $n_{slip} = s \times n_s$ (rpm), in order to develop power, hence the term asynchronous machine. The characteristics in figure 13.28a show an induction motor torque-speed curve, which illustrates the most important mechanical output characteristics.

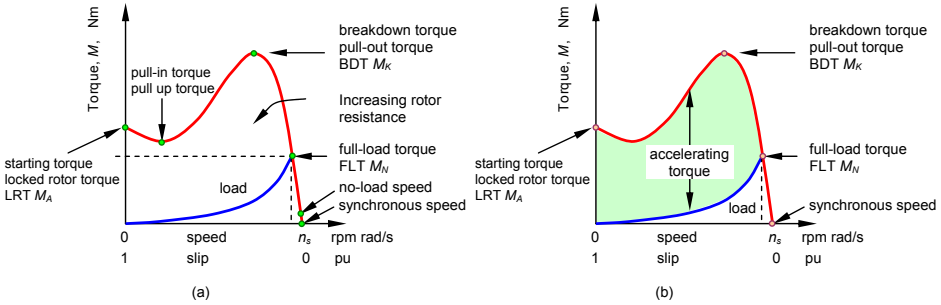


Figure 13.28. *Torque-speed curve for the induction motor showing:*
(a) the coupled load torque requirement and (b) the available accelerating torque.

Any load mechanically coupled to an induction motor has its own particular speed-torque characteristic requirement curve. The acceleration of a motor-load system is due to the difference between the motor developed torque and the load absorbed torque, as shown by the shaded area in figure 13.28b. The larger the torque difference, the higher the acceleration and the quicker full speed is reached, whence the greater the electrical and mechanical stresses experienced by the ac supply and drive system during the acceleration period. An 'ideal' starter accelerates the load with minimal intervention to reach full speed smoothly in a reasonable time, with minimum stress to the supply and drive train.

The motor speed-torque characteristic can be controlled by the rotor cage resistance, where a motor with high rotor resistance can generate its peak torque (pull-out torque) at standstill giving a high break-away torque characteristic, which progressively reduces, as the speed increases, to zero at synchronous speed, NEMA design D in figure 13.29a. A motor with a low rotor resistance will produce a low starting torque but will generate its peak torque closer to the synchronous speed, NEMA designs A and B. Consequently, this type of motor runs at full power with a higher operating efficiency and low slip speed. Induction motors that combine the dual requirements of high starting torque and efficient full-speed operation within a single motor have a double-cage or deep bar design, and this motor characteristic, shown in Figure 13.29a, NEMA design C, is ideal for use with soft starter control. All motors, except class D types, operate at 5% slip or less at full (rated) load.

Notice that all the design classes produce a starting torque and pull-up torque that are greater than the full-load torque level. Full load torque can be developed at any speed, but at the expense high current, low power factor, and low efficiency, hence increased motor heating. The important difference between the classes is the torque per ampere hence efficiency in the normal operating range, viz., at and above rated speed.

A induction motor with two poles often has a lower starting torque than motors with four or more poles, thus oversized motors may be used to ensure that their mechanical load can be started and driven under all operating conditions.

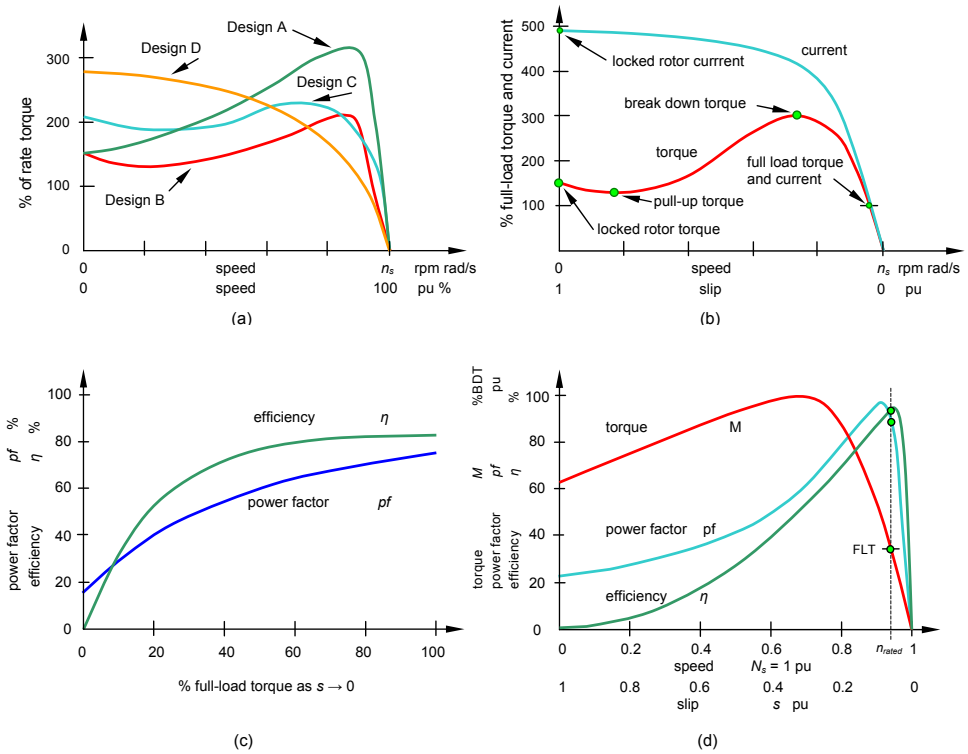


Figure 13.29. *Induction motor characteristics:*
(a) torque-speed curves for various NEMA classes of three-phase ac caged induction motors; (b) speed versus torque and current characteristics; (c) power factor and efficiency versus full-load torque; and (d) torque and efficiency versus rotor speed/slip.

Power factor
Induction motors present a lagging (inductive) power factor to the ac power line. The power factor in large fully loaded high-speed motors can be better than 0.90. At ¾ full-load, for large high-speed motors the power factor can be 92%. The power factor for small low-speed motors can be as low as 0.5. At starting, the power factor can be in the range of 0.1 to 0.25, improving (increasing) as the rotor gains speed.

Power factor (pf) varies considerably with the motor mechanical load as seen in figure 13.29c. An unloaded motor is analogous to a transformer without a secondary load. Little resistance is reflected from the secondary (rotor) to the primary (stator). Thus the ac line sees a reactive load, dominated by the magnetising current, resulting in a pf as low as 0.1 lagging. As the rotor is loaded, an increasing resistive component (representing the developed output power) is reflected from rotor to stator, increasing the power factor.

Efficiency
Large three-phase motors are more efficient than small three-phase motors. Large induction motor efficiency can be as high as 95% at full load, though better than 90% is common. Efficiency for a lightly load or no-load induction motor is poor because most of the current is involved with maintaining the magnetizing flux. As the torque load is increased, more current is consumed in generating torque, while current associated with magnetizing remains fixed. Efficiency at 75% FLT can be slightly higher than that at FLT. Efficiency is decreased a few percent at 50% FLT, and decreases more at 25% FLT. Efficiency only becomes poor below 25% FLT. The variation of efficiency with loading and speed is shown in Figure 13.29, parts c and d.

13.4.6ii Background to induction machine starting

Traditionally there are several ways to start three-phase ac induction motors. Starting via the use of series resistors and chokes/reactors or shunt capacitors are not considered.

Direct-on-line starting DoL

The simplest starting approach is to connect the motor to the ac voltage power supply via contactors and overload relays, and start the motor at full line voltage. This method is called *direct-on-line* (DoL) motor starting. The motor high in-rush current (locked rotor current) can be 5 to 10 times the motor full load current, as seen in figure 13.29b. These high starting currents cause voltage dips and sag in a weak power supply system. DoL starting also causes excessive torques in the mechanical system being driven, causing undesirable shocks among mechanical components, such as gears, belts, sheaves, and connections. Systems exposed to such shock will require frequent inspection and maintenance that lead to costly down-time.

Other viable approaches for electromechanical reduced-voltage starting use either a three-phase step-down auto-transformer or wye-delta starters.

Autotransformer starting

An auto-transformer is a non-isolating transformer that can be tapped to deliver any percentage (including >100%) of full voltage, for example 58%. By starting the motor at reduced voltage, electromechanical starters are able to crudely reduce the in-rush current and developed torque. Although the voltage is increased in limited steps, there are sudden current changes and mechanical shock during transitions.

Motor-start rated circuit breakers (slow opening operation to current surges) replace standard circuit breakers for starting motors of a few kilowatts. This interlocked breaker accepts high over-current for the duration of starting. In figure 13.30a, closure of the start contacts S applies reduced voltage during the start interval. The S contacts open and the run contacts R close after starting. This reduces the starting current to, say, 200% of full-load current. Since the autotransformer is only used for the short start interval, it may be sized considerably smaller than for a continuous duty application. By controlling two phases, which defines the third phase, the three autotransformer version can be reduced to two auto-transformers, as shown in figure 13.30b, if phase current imbalances can be tolerated for the short transient start-up period.

The basic auto-transformer starter has the disadvantage that at the contactor transition instant from 'start' to 'run' the supply to the motor is interrupted, termed an *open transition start*. This means that the electrical system insulation is stressed by the resultant high transient voltages. A *closed transition start* method keeps the motor connected to the supply continuously by means of the connection shown in figures 13.30, parts c and d for two and three controlled phases respectively.

The three start-up sequence of stages, for one phase of the three machine phases, are shown in part e of figure 13.30, which can be described as follows:

- First *Stage-1*, switches L and S close and the motor accelerates at a reduced voltage determined by the autotransformer tapping.
- Then the second *Stage-2*, the star point of the transformer (switch S) is opened so that the motor continues to run with part of the transformer winding in circuit.
- Next, *Stage-3*, this winding section is short-circuited by the 'run' contactor or switch (switch R closes).

The initial starting line current is approximately

$$Starting\ Current = 1.1 \times \left(\frac{applied\ voltage}{full\ voltage} \right)^2 \times \text{standstill current with full volts} \qquad (13.119)$$

The factor of 1.1 compensates for the auto-transformer magnetizing current.

The initial starting torque is approximately

$$Starting\ starting = \left(\frac{applied\ voltage}{full\ voltage} \right)^2 \times \text{standstill torque with full volts} \qquad (13.120)$$

These formulae for initial starting current and torque are approximate because it is assumed for simplicity that the standstill-reactance of a motor is constant at all voltages, that is, the short-circuit current varies in direct proportion to the applied voltage. Owing to magnetic saturation, particularly of the machine slot tips, the standstill reactance tends to be less on full volts than on reduced volts so the current and torque values tend to be less than those given by the formulae.

A fully automatic three-phase starter comprises a triple-pole line contactor, start contactor, running contactor, three single-pole overload relays, auto-transformer with a set of links for tap-changing, a suitable timer, and 'start' and 'stop' pushbuttons. The two auto-transformer version shown in figure 13.30d, reduces the number of poles needed on the start ,S, and run, R, contactors, from three poles to two poles.

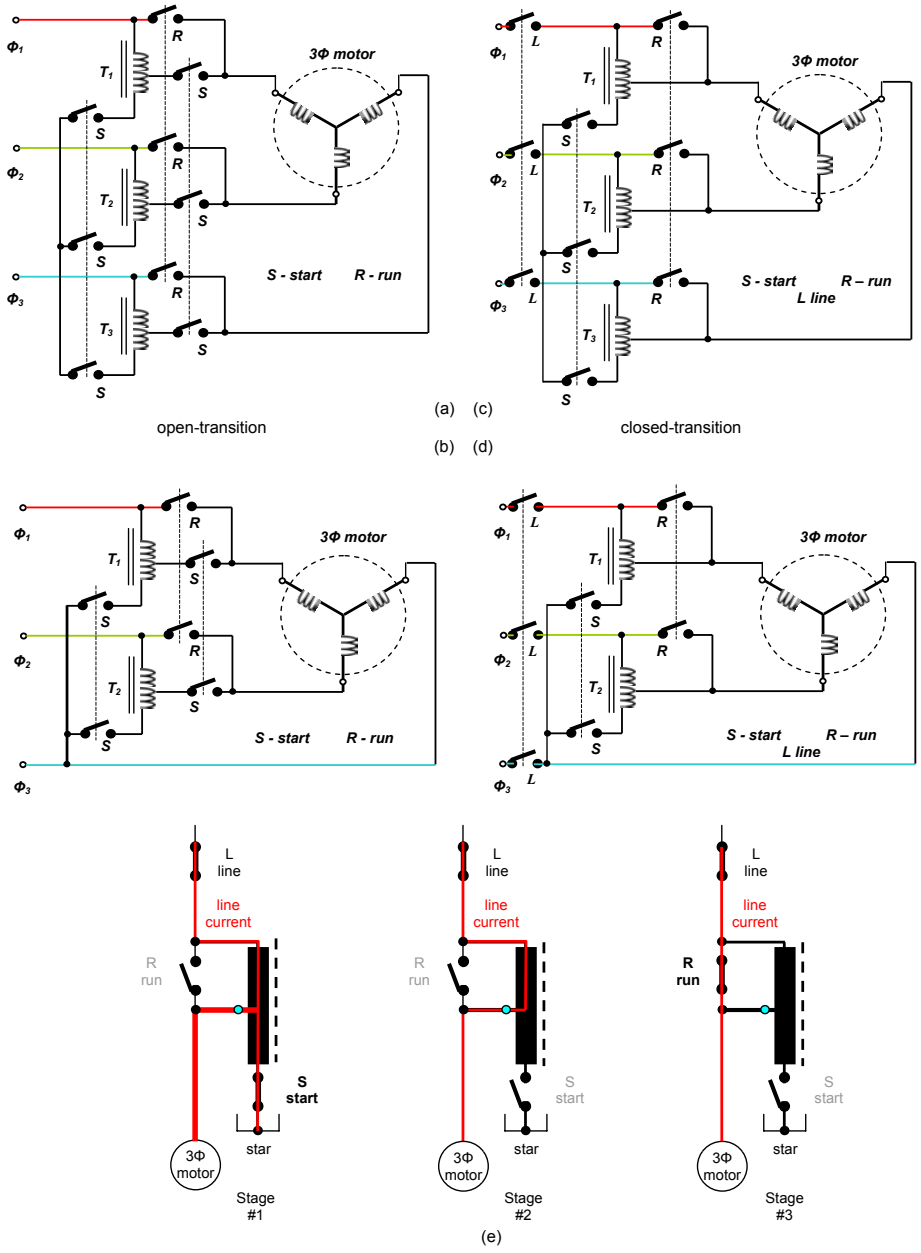


Figure 13.30. *Basic auto-transformer induction motor starter, with two and three-phase independent control: (a) and (b) open transition switching sequence, (c) and (d) closed transition switching sequence, and (e) three stages of the closed transition switching sequence.*

Star-delta or wye-delta starter

From equations (13.119) and (13.120), reduced voltage starting utilises the fact that motor torque and current are proportional to the square of the terminal voltage. This is exploited in the most familiar type of reduced-voltage starter, namely the star-delta or wye-delta starter, shown in figure 13.31. The star-delta starter consists of three contactors and a time switch (which can be mechanical, pneumatic, electrical or electronic). With the wye-delta starter, the special winding terminated motor (access to three individual windings) is first run as a wye motor so each motor winding only experiences 58% of full voltage, 1/√3. After a set period-of-time, the starter switches to run the motor as a delta connected motor and the motor windings experience the full ac line voltage. The change-over is controlled by the timer switch and is usually arranged to switch at 80% of full speed. The effect of starting in star, with each stator winding voltage reduced to 58%, 1/√3, of normal, is a reduction in the starting torque to a third of locked rotor torque (LRT) with a consequential reduction in starting current and acceleration force. Effectively, the voltage is reduced by a 1.732 factor,√3. The impedance seen by the power system is 3 times the impedance of the delta run connection. The starting sequence and resultant torque/current characteristics are shown in figure 13.36. For Wye Start, Delta Run, the starting characteristics are:

- starting current is approximately 30% of that obtained with the normal delta connection.
- starting torque is approximately 25-30% of that realised with the normal delta connection.

This starting method is only viable when the system is light loaded during the start. Although an improvement over the DoL system, disadvantages remain.

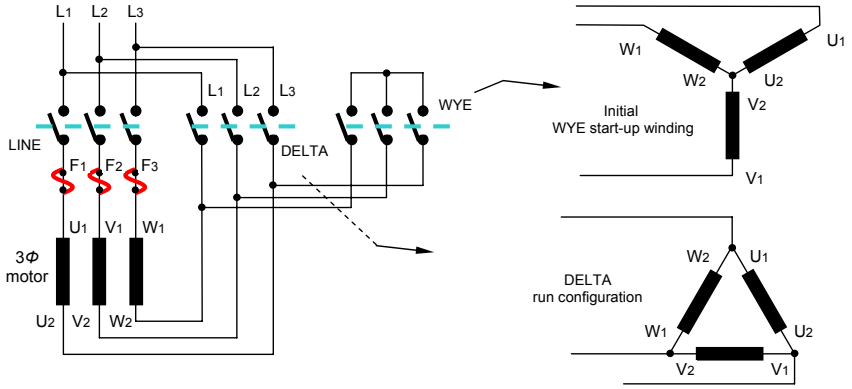


Figure 13.31. *Basic wye-start, delta-run connection configuration for induction motor starting.*

The transfer from a star to a delta connection momentarily disconnects the motor from the supply. During this time it is under the mechanical influence of the rotating load and, at the instant of disconnection, current continues to flow in the rotor bars due to the long time delay necessary for the magnetic flux to die away. Therefore, there is a residual flux 'frozen' on the surface of the rotating rotor, which cuts the stator windings, generating a voltage whose frequency depends on rotor speed. If the load inertia is small, such as in a pump, or if the friction is high, there could be a significant loss of speed during the time the supply is disconnected. In this case, when the delta run connection is made, a large phase differential can exist between the supply and the rotor fluxes. This can give rise to large current surges, possibly more than the full-voltage locked rotor current, together with large transient torque oscillations, as much as five times full-load torque. Although the effects are only transitory, typically one fifth of a second, they are sources of stress and potential damage to the drive system, and where frequent starting is necessary, incur high maintenance costs.

There are methods of control, for example, the closed transition starter, which eliminate or reduce the reconnection transients. However, such starters are expensive and have reliability implications; for these reasons, they are not widely utilised. The star-delta starter also has disadvantages due to the restricted starting torque available. If 40% LRT is needed to breakaway, the motor size must be increased or direct-on-line starting is re-employed. Combined with the severe effects of the re-switching surges and the additional costs of bringing six cables from the motor to the starter instead of only three, star-delta starting offers a less than ideal solution to the problem of induction motor starting.

As a starting alternative, auto-transformers and wye-delta starters are large and require extra wiring.

Solid-state starting technology can overcome many of the problems associated with mechanical based starters and can provide stepless soft-starting of three-phase ac caged induction motors.

13.4.6iii Solid-state soft-starter

The solid-state switches, in figure 13.18 for example, are phase controlled in a similar manner to a light dimmer, in that they are turned on for a part of each ac cycle. The rms voltage is controlled by varying the conduction angle of the switches. Decreasing the delay angle, α (increasing the conduction angle), as shown in figure 13.32, increases the rms output voltage. Controlling the rms output voltage by means of solid-state switches has a number of advantages, one being the improvement in system efficiency, due to the low on-state voltage of SCR solid-state switches. Another advantage of the solid-state starter is that the rms voltage can be easily altered to suit the required starting conditions. By varying the conduction angle, the output voltage can be increased or reduced, and this can be achieved automatically by the control electronics. The control electronics can be pre-programmed to provide a particular output voltage contour based on a timed sequence (open loop), or can dynamically control the output voltage to achieve an output profile based on measurements of characteristics such as current and speed (closed loop).

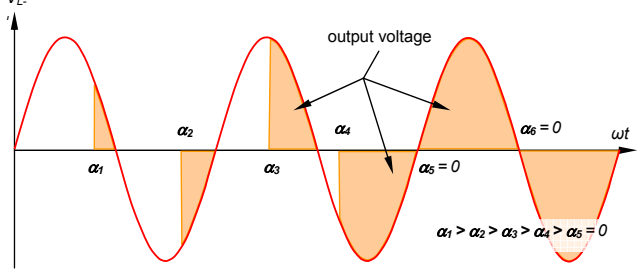


Figure 13.32. *Smoothly ramped-up motor voltage by controlling the SCR's firing angle, α.*

Switching elements

Voltage control is achieved by means of solid-state ac switches in series with one or more phases. The ac switch possibilities comprise any of the combinations in figure 13.33.

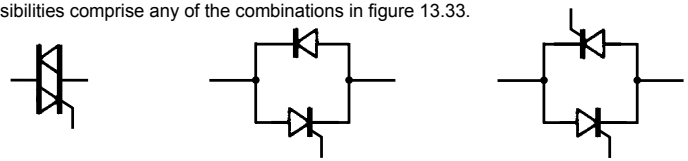


Figure 13.33. *Possible ac switch combinations per phase: (a) a triac; (b) SCR and diode reverse parallel; and (c) reverse parallel connected SCRs.*

The switching elements must be able to control the current applied to the motor at line voltage. In order to maintain high reliability, the switching elements need to be rated at least three times the line voltage. On a 400V ac supply, this means that the requirement is for 1200V (dc, bidirectional) devices, and 600V devices on a 200V ac supply. It is also important that the switching elements have a high transient current overload capacity. 1200V triacs with robust current transient overload characteristics are not readily available, so the choice is between the SCR-Diode and SCR-SCR for 400V ac applications.

The major differences between the SCR-SCR option in figure 13.33c and the SCR-Diode options in figure 13.33b are cost and the harmonic content of the output voltage. The SCR-SCR approach provides a symmetrical output which is technically desirable from the point of supply disturbances and harmonics, while the SCR-Diode combination is inferior technically, it is commercially more effective and easier to implement. Harmonic regulation requirements have drastically reduced the viability of SCR-Diode type soft starters.

The solid-state soft-starter can be designed to control

- one phase, reducing the torque but not the current in two phases, (SCR/Diode cannot be used in this connection), Fig. 13.34a, or
- two phases reducing the torque but the current will not be optimally reduced or balanced, there will be negative sequence currents heating the rotor and reducing the torque per unit start current, (SCR/Diode cannot be used in this connection), Fig. 13.34b, or
- three phases, reducing current and torque, providing the optimum results for torque generated per unit of start current, Fig. 13.34c. The SCR/diode combination can be used.

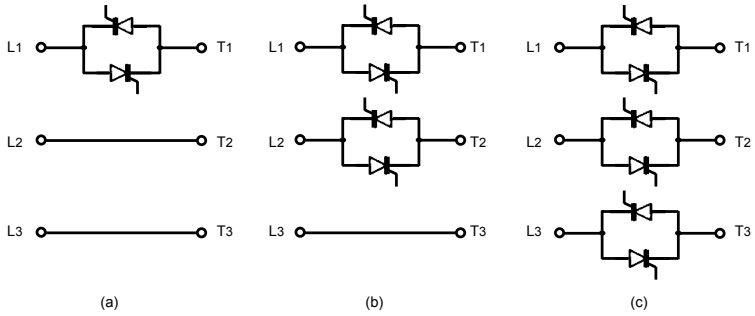


Figure 13.34. *Three possible line configurations:*
(a) single phase control; (b) control of two phases; and (c) three-phase fully-controlled regulator.

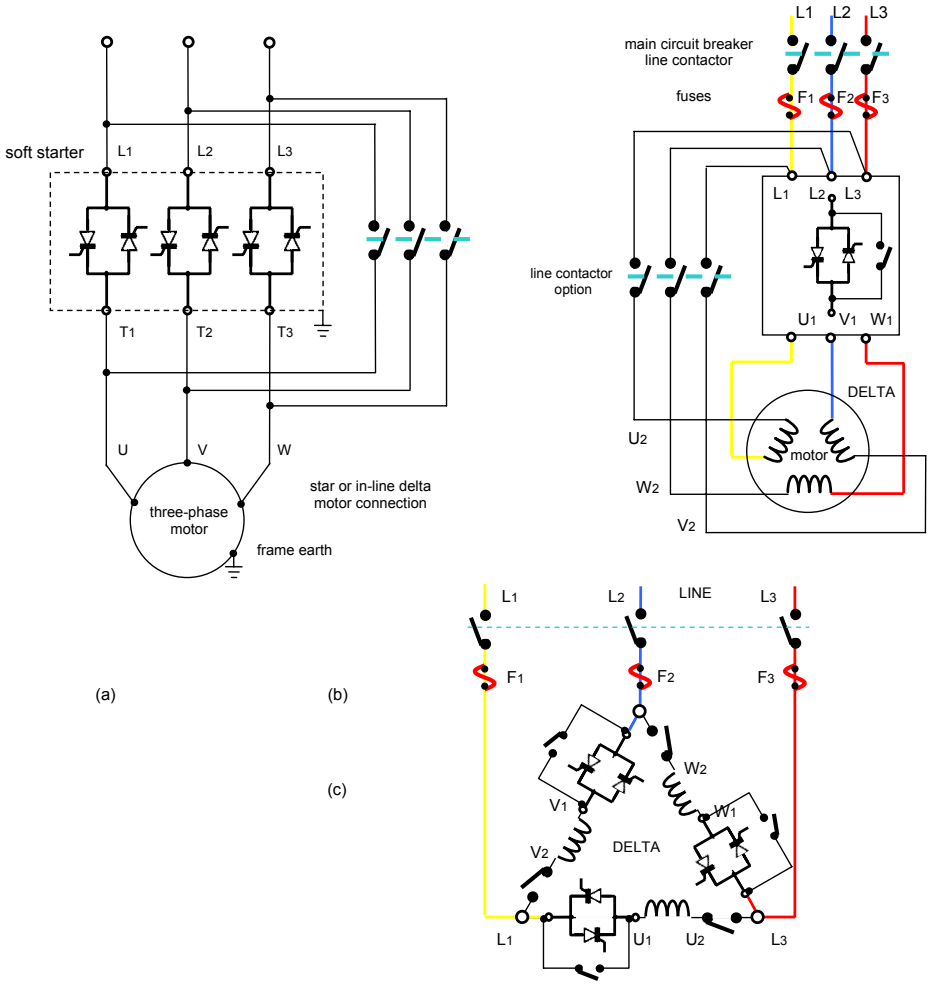


Figure 13.35. *Three-phase voltage control of caged three-phase ac induction motor: (a) line delta or star controlled and (b) and (c) control within an in-delta configuration, both with a bypass relay.*

Solid-state soft-starter arrangements

A reduced ac voltage can be delivered to a motor by controlling an SCR's firing angle, as illustrated in Figure 13.32. This SCR's firing angle control can ramp the voltage smoothly to full rating with the motor-controller connection configurations shown in Figure 13.35. Figure 13.36 shows that the in-rush current with electronic soft starters is much lower, as is the starting torque available from the motor, compared with DoL ans star-delta starter. Therefore, both voltage dips and mechanical shock are reduced considerably with solid-state soft starters.

In the in-delta circuit configuration in figures 13.35b/c the individual phases of the switching devices are connected in series with the individual motor windings (6 conductor connections as with the star-delta starter). The soft starter conducts about 58 % of the rated motor current. This allows the use of a significantly smaller device than the in-line approach, which only requires three motor connections, as shown in figure 13.35a.

By using a by-pass contactor across the semiconductor switches, as shown in figure 13.35a, the soft-starter power losses are reduced. Additionally, since the starter is only functional during in short infrequent start/stop periods, it is possible to reduce the starter enclosure size and use a higher IP-class since air ventilation is not required.

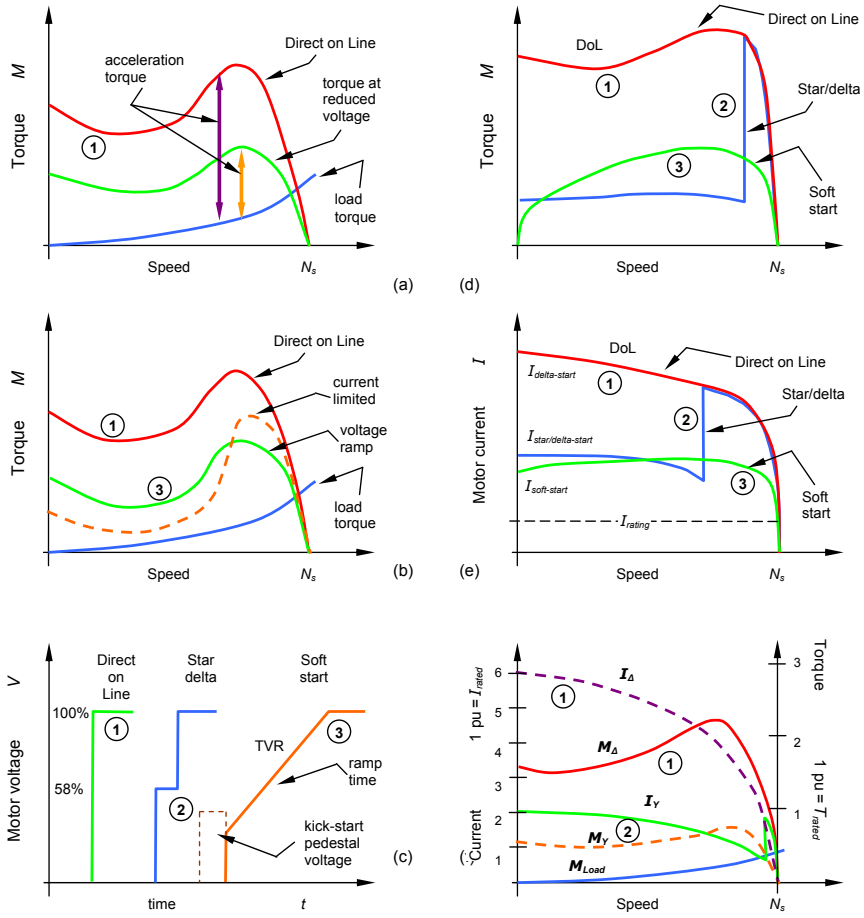


Figure 13.36. *Characteristics showing why solid-state soft starters significantly reduce voltage dip and mechanical shock.*

Solid-state soft starters can be connected in series with the line voltage applied to the motor (three-wire or standard connection) as in figure 13.35a, or can be connected inside the delta loop of a delta connected motor, controlling the voltage applied to each winding (six-wire or inside delta connection) as in figure 13.35b/c. There are usually three pairs of SCRs to control the voltage to a three-phase ac induction motor, that is, one pair for each phase, figure 13.35. Because SCRs are power components, they generate approximately 1W/A per phase of heat when on and the heat sink must dissipate the heat generated. Both the SCRs and the heat sink are components that add to the costs.

Because a three-phase motor is a three-wire system, the Kirchhoff sum of its three-phase currents is constrained to zero at any instant. If the currents in two of the three phases are reduced, the current to the third phase will be reduced as well, even when the third phase is directly connected to the full line voltage. A two-phase-controlled solid-sate soft starter based on figure 13.34b, is able to control the three-phase currents and since only two pairs of SCRs are used, it has a smaller heat sink. Fewer SCRs and a smaller heat sink reduce the cost and size. Compared to an electromechanical starter, it offers superior performance in a compact size, where the cost of parts, installation, and maintenance are lower.
A two-phase-controlled soft-starter can cause undesirable acoustical noise on larger motors at voltages less than 50%. The cause of the audible noise is related to the dc component in the phase current, which causes additional heating. Because of shorter starting times, motor heating is minimal. Polarity balancing control, balances the current in positive half and negative half cycles, eliminating the dc components. Then motors can be started at voltages less than 50% of full rating. This feature is particular applicable when soft starting a fan motor or pump motor at light load or no-load during the start period. Polarity balancing cannot balance the currents among the three phases and the phase without SCR control will have higher current. The imbalance between the three-phase currents is intrinsic to two-phase control and cannot be influenced. Because the imbalance among the three currents is generally within 10 to 25%, it is not critical in applications where the motor load reaches full speed quickly.
A three-phase fully-controlled soft-starter is applicable if balanced phase currents are essential to within 10%.

The functional block diagram in figure 13.37 shows a three-phase controlled soft-starter which offers features that include:

- polarity balancing control allowing the motor to start at less than 50% full voltage
- integrated ac motor thermal protection
- selectable motor overload trip level
- adjustable current limiting, start time, stop time, and starting voltage
- built in by-pass contactors
- detection of phase failure, faulty control voltage, locked rotor, SCR overheating, etc.

Design features particular to power electronics knowhow include the use of isolating pulse transformers for SCR triggering and R-C snubbers across the SCRs. Series snubber resistors are used not so much because of the necessary power rating but to achieve the required 1200V voltage rating. Two series resistors allows low cost, low voltage, low inductance resistors to be used.

13.4.6iv Soft-starter control and application

Open-loop control and the start voltage profile
Open-loop soft starters produce a start voltage profile which is independent of the current drawn, or the speed of the motor.
The start voltage profile is programmed to follow a predetermined contour against time, as shown in figure 13.36c. A basic Timed Voltage Ramp (TVR) system operates by applying an initial voltage to the motor, possibly involving a kick-start pedestal voltage, then slowly ramps from this voltage up to full voltage. On basic systems, the initial start voltage is not adjustable, but the ramp time is and may be a simple linear ramp or a complex shape to emulate a controlled current start. The voltage ramp time is referred to as the acceleration ramp time and is calibrated in seconds. This is not an accurate description, as it does not directly control the acceleration of the motor. Technically, it should be referred to as the voltage ramp time. On more sophisticated controllers, the start voltage is pre-setable, typically from 10% to 70% of full line voltage, and is set to achieve at least breakaway torque for the motor at start. There is no advantage in the motor stalling or straining to start due to insufficient torque. Eventually full voltage is applied under locked rotor load conditions, producing locked rotor torque and current which increases the heat dissipated in the motor, until any protection trips or failure. The starter does not have any programmed knowledge of the connected motor, so is unable to deliver a prescribed amount of torque under open loop conditions. The actual start torque produced is given by equation (13.119). The motor LRT can vary from as low as 60% FLT to as high as 350% FLT which is a range of almost 6 to 1.

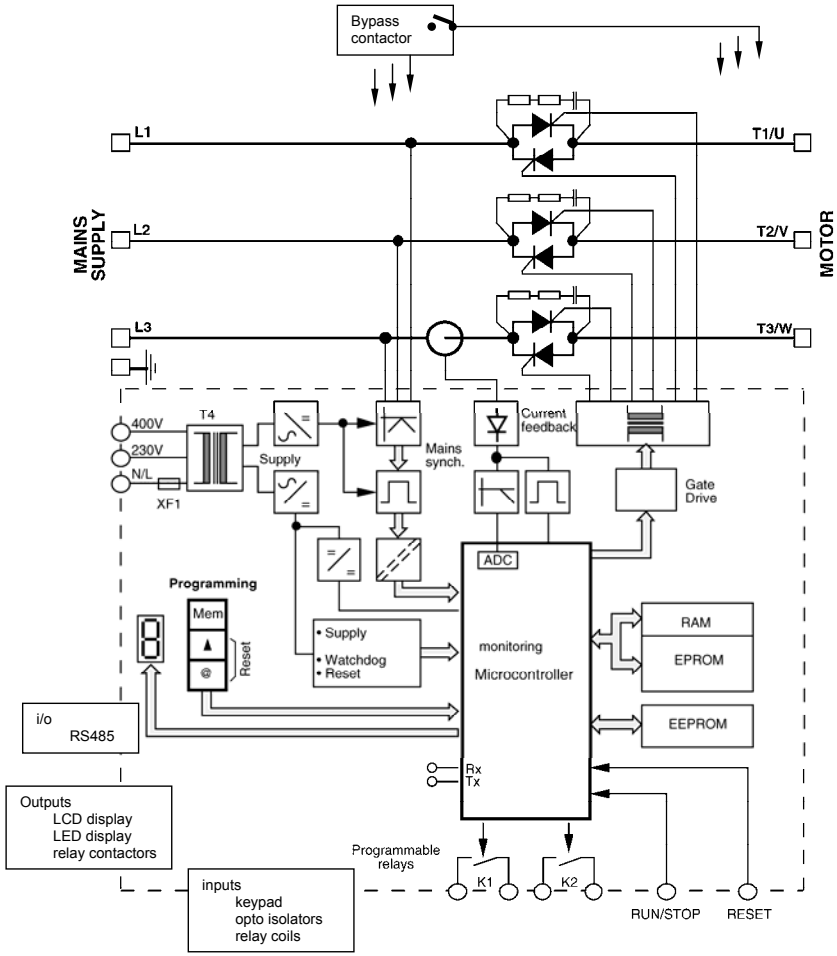


Figure 13.37. *Functional block diagram of three-phase controlled SCR based soft starter, with voltage and current feedback control, with optional bypass contactor.*

Closed-Loop Control
Closed-loop soft-starters monitor an output characteristic or effect from the starting action and dynamically modify the start voltage profile to cause the desired response. The most common closed loop soft starter is the controlled current soft starter where the current drawn by the motor during start is monitored and controlled to give either a constant current, as shown in figure 13.36b, or a current ramp soft start. Another closed loop strategy is the constant acceleration soft start where the motor speed is monitored by a tacho-generator or shaft encoder and the voltage is controlled to maintain a constant rate of acceleration or a linear increase in motor speed. Closed-loop control can take the following forms.

i. In basic closed loop systems, the soft starter is essentially a standard TVR soft starter with a ramp option where the current in one phase is monitored and compared to a set point. If the current exceeds the set point, the ramp is frozen until the current drops below that set point. At the other complexity extreme, a comprehensive closed loop soft starter monitors the current in two phases (effectively in all three phases) and dynamically changes the output voltage to correct the start current to the required profile. This system is able to both increase and reduce the start voltage to suit the control needs, and attempts to minimise any dc current component.

ii. A constant current starter commences at zero volts and rapidly increases the output voltage until the required current is delivered to the motor, and then adjusts the output voltage during motor starting until either full voltage is reached, or the motor overload protection operates. With a controlled current soft starter, the voltage reduction reduces as the motor accelerates due to the rising motor impedance. As the motor approaches full speed, the voltage rises quickly (against speed) to full voltage. When the torque curve for a motor started by a constant current starter is compared with that of a constant voltage starter such as an autotransformer starter, there is an increase in the torque as the motor accelerates with a constant current start. This is ideal because as the motor increases in speed, the actual load on the motor shaft increases. This characteristic enables a load to be started with a lower current on a soft starter than traditional starter methods. Constant current starters are ideal for high inertia loads, or loads with a near constant starting torque load requirements.

iii. The current ramp soft starter operates in the same manner as the constant current soft starter except that the current is ramped from an initial start current to a current limit setting over a period of time. The initial start current, current limit, and ramp time are all user adjustable to suit the application. Machines requiring a varying starting torque, such as load conveyers, or applications requiring a reduced initial torque such as pumping applications, or genset applications where the relatively slow application of current load will allow the genset to track the load, are examples where the current ramp soft start can be used to advantage over a constant-current soft-starter.

iv. In a torque control starter, the controller models the motor under high-slip and low-slip conditions and uses a mathematical model to calculate the current and shaft torque being produced by the motor. These are then used as a feed back source with the square law reduced start torque and current curves, given by equations (13.119) and (13.120), being used to control the start voltage applied to the motor. The torque curve generated by equation (13.119) can be superimposed onto the load speed torque curve, and provided the torque developed at all speeds exceeds the load torque, the motor accelerates to full speed. If the curves cross, the start current (or voltage) are increased to increase the motor starting torque. The difference between the developed and the load torques is the acceleration torque that accelerates the machine to full speed. A high acceleration torque may be desirable for a high inertia load in order to minimize the starting time.

Methods of stopping

Soft-stop

Soft-starters inherently incorporate soft-stop, which is the opposite to soft-start. The voltage is gradually reduced, reducing the torque capacity of the motor. The reduction of available torque causes the motor to decelerate when the motor shaft torque is less than the torque required by the load. As the torque is reduced, the speed of the load is reduced to the point where the load torque equals the shaft torque.
Typically, soft stop is achieved using an open loop voltage ramp, but a torque control soft stop system can use torque feedback to provide better deceleration control.
Open loop soft stop performance is dependent on the characteristics of the motor and driven load. On larger machines this can be non-linear, hence provides poor performance.
Soft slope effectively adds inertia to the load and extends the braking time. It should only be applied to installations where the stopping time is too short and needs to be extended. Soft stop does not provide braking, and occurs over a period longer than it would take the rotational system to coast to standstill without any power applied.

DC braking

DC braking is used to apply a braking torque to the motor and load, making them stop quicker.
Software controlled dc braking is possible for soft starters, but is not as effective as the braking that can be achieved with a specific dc brake electronic circuit.

Software

DC braking using the soft starter is achieved by turning on a positive SCR in one phase and a negative SCR on in a second phase for a small angle of each cycle. This causes a high pulse of dc current to flow through the motor windings and creates a stationary torque field in the stator. This causes the motor to slow down. The short pulses at line frequency also produce a synchronous component in the torque field that can limit the effectiveness close to synchronous speed. In some cases, a shorting contactor is connected across a motor winding to prolong the period of current flow and reduce the line frequency component.
During dc braking, the energy of the driven load is dissipated in the rotor of the motor.

Hardware

A stop button, with two interlocked contactors initiates braking, as shown in figure 13.38, where a dedicated thyristor/diode braking circuit is shown.

An induction motor with high-inertial load can be quickly stop by circulating dc current in a stator winding, where any two stator terminals can be connected to a dc source such that the resultant dc current produces stationary N-S poles in the stator. Since the number of stationary poles is the same as the number of rotating poles normally produced with ac currents, as the rotor bars sweeps past the dc field, an ac rotor voltage is induced.
The I^2R loss produced in the rotor circuit is converted kinetic energy stored previously in the rotating masses, hence the motor comes to rest by dissipating all the kinetic energy as heat.
The benefit of dc braking is that efficient heat is produced, since the dissipated rotor losses are equal to the kinetic energy of the rotating masses and are independent of the dc current magnitude, while the braking torque is proportional to the square of the dc braking current.
DC injection duration and frequency of occurrence should be minimised in order to minimise motor heating.

The dc injection braking procedure shown in figure 13.38b, is as follows:

- The motor contactor C_{motor} is opened, then after a delay of 200ms to 2.5s (increased time as motor rating increases), the braking contactor C_{brake} is closed, which allows the motor back emf to reduce. Any overlap between C_{motor} and C_{brake} is prevented by using interlocked contactors.
- After a further 50ms delay, dc current is injected into two motor winding by firing the braking thyristor, until rotation stops. This second 50ms delays allows an ac breaking contactor C_{brake} to be used, since dc current switching is avoided. The braking torque is a function of dc current, which is controlled by the thyristor firing angle.
- After the thyristor triggering is removed (always before or when to rotor comes to a standstill), a delay of 200ms to 2s (increased time as motor rating increases) is allowed before the braking contactor C_{brake} is opened, in order to avoid braking a dc current.
- 200ms after the braking contactor C_{brake} is opened, the motor can be enabled by closing the motoring contactor, C_{motor}.

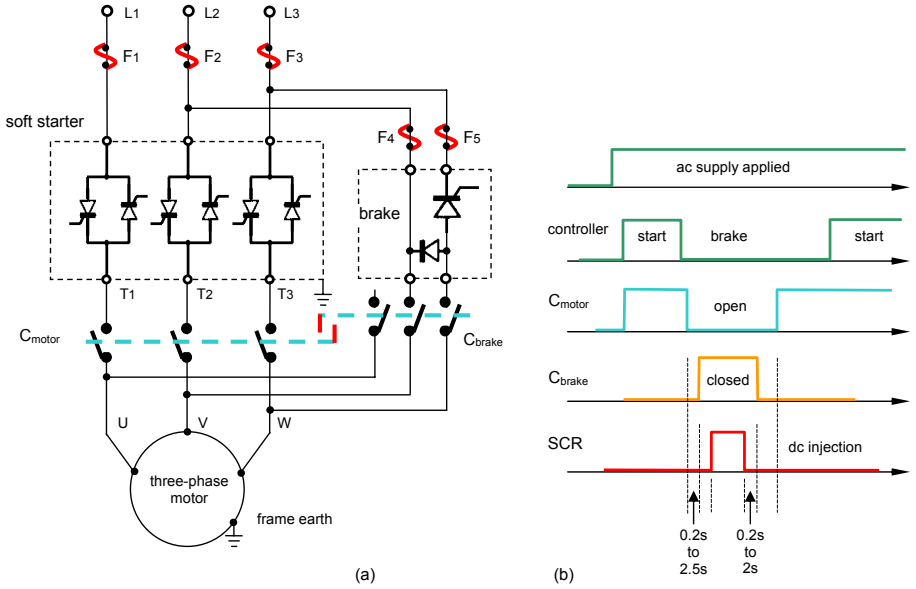


Figure 13.38. *Braking circuit: (a) circuit connection and (b) timing sequence and delay times.*

Reversing and plugging

A mechanical contactor based reversing arrangement is shown in figure 13.39a, which uses two interlocked contactors. Contactors and fusing for reversing circuits are only used between the line and the soft-stater.
It is required, with this contactor arrangement, to insert a 150 to 350ms delay between the opening of one contactor and the closing of the other, to allow any residual flux in the rotor to die away.

Figure 13.39c shows the machine torque and current conditions when the machine is reversed whilst operating at or near rated torque/speed, point A. On reversing two phase connections, the operating point A moves to point B, which represents a deceleration torque with a high machine current, point C. The machine accelerates to operating point D as a result of stopping and reversing rotation in a direction opposite to the original rotating direction. In attempting to traverse from point A to point D, the rotation passes through zero speed point E, at which time the controller is phased back to zero thyristor conduction. When plug-braking, (phase reversal at full voltage) there should be some form of zero-speed detection to stop the drive after braking has been completed, otherwise the drive may either accelerate the motor in the reverse direction or switch off before zero speed has been reached.
A solid state solution to reversing and plugging is shown in figure 13.38b, which requires two extra sets of back to back parallel connected thyristors.

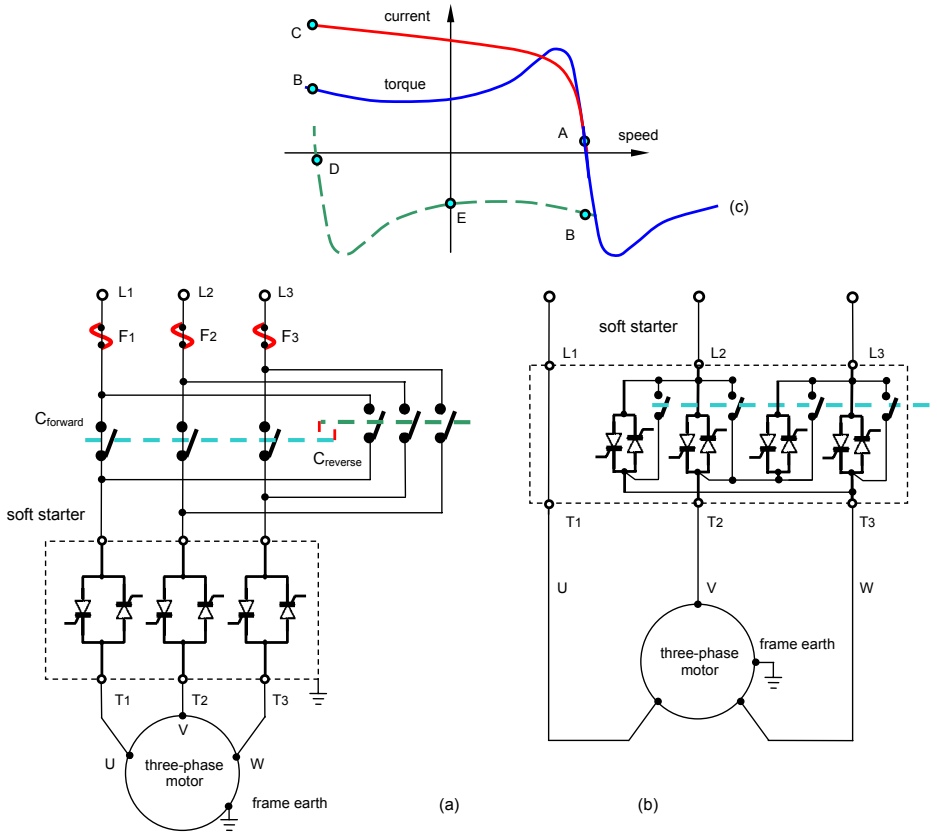


Figure 13.39. *Reversing circuit:*
(a) mechanical contactor reversing; (b) electronic reversing (4 quadrant) with mechanical contactor bypassing; and (c) torque and current characteristic of reversal of two phase voltages.

Ratings
As the rating of the soft-starter is essentially thermal, there is a strong relationship between the start time, start current, start/stop frequency of occurrence, ambient temperature, off-time, and the rating of the starter. Typically, the thermal inertia of the SCR heatsink assembly is quite long so there is not a large variation in the rating between say a 10-second rating and a 30-second rating.
At altitudes above 1000m the rated current is usually derated at about 7% per 1000m increase in altitude, up to 4000m.

Harmonics
Harmonics are unwanted voltages and currents existing in almost every electrical system and are a multiple of the rated ac mains frequency. Typical harmonics are odd, viz., 3rd, 5th, 7th, 9th etc., which contribute to the unnecessary heating of motors, cables and other equipment and may shorten the lifetime of these devices if exposed for a long period of time.
The resultant EMC may disturb other local electronic systems. Soft-starters generally fulfil EMC directives (EN 60947-4-2) on emission (EN55011 Class A) and immunity (IEC 6 1000-4/ 2 to 6) since their operation is non-continuous and intermittent.

Fuses
Standard IEC 60947-4-2 defines two types of co-ordination according to the expected level of service continuity. Co-ordination requires that, under short-circuit conditions:

> ***Type 1***: The soft-starter device shall cause no danger to persons or installation and may not be suitable for further service without repair and replacement of parts.
>
> ***Type 2***: The soft-starter device shall cause no danger to persons or installation and shall be suitable for continued use. For hybrid controllers and starters, contact welding is a possibility, in which case equipment maintenance is required.

Semiconductor fuse curves do not follow the ratings curves for soft starters and only offer short circuit protection. Semi-conductor fuses (high speed fuses) are the only type of fuses that are fast enough to achieve type 2 co-ordination when using a softstarter. A separate overload relay for motor protection is required in combination with this type of fuse. If replacing the semi-conductor fuses with an MCB, protection reverts to type 1 co-ordination.

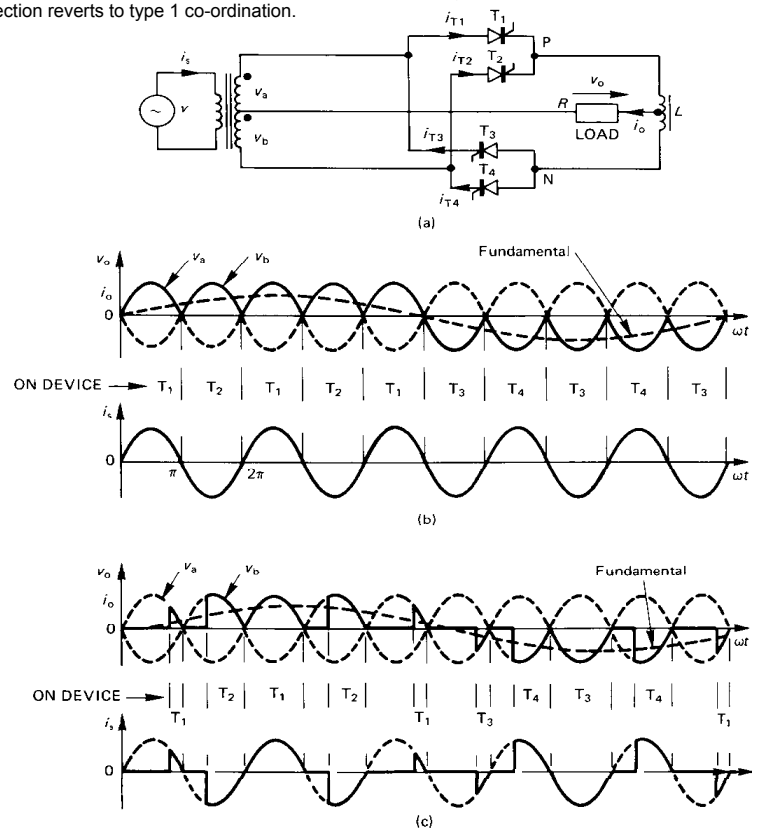


Figure 13.40. *Single-phase cycloconverter ac regulator: (a) circuit connection with a purely resistive load; (b) load voltage and supply current with 180° conduction of each thyristor; and (c) waveforms when phase control is used on each thyristor.*

13.5 Cycloconverter

The simplest cycloconverter is a single-phase, two-pulse, ac input to single-phase ac output circuit as shown in figure 13.40a. It synthesises a low-frequency ac output from selected portions of a higher-frequency ac voltage source and consists of two converters connected back-to-back. Thyristors T_1 and T_2 form the positive converter group P, while T_3 and T_4 form the negative converter group N.

Figure 13.40b shows how an output frequency of one-fifth of the input supply frequency is generated. The P group conducts for five half-cycles (with T_1 and T_2 alternately conducting), then the N group conducts for five half-cycles (with T_3 and T_4 alternately conducting). The result is an output voltage waveform with a fundamental of one-fifth the supply with continuous load and supply current.
The harmonics in the load waveform can be reduced and rms voltage controlled by using phase control as shown in figure 13.40c. The phase control delay angle is greater towards the group changeover portions of the output waveform. The supply current is now distorted and contains a subharmonic at the cycloconverter output frequency, which for figure 13.40c is at one-fifth the supply frequency.

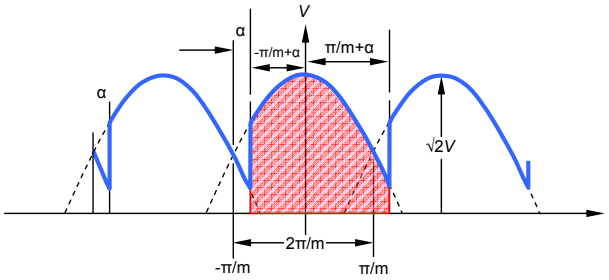


Figure 13.41. *Output voltage from an m-phase converter with firing delay angle α.*

In figure 13.41, if the firing delay angle α, the conduction period is from $-\pi/m + \alpha$ to $+\pi/m + \alpha$, such that the conduction period is 2 π/m. The average output voltage is

$$V_{dc} = \frac{m}{2\pi}\int_{-\frac{\pi}{m}+\alpha}^{\frac{\pi}{m}+\alpha}\sqrt{2}V\cos\omega t\, d\omega t = \sqrt{2}V\frac{m}{\pi}\sin\frac{m}{\pi}\times\cos\alpha$$

For the cycloconverter output

$$v_o(t) = \sqrt{2}V_o\sin\omega_o t$$

Equating gives

$$\cos\alpha = \frac{\sqrt{2}V_o}{\sqrt{2}V\frac{m}{\pi}\sin\frac{m}{\pi}}\sin\omega_o t = \rho\sin\omega_o t$$

That is

$$\alpha = \cos^{-1}(\rho\sin\omega_o t)$$

where ρ is the output voltage modulation factor.

With inductive loads, one blocking group cannot be turned on until the load current through the other group has fallen to zero, otherwise the supply will be short-circuited. An intergroup reactor, *L*, as shown in figure 13.40a can be used to limit any inter-group circulating current, and to maintain a continuous load current.

A single-phase ac load fed from a three-phase ac supply, and three-phase ac load cycloconverters can also be realised as shown in figures 13.42a and both of 13.42b and c, respectively. A transformer is needed in figure 13.42a, if neutral current is to be avoided. The three-pulse per ac cycle cycloconverter in figure 13.42b uses 18 thyristors, while the 6-pulse cycloconverter in figure 13.42c uses 36 thyristors (inter-group reactors are not shown), where the load (motor) neutral connection is optional. The output frequency, with considerable harmonic content, is limited to about 40% of the input frequency, and motor reversal and regeneration are achievable.

If a common neutral is used, no transformer is necessary. Most cycloconverters are 6-pulse, and the neutral connection in figure 13.42c removes the zero sequence component.

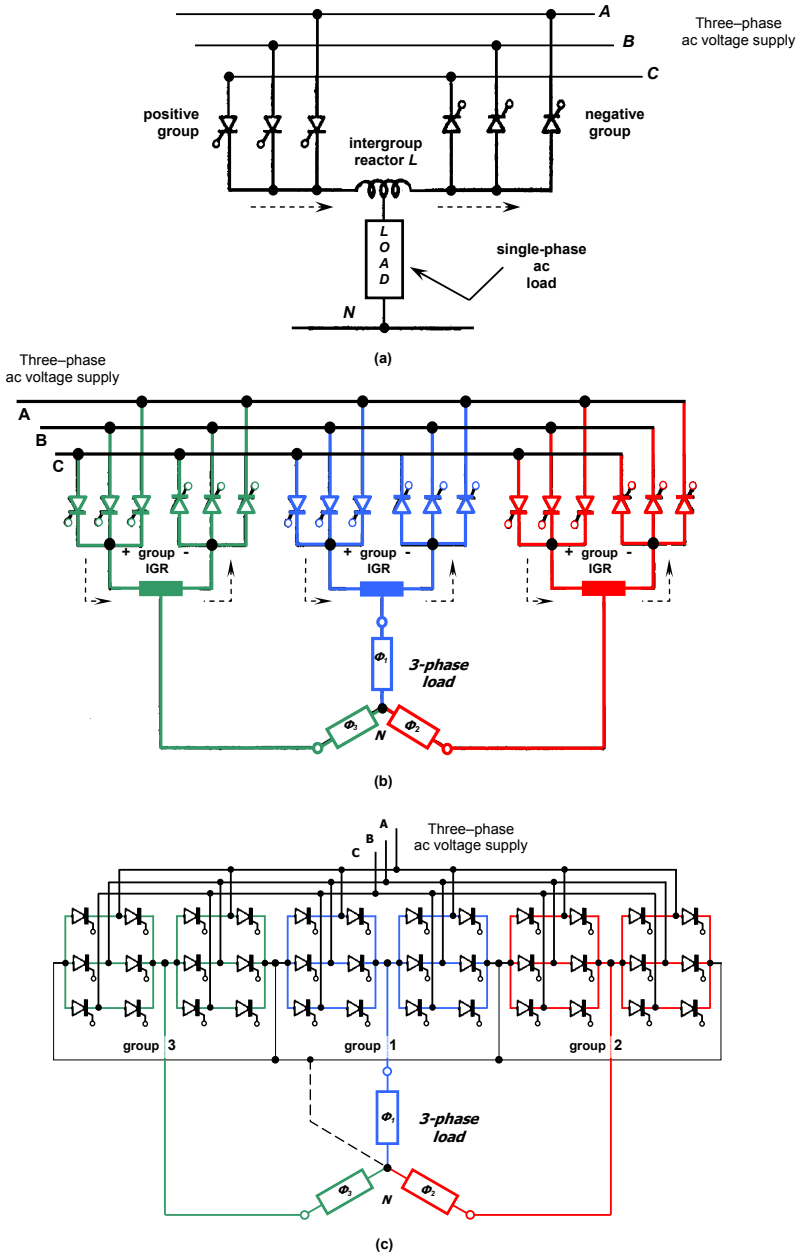


Figure 13.42. *Cycloconverter ac regulator circuits:*
(a) three-phase to single-phase; and three–phase supply to three-phase load (b) 3-pulse without neutral connection; and (c) 6-pulse with optional load neutral connection.

The positive features of the cycloconverter are

- Natural commutation
- No intermediate energy storage stage
- Inherently reversible current and voltage

The negative features of the cycloconverter are

- High harmonics on the input and output
- Requires at least 18 thyristors usually 36
- High reactive power

13.6 The matrix converter

Commutation of the cycloconverter switches is restricted to natural commutation instances dictated by the supply voltages. This usually results in the output frequency being significantly less than the supply frequency if a reasonable low harmonic output is required. In the matrix converter in figure 13.43c, the thyristors in figure 13.42b are replaced with fully controlled, bidirectional switches, like those shown in figures 13.43a and b. Rather than eighteen switches and eighteen diodes, nine switches and thirty-six diodes can be used if a unidirectional voltage and current switch in a full-bridge configuration is used as shown in figure 6.11. These switch configurations allow converter current commutation as and when desired, provide certain conditions are fulfilled. These switches allow any one input supply ac voltage and current to be directed to any one or more of the output lines. At any instant, only one of the three input voltages can be connected to a given output. This flexibility implies a higher quality output voltage can be attained, with enough degrees of freedom to ensure the input currents are sinusoidal and with unity (or adjustable) power factor. If the inputs are voltage sources, the outputs must be current sources, and vice versa. The input *L-C* filter prevents matrix modulation frequency components from being injected into the input three-phase ac supply system.

In the usual case of voltage source inputs and current source outputs, the switch conditions must:

- Never short-circuit two or more input phase voltages
- Never open circuit any output line current

Generally, the relationship between the n output voltages (v_a, v_b, v_c, .. v_n) and the m input voltages (v_A, v_B, v_C, .. v_M) is determined by the states of the $M \times n$ bidirectional switches ($S_{i,j}$), where $S_{i,j} = 1$ = closed, $S_{i,j} = 0$ = open, according to

$$\begin{bmatrix} v_a \\ v_b \\ v_c \\ \vdots \\ v_n \end{bmatrix} = \begin{bmatrix} S_{Aa} & S_{Ba} & S_{Ca} & \cdots & S_{Ma} \\ S_{Ab} & S_{Bb} & S_{Cb} & \cdots & S_{Mb} \\ S_{Ac} & S_{Bc} & S_{Cc} & \cdots & S_{Mc} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ S_{An} & S_{Bn} & S_{Cn} & \cdots & S_{Mn} \end{bmatrix} \begin{bmatrix} v_A \\ v_B \\ v_C \\ \vdots \\ v_M \end{bmatrix} \qquad \text{(V)} \qquad V_{out} = S V_{in} \qquad (13.121)$$

where S is the switch connection matrix and $i = A, B, \ldots M$ and $j = a, b, \ldots n$.

If the n inputs are voltage sources, then the switches must satisfy

$$\sum_{i=A}^{M} S_{ia} = \sum_{i=A}^{M} S_{ib} = \sum_{i=A}^{M} S_{ic} = \ldots\ldots \sum_{i=A}^{M} S_{in} = 1 \qquad \sum_{i=A}^{M} S_{ia} \sum_{j=a}^{n} S_{ij} = n \qquad (13.122)$$

The first set of equalities in equation (13.122) ensure that each output can only be connected (at most) to one input voltage supply, thus avoiding shorting two or more voltage inputs. Since the inputs are voltage sources, the output must be current sources, thus the second equality ensures a path for current in each of the n output phases. The relationships between the input and output currents, in terms of the switch connection matrix S, are given by

$$\begin{bmatrix} i_A \\ i_B \\ i_C \\ \vdots \\ i_M \end{bmatrix} = \begin{bmatrix} S_{Aa} & S_{Ba} & S_{Ca} & \cdots & S_{Ma} \\ S_{Ab} & S_{Bb} & S_{Cb} & \cdots & S_{Mb} \\ S_{Ac} & S_{Bc} & S_{Cc} & \cdots & S_{Mc} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ S_{An} & S_{Bn} & S_{Cn} & \cdots & S_{Mn} \end{bmatrix}^T \begin{bmatrix} i_a \\ i_b \\ i_c \\ \vdots \\ i_n \end{bmatrix} \qquad \text{(A)} \qquad I_{in} = S^T V_{out} \qquad (13.123)$$

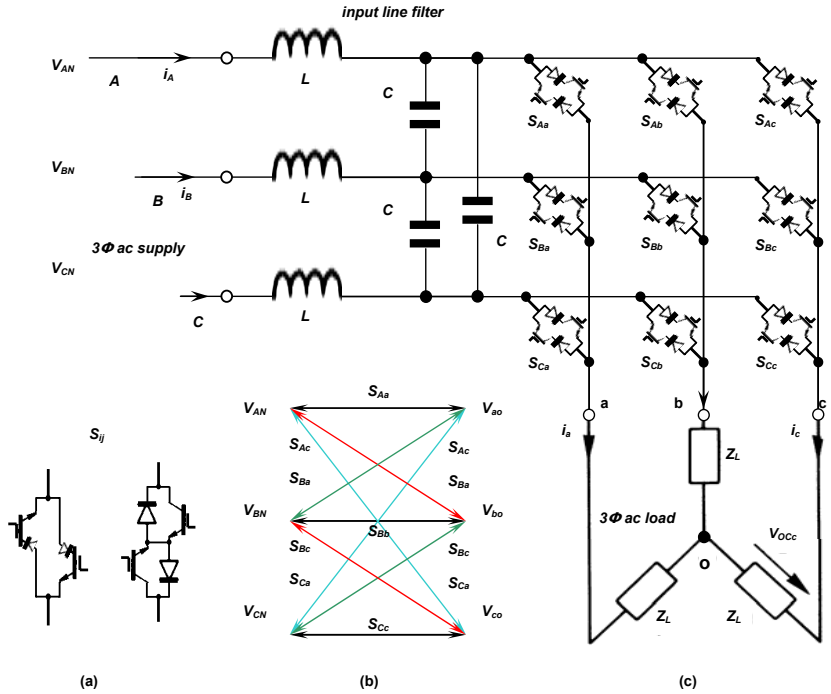


Figure 13.43. *Three-phase input to three-phase output matrix converter circuit: bidirectional switches (a) reverse blocking igbts conventional igbts; (b) switching matrix; and (c) three–phase ac supply to three-phase ac load.*

For the three-phase voltage input to three-phase current output matrix converter, the relationship between the output voltages (v_a, v_b, v_c) and the input voltages (v_A, v_B, v_C) is determined by the states of the nine bidirectional switches ($S_{i,j}$), where $S_{i,j} = 1$ = closed, $S_{i,j} = 0$ = open, according to

$$\begin{pmatrix} v_a \\ v_b \\ v_c \end{pmatrix} = \begin{pmatrix} S_{Aa} & S_{Ba} & S_{Ca} \\ S_{Ab} & S_{Bb} & S_{Cb} \\ S_{Ac} & S_{Bc} & S_{Cc} \end{pmatrix} \begin{pmatrix} v_A \\ v_B \\ v_C \end{pmatrix} \qquad \text{(V)} \qquad V_{out} = S V_{in} \qquad (13.124)$$

From Kirchhoff's voltage law, the number of switches on in each row must be either one or none, otherwise at least one input supply is shorted, that is (i refers to the input and j refers to the output)

$$\sum_{i=1}^{3} S_{ij} \le 1 \quad \text{for any } j \qquad (13.125)$$

With the balanced star load shown in figure 13.43c, the load neutral voltage v_o is given by

$$v_o = \tfrac{1}{3}\left(v_a + v_b + v_c\right) \qquad (13.126)$$

The line-to-neutral and line-to-line voltages are the same as those applicable to svm (space voltage modulation, Chapter 14.1.3vii), namely

$$\begin{pmatrix} v_{ao} \\ v_{bo} \\ v_{co} \end{pmatrix} = \frac{1}{6} \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix} \begin{pmatrix} v_a \\ v_b \\ v_c \end{pmatrix} \qquad \text{(V)} \qquad (13.127)$$

from which

$$\begin{pmatrix} v_{ab} \\ v_{bc} \\ v_{ca} \end{pmatrix} = \tfrac{1}{2}\begin{pmatrix} 1 & -1 & -1 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}\begin{pmatrix} v_a \\ v_b \\ v_c \end{pmatrix} \quad \text{(V)} \tag{13.128}$$

Similarly the relationship between the input line currents (i_A, i_B, i_C) and the output currents (i_a, i_b, i_c) is determined by the states of the nine bidirectional switches ($S_{i,j}$), according to

$$\begin{pmatrix} i_A \\ i_B \\ i_C \end{pmatrix} = \begin{pmatrix} S_{Aa} & S_{Ba} & S_{Ca} \\ S_{Ab} & S_{Bb} & S_{Cb} \\ S_{Ac} & S_{Bc} & S_{Cc} \end{pmatrix}^T \begin{pmatrix} i_a \\ i_b \\ i_c \end{pmatrix} \quad \text{(A)} \qquad I_{in} = S^T I_{out} \tag{13.129}$$

where the switches S_{ij} are constrained such that no two or three switches short between the input lines or cause discontinuous output current. Discontinuous output current must not occur since no natural default current freewheel paths exist. The input short circuit constraint is complied with by ensuring that only one switch in each row of the 3×3 matrix in equation (13.124) (hence row in equation (13.129)) is on at any time, viz., equation (13.125), while continuous load current in equation (13.129) (hence column in equation (13.124)) is ensured by Kirchhoff's current law, that is

$$\sum_{j=1}^{3} S_{ij} \geq 1 \quad \text{for at least any two } i \tag{13.130}$$

More than one switch on in a column implies that an input phase is parallel feeding more than one output phase, which is allowable.

If each switch *Sij* is modulated *m(t)* (usually sinusoidally in time), the low frequency relationships between the input and output currents and voltages are given by (*M* is the low frequency transfer matrix)

$$\begin{pmatrix} v_a \\ v_b \\ v_c \end{pmatrix} = \begin{pmatrix} m_{Aa}(t) & m_{Ba}(t) & m_{Ca}(t) \\ m_{Ab}(t) & m_{Bb}(t) & m_{Cb}(t) \\ m_{Ac}(t) & m_{Bc}(t) & m_{Cc}(t) \end{pmatrix}\begin{pmatrix} v_A \\ v_B \\ v_C \end{pmatrix} \quad \text{(V) or} \quad V_{out} = M(t)v_i \tag{13.131}$$

$$\begin{pmatrix} i_A \\ i_B \\ i_C \end{pmatrix} = \begin{pmatrix} m_{Aa}(t) & m_{Ba}(t) & m_{Ca}(t) \\ m_{Ab}(t) & m_{Bb}(t) & m_{Cb}(t) \\ m_{Ac}(t) & m_{Bc}(t) & m_{Cc}(t) \end{pmatrix}^T \begin{pmatrix} i_a \\ i_b \\ i_c \end{pmatrix} \quad \text{(A) or} \quad I_{in} = M(t)^T i_o \tag{13.132}$$

Thus given Kirchhoff's voltage and current law constraints, not all the 512 (2^9) states for nine switches can be used, and only 27 states, in three groups as summarized in Table 13.2, of the switch matrix can be utilised.

- The first group, of six combinations, allows each output phase to be connected to a different input phase.
- The second group (with 3 subgroups, each with 6 combinations), of 3x6 = 18 combinations, is when two output phases connect to the same input phase (two output phases shorted).
- The third group, of three combinations, is when all the output line voltages are zero, shorted.

For sinusoidal input phase voltages of frequency ω_i and maximum voltage V_{imax}

$$v_i = \begin{pmatrix} v_A \\ v_B \\ v_C \end{pmatrix} = V_{i\max}\begin{pmatrix} \cos\omega_i t \\ \cos\omega_i t - 120^\circ \\ \cos\omega_i t + 120^\circ \end{pmatrix} \tag{13.133}$$

If sinusoidal output line to line voltages are generated at frequency ω_o and relative displacement φ_o, then, neglecting non-fundamental components

$$v_o = \begin{pmatrix} v_{ab} \\ v_{bc} \\ v_{ca} \end{pmatrix} = \sqrt{3}\, V_{o\max}\begin{pmatrix} \cos\omega_o t - \varphi_o + 30^\circ \\ \cos\omega_o t - \varphi_o + 30^\circ - 120^\circ \\ \cos\omega_o t - \varphi_o + 30^\circ + 120^\circ \end{pmatrix} \tag{13.134}$$

The instantaneous relationship between the input phase voltages and output line voltages is

$$\begin{aligned} v_o &= v_i T_{ph-L} \\ v_o &= v_i\, m \begin{pmatrix} \cos\omega_o t - \varphi_o + 30^\circ \\ \cos\omega_o t - \varphi_o + 30^\circ - 120^\circ \\ \cos\omega_o t - \varphi_o + 30^\circ + 120^\circ \end{pmatrix}\begin{pmatrix} \cos\omega_i t \\ \cos\omega_i t - 120^\circ \\ \cos\omega_i t + 120^\circ \end{pmatrix}^T \\ &= V_{i\max}\, m \begin{pmatrix} \cos\omega_o t - \varphi_o + 30^\circ \\ \cos\omega_o t - \varphi_o + 30^\circ - 120^\circ \\ \cos\omega_o t - \varphi_o + 30^\circ + 120^\circ \end{pmatrix} \times \tfrac{3}{2}\cos\varphi_i \\ &= \sqrt{3}\, V_{o\max}\begin{pmatrix} \cos\omega_o t - \varphi_o + 30^\circ \\ \cos\omega_o t - \varphi_o + 30^\circ - 120^\circ \\ \cos\omega_o t - \varphi_o + 30^\circ + 120^\circ \end{pmatrix} \end{aligned} \tag{13.135}$$

where *m* is the modulation index, φ_i is the input displacement factor, and *T* is the instantaneous transfer matrix.

Table 13.2: Three-phase voltage to three-phase current matrix converter switch combinations

Group	S_{Aa} S_{Ab} S_{Ac}	S_{Ba} S_{Bb} S_{Bc}	S_{Ca} S_{Cb} S_{Cc}	A B C	v_{ab} v_{bc} v_{ca}	i_A i_B i_C
I	1 0 0	0 1 0	0 0 1	A B C	v_{AB} v_{BC} v_{CA}	i_a i_b i_c
I	1 0 0	0 0 1	0 1 0	A C B	$-v_{AB}$ $-v_{CA}$ $-v_{BC}$	i_b i_a i_c
I	0 1 0	1 0 0	0 0 1	B A C	$-v_{AB}$ $-v_{CA}$ $-v_{BC}$	i_b i_a i_c
I	0 1 0	0 0 1	1 0 0	B C A	v_{BC} v_{CA} v_{AB}	i_c i_a i_b
I	0 0 1	1 0 0	0 1 0	C A B	v_{CA} v_{AB} v_{BC}	i_b i_c i_a
I	0 0 1	0 1 0	1 0 0	C B A	$-v_{BC}$ $-v_{AB}$ $-v_{CA}$	i_c i_b i_a
II-A	1 0 0	0 0 1	0 0 1	A C C	$-v_{CA}$ 0 v_{CA}	i_a 0 $-i_a$
II-A	0 1 0	0 0 1	0 0 1	B C C	v_{BC} 0 $-v_{BC}$	0 $-i_a$ i_a
II-A	0 1 0	1 0 0	1 0 0	B A A	$-v_{AB}$ 0 $-v_{AB}$	i_a i_a 0
II-A	0 0 1	1 0 0	1 0 0	C A A	v_{CA} 0 $-v_{CA}$	$-i_a$ 0 i_a
II-A	0 0 1	0 1 0	0 1 0	C B B	$-v_{BC}$ 0 v_{BC}	0 $-i_a$ i_a
II-A	1 0 0	0 1 0	0 1 0	A B B	v_{AB} 0 $-v_{AB}$	i_a $-i_a$ 0
II-B	0 0 1	1 0 0	0 0 1	C A C	$-v_{CA}$ $-v_{CA}$ 0	i_b 0 $-i_b$
II-B	0 0 1	0 1 0	0 0 1	C B C	$-v_{BC}$ v_{BC} 0	0 i_b - i_b
II-B	1 0 0	0 1 0	1 0 0	A B C	v_{AB} $-v_{AB}$ 0	$-i_b$ i_b 0
II-B	1 0 0	0 0 1	1 0 0	A C A	$-v_{CA}$ v_{CA} 0	$-i_a$ 0 i_b
II-B	0 1 0	0 0 1	0 1 0	B C B	v_{BC} $-v_{BC}$ 0	0 $-i_b$ i_b
II-B	0 1 0	1 0 0	0 1 0	B A B	$-v_{AB}$ v_{AB} 0	i_a $-i_b$ 0
II-C	0 0 1	0 0 1	1 0 0	C C A	0 v_{CA} $-v_{CA}$	i_c 0 $-i_c$
II-C	0 0 1	0 0 1	0 1 0	C C B	0 $-v_{BC}$ v_{BC}	0 i_c $-i_c$
II-C	1 0 0	1 0 0	0 1 0	A A B	0 v_{AB} $-v_{AB}$	$-i_c$ i_c 0
II-C	1 0 0	1 0 0	0 0 1	A A C	0 $-v_{CA}$ v_{CA}	$-i_c$ 0 i_c
II-C	0 1 0	0 1 0	0 0 1	B B C	0 v_{BC} $-v_{BC}$	0 $-i_c$ i_c
II-C	0 1 0	0 1 0	1 0 0	B B A	0 $-v_{AB}$ v_{AB}	i_c $-i_c$ 0
III	1 0 0	1 0 0	1 0 0	A B C	0 0 0	0 0 0
III	0 1 0	0 1 0	0 1 0	A B C	0 0 0	0 0 0
III	0 0 1	0 0 1	0 0 1	A B C	0 0 0	0 0 0

By equating co-efficients in equation (13.135), the magnitude of the output line to line voltage, in terms of the input phase voltage magnitude is given by

$$V_{o\max} = \tfrac{1}{2}\sqrt{3} \times m \times V_{i\max}\cos\varphi_i \tag{13.136}$$

The maximum voltage gain, when *m* = 1 and unity input displacement pf, $\cos\varphi_i = 1$, the ratio of the peak fundamental ac output voltage to the peak ac input voltage is ½√3 = 0.866. Above this level, called over-modulation, distortion of the input current occurs.

Expressions similar to equations (13.133) to (13.136) are applicable to the input and output currents.

The output current is

$$i_o = \begin{pmatrix} i_{ab} \\ i_{bc} \\ i_{ca} \end{pmatrix} = \frac{I_{omax}}{\sqrt{3}} \begin{pmatrix} cos\omega_o t - \varphi_o + 30° - \varphi_L \\ cos\omega_o t - \varphi_o + 30° - 120° - \varphi_L \\ cos\omega_o t - \varphi_o + 30° + 120° - \varphi_L \end{pmatrix} \quad (13.137)$$

where $\cos\varphi_L$ is the load power factor and I_{omax} is the amplitude of the output line current.

The relationship between the input and output line currents is

$$i_i = \begin{pmatrix} i_A \\ i_B \\ i_C \end{pmatrix} = T^T_{ph-L} i_o$$
$$= ½\sqrt{3}\, I_{omax} \cos\varphi_L \begin{pmatrix} cos\omega_o t - \varphi_i \\ cos\omega_o t - \varphi_i - 120° \\ cos\omega_o t - \varphi_i + 120° \end{pmatrix} \quad (13.138)$$

By equating co-efficients, the relationship between the input and output current magnitudes is

$$I_{imax} = ½\sqrt{3}\, I_{omax} \cos\varphi_L \quad (13.139)$$

Additionally the input power must equal the output power, that is

$$P_i = \sqrt{3} v_i i_i \cos\varphi_i = \sqrt{3} v_o i_o \cos\varphi_o = P_o \quad (13.140)$$

Since the switches are bidirectional and fully controlled, power flow can be bidirectional. Control involves the use of a modulation index, $0 \le m \le 1$, that varies sinusoidally.

Modulation strategies

If the input currents and output voltages are to be sinusoidal, from equations (13.138) and (13.135), respectively, the voltage gain $q=V_{omax}/V_{imax}$ between the output and input voltages is given by

$$M_1 = ⅓ \begin{pmatrix} 1+2q\cos\omega_m t & 1+2q\cos\omega_m t - 120° & 1+2q\cos\omega_m t + 120° \\ 1+2q\cos\omega_m t + 120° & 1+2q\cos\omega_m t & 1+2q\cos\omega_m t - 120° \\ 1+2q\cos\omega_m t - 120° & 1+2q\cos\omega_m t + 120° & 1+2q\cos\omega_m t \end{pmatrix} \quad (13.141)$$

where $\omega_m = \omega_o - \omega_i$ such that $\varphi_i = \varphi_o$ and

$$M_2 = ⅓ \begin{pmatrix} 1+2q\cos\omega_m t & 1+2q\cos\omega_m t - 120° & 1+2q\cos\omega_m t + 120° \\ 1+2q\cos\omega_m t - 120° & 1+2q\cos\omega_m t + 120° & 1+2q\cos\omega_m t \\ 1+2q\cos\omega_m t + 120° & 1+2q\cos\omega_m t & 1+2q\cos\omega_m t - 120° \end{pmatrix} \quad (13.142)$$

where $\omega_m = -(\omega_o - \omega_i)$ such that $\varphi_i = -\varphi_o$

The solution $\varphi_i = \varphi_o$ gives the same phase displacement at the input and output ports whereas the solution $\varphi_i = -\varphi_o$ gives the reversed phase displacement. Combining both solutions provides for input displacement factor control. The maximum voltage ratio is when $q = ½$.

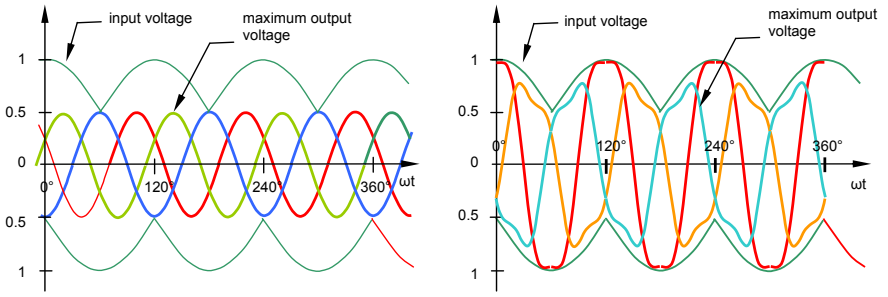


Figure 13.44. *Three-phase input voltages and three-phase output voltages showing: (a) maximum output magnitude of 0.5pu and (b) output voltage increased to 0.866 pu with triplen injection.*

As with standard sinusoidal PWM, by adding triplens (3rd harmonics from the input and output) q can be increased from ½ to ½√3 = 0.866, as illustrated in part b of figure 13.44.

$$v_o = qV_{imax} \begin{pmatrix} \cos\omega_o t & -⅙\cos 3\omega_o t + \frac{1}{2\sqrt{3}}\cos 3\omega_i t \\ \cos(\omega_o t + 120°) & -⅙\cos 3\omega_o t + \frac{1}{2\sqrt{3}}\cos 3\omega_i t \\ \cos(\omega_o t + 120°) & -⅙\cos 3\omega_o t + \frac{1}{2\sqrt{3}}\cos 3\omega_i t \end{pmatrix} \quad (13.143)$$

Since no intermediate energy storage stage is involved, such as a dc link, this so called total silicon solution to ac to ac conversion provides no ride-through, thus is not well suited to ups application. The advantage of the matrix converter over a dc link approach to ac to ac conversion lies not in the fact that a dc link capacitor is not required. Given the matrix converter requires an input *L-C* filter, capacitor size and cost requirements are similar. The key feature of the matrix converter is that the capacitor voltage requirement is ac. For a given temperature, ripple current, etc., the lifetime of an ac capacitor is significantly longer than a dc voltage electrolytic capacitor, as is required for a dc link. The use of oil impregnated paper bipolar capacitors to improve dc-link inverter reliability, significantly increases capacitor volume and cost for a given capacitance and voltage.

The key limitations of the matrix converter, hampering its exploitation are

- The ac output voltage is restricted to 86.6% of the ac input voltage (without distortion)
- The need for a capacitive over voltage 3Φ clamping circuit due commutation spikes
- Inter dependence between the input and output voltage and current harmonics
- The need for reverse blocking bidirectional current and voltage switches
- Minimal ride-through capability

13.6.1 High frequency resonant dc to ac matrix converter

A combination of integral cycle control with a high-frequency single-phase to three-phase matrix converter is shown in figure 13.45. High frequency ac is produced by a H-bridge parallel resonant voltage converter, which is transformer coupled to the matrix converter. A key feature is that both the H-bridge and matrix converter switches, can be soft-switched for low switching losses. Figure 13.46 shows the output voltage waveforms constituted from half-sine resonant voltage pulse components.

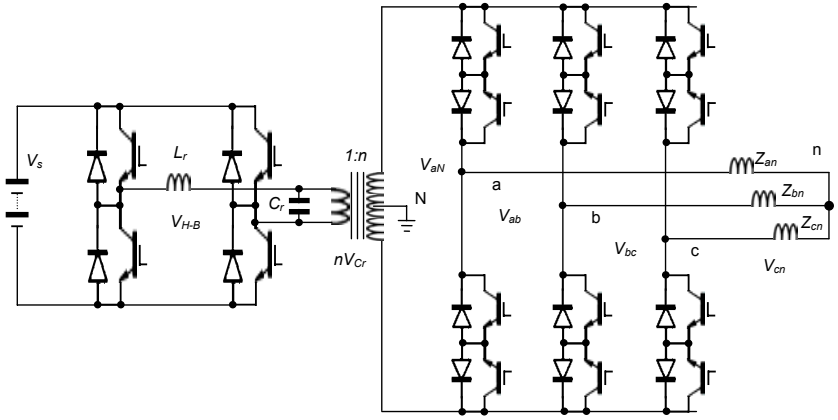


Figure 13.45. *Twelve switch high frequency ac to ac converter.*

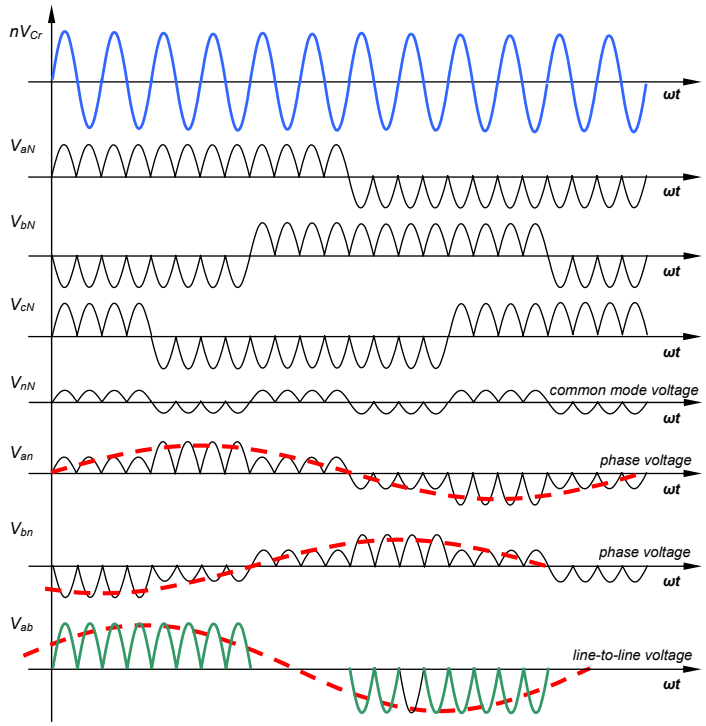


Figure 13.46. *Quasi-square generated voltages of the twelve switch high frequency ac to ac converter.*

13.7 Power Quality: load efficiency and supply current power factor

One characteristic of ac regulators is non-sinusoidal load current, hence supply current as illustrated in figure 13.1b. Difficulty therefore exists in defining the supply current power factor and the harmonics in the load current may detract from the load efficiency. For example, with a single-phase motor, current components other than the fundamental detract from the fundamental torque and increase motor heating, noise, and vibration. To illustrate the procedure for determining load efficiency and supply power factor, consider the circuit and waveforms in figure 13.1.

13.7.1 Load waveforms

The load voltage waveform is constituted from the sinusoidal supply voltage v and is defined by

$$v_o(\omega t) = \sqrt{2}V \sin \omega t \quad \text{(V)} \qquad \alpha \le \omega t \le \beta \qquad \pi+\alpha \le \omega t \le \pi+\beta \tag{13.144}$$

and $v_o = 0$ elsewhere.

Fourier analysis of v_o yields the load voltage Fourier coefficients v_{an} and v_{bn} such that

$$v_o(\omega t) = \sum_{\forall n} \{ v_{an} \cos n\omega t + v_{bn} \sin n\omega t \} \quad \text{(V)} \tag{13.145}$$

for all values of n.
The load current can be evaluated by solving

$$Ri_o + L\frac{di_o}{dt} = \sqrt{2}\, V \sin \omega t \quad \text{(V)} \tag{13.146}$$

over the appropriate bounds and initial conditions. From Fourier analysis of the load current i_o, the load current coefficients i_{an} and i_{bn} can be derived.

Derivation of the current waveform Fourier coefficients may prove complicated because of the difficulty of integrating an expression involving equation (13.2), the load current. An alternative and possibly simpler approach is to use superposition and the fact that each load Fourier voltage component produces a load current component at the associated frequency but displaced because of the load impedance at that frequency. That is

$$i_{an} = \frac{v_{an}}{R}\cos\phi_n \quad \text{(A)} \qquad i_{bn} = \frac{v_{bn}}{R}\cos\phi_n \quad \text{(A)} \tag{13.147}$$

where $\phi_n = \tan^{-1} n\omega L / R$

The load current i_o is given by

$$i_o(\omega t) = \sum_{\forall n} \left\{ i_{an}\cos\left(n\omega t - \phi_n\right) + i_{bn}\sin\left(n\omega t - \phi_n\right) \right\} \quad \text{(A)} \tag{13.148}$$

The load efficiency, η, which is related to the power dissipated in the resistive component R of the load, is defined by

$$\eta = \frac{\text{fundamental active power}}{\text{total active power}} = \frac{\tfrac{1}{2}\left(i_{a1}^2 R + i_{b1}^2 R\right)}{\tfrac{1}{2}\sum_{\forall n}\left(i_{an}^2 R + i_{bn}^2 R\right)} = \frac{i_{a1}^2 + i_{b1}^2}{\sum_{\forall n}\left(i_{an}^2 + i_{bn}^2\right)} \tag{13.149}$$

In general, the total load power is $\sum_{\forall n} v_{n\,rms} \times i_{n\,rms} \times \cos\phi_n$.

13.7.2 Supply waveforms

Linear load:
For sinusoidal single and three-phase ac supply voltages feeding a linear load, the load power and apparent power are given by

$$P = V_s I_s \cos\phi \qquad S = V_s I_s$$
$$P = \sqrt{3}\, V_{LL} I_s \cos\phi \qquad S = V_{LL} I_s \tag{13.150}$$

and the supply power factor is

$$\cos\phi = \frac{P}{S} \tag{13.151}$$

Non-linear loads (e.g. rectification)**:**

i. The supply distortion factor μ, displacement factor $\cos\psi$, and power factor λ give an indication of the adverse effects that a non-sinusoidal load current has on the supply as a result of SCR phase control.
In the circuit of figure 13.1a, the load and supply currents are the same and given by equation (13.2). The supply current Fourier coefficients i_{san} and i_{sbn} are the same as for the load current Fourier coefficients i_{sa} and i_{sb} respectively, as previously defined.
The total supply (input) power factor λ can be defined as

$$\lambda = \frac{\text{real power}}{\text{apparent power}} = \frac{\text{total mean input power}}{\text{total rms input VA}} = \frac{v_{1rms} i_{1rms} \cos\psi_1}{V_{rms} I_{rms}} = \frac{\frac{1}{\sqrt{2}}\sqrt{v_{sa1}^2 + v_{sb1}^2} \times \frac{1}{\sqrt{2}}\sqrt{i_{sa1}^2 + i_{sb1}^2} \times \cos\psi_1}{v \times \frac{1}{\sqrt{2}}\sqrt{i_{a1}^2 + i_{b1}^2}} \tag{13.152}$$

The supply voltage is sinusoidal hence supply power is not associated with the harmonic non-fundamental currents.

$$\lambda = \frac{v\sqrt{\tfrac{1}{2}\left(i_{sa1}^2 + i_{sb1}^2\right)}\,\cos\psi_1}{v\, I_{rms}} = \frac{\sqrt{\tfrac{1}{2}\left(i_{sa1}^2 + i_{sb1}^2\right)}\,\cos\psi_1}{I_{rms}} = \frac{i_{s1}}{I_{rms}} \times \cos\psi_1 = DF_{i1} \times DPF \tag{13.153}$$

where cos ψ, termed the displacement power factor, *DPF*, is the fundamental power factor defined as

$$\cos\psi_1 = \cos\left(-\tan^{-1} i_{sa1} / i_{sb1}\right) \tag{13.154}$$

Equating with equation (13.153), the total supply power factor is defined as

$$\lambda = \mu \cos\psi_1 \qquad 0 \ge \lambda \ge 1 \tag{13.155}$$

The supply current distortion factor μ is the ratio of fundamental rms current to total rms current i_{srms}, that is

$$\mu = \frac{\sqrt{\tfrac{1}{2}\left(i_{sa1}^2 + i_{sb1}^2\right)}}{i_{rms}} = \frac{i_{s1}}{I_{rms}} \tag{13.156}$$

ii. (a) The supply fundamental harmonic factor ρ_F is defined as

$$\rho_F = \frac{\text{total harmonic (non-fundamental) rms current (or voltage)}}{\text{fundamental rms current (or voltage)}} = I_h \Big/ i_{s1} = I_h \Big/ \sqrt{i_{sa1}^2 + i_{sb1}^2} = \sqrt{\frac{I_{rms}^2}{i_{sa1}^2 + i_{sb1}^2} - 1} = \sqrt{\frac{1}{\mu^2} - 1} \tag{13.157}$$

where I_h is the total harmonic (non-fundamental) current (assuming no dc component)

$$I_h = \sqrt{I_{rms}^2 - i_{s1}^2} = \sqrt{\sum_{n\neq 1}^{\infty} I_{nrms}^2} = \frac{1}{\sqrt{2}}\sqrt{\sum_{n\neq 1}^{\infty} i_{san}^2 + i_{sbn}^2} \tag{13.158}$$

The general relationships between the various current forms can be summarised as

$$I_{rms} = \sqrt{I_{dc}^2 + I_{1rms}^2 + I_{2rms}^2 + I_{3rms}^2 + \ldots} = \sqrt{I_{dc}^2 + I_{1rms}^2 + I_{hl}^2} \quad \text{or} \quad = \sqrt{I_{dc}^2 + I_{ac}^2} \tag{13.159}$$

(b) Alternatively, the supply total rms harmonic factor ρ_{RMS} is defined as:

$$\rho_{RMS} = \frac{\text{total harmonic (non-fundamental) rms current (or voltage)}}{\text{total rms current (or voltage)}} = I_h \Big/ i_{rms} = \sqrt{1 - \frac{i_{s1}^2}{I_{rms}^2}} = \sqrt{1-\mu^2} \tag{13.160}$$

iii. The supply crest factor δ is defined as the ratio of peak supply current $\hat{i}_s$ to the total rms current:

$$\delta = \hat{i}_s / I_{rms} \tag{13.161}$$

iv. The energy conversion factor υ is defined by

$$\upsilon = \frac{\text{fundamental output power}}{\text{fundamental input power}} = \frac{\tfrac{1}{\sqrt{2}}\sqrt{v_{a1}^2 + v_{b1}^2} \times \tfrac{1}{\sqrt{2}}\sqrt{i_{a1}^2 + i_{b1}^2} \times \cos\phi_1}{v \times \tfrac{1}{\sqrt{2}}\sqrt{i_{sa1}^2 + i_{sb1}^2} \times \cos\psi_1} \tag{13.162}$$

Example 13.7: ***Power quality - load efficiency***

If a purely resistive load R is fed with a voltage

$$v_o = \sqrt{2}V\sin\omega t + \frac{\sqrt{2}V}{3}\sin 3\omega t$$

what is the fundamental load efficiency?

Solution

The load current is given by

$$i_o = v_o / R = \frac{\sqrt{2}V}{R}\left(\sin\omega t + \frac{1}{3}\sin 3\omega t\right)$$

The load efficiency is given by equation (13.149), that is

$$\eta = \frac{\left(\frac{\sqrt{2}V}{R}\right)^2 R}{\left(\frac{\sqrt{2}V}{R}\right)^2 R + \left(\frac{\sqrt{2}V}{3R}\right)^2 R} = \frac{1}{1 + 1/9} = 0.90$$

The introduced third harmonic component decreases the load efficiency by 10%.

♣

Example 13.8: ***Power quality - sinusoidal source and constant current load***

A half-wave rectifier with a load freewheel diode as shown in figure 11.3 has a 10A constant current load, I_o. If rectifier circuit is supplied from the ac mains with voltage $v(\omega t) = \sqrt{2}\,230\times\sin 2\pi 50t$ determine:

i. the supply apparent power and average load power
ii. the total supply power factor, λ, hence distortion, μ, and displacement factors
iii. the average and rms current rating of each diode and diode reverse voltage requirements

Solution

The rms supply voltage is 230V, at 50Hz. The supply current is a 10ms, 10A current block occurring every 20ms. The rms supply current is therefore $10/\sqrt{2} = 7.07$A.

i. The supply apparent power is

$$S = V_{rms} I_{rms} = 230\text{V} \times 7.07\text{A} = 1626.1\text{VAr}$$

The average load voltage is that for half wave rectification, viz.,

$$V_o = \frac{\sqrt{2}V}{\pi} = 103.5\text{V}$$

The average load power, which must be equal to the input power from the 50Hz source, is

$$P_o = P_{in} = V_o I_o = 103.5\text{V} \times 10\text{A} = 1035\text{W}$$

The fundamental of a square wave, with a dc offset of half the magnitude is

$$I_{1rms} = \frac{1}{\sqrt{2}}\hat{I}_1 = \frac{1}{\sqrt{2}} \times \frac{2}{\pi} \times 10\text{A} = 4.50\text{A}$$

which is in phase with the ac supply, that is $\cos\phi_{50\text{Hz}} = 1$.

Alternately, the load power, hence input power, which is at the supply voltage frequency of 50Hz, can be confirmed by

$$P_{in} = V_{rms} I_{rms} \cos\phi_{50\text{Hz}} = 230\text{V} \times 7.07\text{A} \times 1 = 1035\text{W}$$

ii. The power factor is

$$pf = \lambda = \frac{P_{in}}{S} = \frac{1035\text{W}}{1626.1\text{VAr}} = 0.64$$

The current distortion factor is

$$DF = \mu = \frac{I_{1rms}}{I_{rms}} = \frac{4.50A}{7.07A} = 0.64$$

which, since the supply is single frequency sinusoidal, confirms that the displacement factor for the fundamental current is

$$\cos\psi_1 = \frac{\lambda}{\mu} = \frac{0.64}{0.64} = 1 = \cos\phi_{50\text{Hz}}$$

that is $\phi_{50\text{Hz}} = 0°$

iii. The average and rms current ratings of both the rectifying diode and the freewheel diode are the same, viz.,

$$\bar{I}_D = \frac{I_o}{2} = \frac{10\text{A}}{2} = 5\text{A} \qquad I_{D\,rms} = \frac{I_o}{\sqrt{2}} = \frac{10\text{A}}{\sqrt{2}} = 7.07\text{A}$$

In reverse bias, each diode experiences alternate ac supply peak voltages of $\sqrt{2}\,230\text{V} = 325.3\text{V}$

♣

Example 13.9: ***Power quality - sinusoidal source and non-linear load***

An unbalanced single-phase rectifier circuit is supplied from the ac mains with voltage $v(\omega t) = \sqrt{2}\,230\times\sin 2\pi 50t$. The dominant resultant harmonics in the supply current are

$$i(\omega t) = 10 + 15\times\sin\left(\omega t + \tfrac{1}{6}\pi\right) + 3\times\sin\left(2\omega t + \tfrac{1}{4}\pi\right) + 2\times\sin\left(4\omega t - \tfrac{1}{4}\pi\right)$$

Determine

i. the fundamental power factor hence power delivered from the supply
ii. the total supply power factor, hence distortion factor
iii. the harmonic current and the ac current

iv. the total harmonic distortion with respect to the fundamental current and the total rms current
v. the current crest factor.

Solution

i. The power from the supply delivered to the load is only at the supply frequency

$$P_{50Hz} = V_{s50Hz} I_{s50Hz} \cos\phi_{50Hz}$$
$$= 230\text{V} \times \frac{15\text{A}}{\sqrt{2}} \times \cos \tfrac{1}{6}\pi = 2113\text{W}$$

The fundamental power factor $\cos \tfrac{1}{6}\pi$ = 0.866, leading.

ii. The total supply power factor is

$$pf = \lambda = \frac{P_{50Hz}}{S} = \frac{V_{s50Hz} I_{s50Hz} \times \cos\phi_{50Hz}}{V_{s50Hz} \times I_s} = \frac{I_{s50Hz}}{I_s} \times \cos\phi_{50Hz} = \mu \cos\psi_1$$

The supply rms current I_s is

$$I_s = \sqrt{10\text{A}^2 + \left(\frac{15\text{A}}{\sqrt{2}}\right)^2 + \left(\frac{3\text{A}}{\sqrt{2}}\right)^2 + \left(\frac{2\text{A}}{\sqrt{2}}\right)^2} = 14.8\text{A}$$

Hence

$$\lambda = \frac{P_{50Hz}}{S} = \frac{2113\text{W}}{230\text{V} \times 14.8\text{A}} = 0.62$$
$$= \frac{15\text{A}/\sqrt{2}}{14.8\text{A}} \times 0.866 = 0.717 \times 0.866 = \mu \cos\psi_1$$

The total supply power factor λ is 0.62 and the current distortion factor μ is 0.717.

iii. From equation (13.158) the supply harmonic (non 50Hz) current is

$$I_h = \sqrt{I_{rms}^2 - i_{s1}^2}$$
$$= \sqrt{14.8^2 - \left(\frac{15}{\sqrt{2}}\right)^2} = 10.3\text{A}$$

and from equation (13.159) the ac supply current (non-dc) is

$$I_{ac} = \sqrt{I_{rms}^2 - I_{dc}^2}$$
$$= \sqrt{14.8\text{A}^2 - 10\text{A}^2} = 10.9\text{A}$$

iv. From equations (13.157) and (13.160), total harmonic distortions on the supply current are

$$\rho_F = \frac{\text{total harmonic (non-50Hz) rms current}}{\text{fundamental rms current}}$$
$$= I_h / i_{s1} = \sqrt{\frac{1}{\mu^2} - 1} = \sqrt{\frac{1}{0.717^2} - 1} = 0.97$$

and

$$\rho_{RMS} = \frac{\text{total harmonic (50Hz) rms current}}{\text{total rms current}}$$
$$= I_h / i_{rms} = \sqrt{1 - \mu^2} = \sqrt{1 - 0.717^2} = 0.70$$

v. The current crest factor is given by equation (13.161), namely $\delta = \hat{i}_s / I_{rms}$. The maximum supply current will be dominated by the dc and 50Hz components thus the maximum will be near $\omega t = \frac{1}{6}\pi = \frac{1}{2}\pi$, $\omega t = \frac{1}{3}\pi$. Iteration around $\omega t = \frac{1}{3}\pi$ gives $\hat{i}_s$ =28.85A at ωt =0.83 rad.

$$\delta = \frac{\hat{i}_s}{I_{rms}} = \frac{28.85\text{A}}{14.8\text{A}} = 1.95$$

♣

Reading list

Hart, D.W., *Introduction to Power Electronics,*
Prentice-Hall, Inc, 1994.

Rombaut, C., et al., *Power Electronic Converters – AC/AC Conversion*,
North Oxford Academic Publishers, 1987.

Problems

13.1. Determine the rms load current for the ac regulator in figure 13.24, with a resistive load *R*. Consider the delay angle intervals 0 to ½π, ½π to ⅔π, and ⅔π to 7π /6.

13.2. The ac regulator in figure 13.24, with a resistive load *R* has one thyristor replaced by a diode. Show that the rms output voltage is

$$V_{rms} = \left[\frac{1}{2\pi}(2\pi - \alpha + \tfrac{1}{2}\sin 2\alpha)\right]^{\frac{1}{2}}$$

while the average output voltage is

$$\overline{V}_o = \frac{\sqrt{2}V}{2\pi}(\cos\alpha - 1)$$

13.3. Plot the load power for a resistive load for the fully controlled and half-controlled three-phase ac regulator, for varying phase delay angle, α. Normalise power with respect to $\hat{V}^2/R$.

13.4. For the tap changer in figure 13.15, with a resistive load, calculate the rms output voltage for a phase delay angle α. If v_2 = 200V ac and v_1 = 240V ac, calculate the power delivered to a 10 ohm resistive load at delay angles of ¼π, ½π, and ¾π. What is the maximum power that can be delivered to the load?

13.5. A. 0.01H inductance is added in series with the load in problem 13.4. Determine the load voltage and current waveforms at a firing delay angle of ½π. Assuming a 50 Hz supply, what is the minimum delay angle?

13.6. The thyristor T_2 in the single-phase controller in figure 13.1a is replaced by a diode. The supply is 240V ac, 50 Hz and the load is 10Ω resistive. For a delay angle of α = 90°, determine the
i. rms output voltage
ii. supply power factor
iii. mean output voltage
iv. mean input current.
[207.84 V; 0.866 lagging; 54 V; 5.4 A]

13.7. The single-phase ac controller in figure 13.6 operating on the 240 V, 50 Hz mains is used to control a 10Ω resistive heating load. If the load is supplied repeatedly for 75 cycles and disconnected for 25 cycles, determine the
i. rms load voltage,
ii. input power factor, λ, and
iii. the rms thyristor current.

13.8 The ac controller in problem 13.3 delivers 2.88kW. Determine the duty cycle, *m*/*N*, and the input power factor, λ.

13.9 A single-phase ac controller with a 240Vac 50Hz voltage source supplies an *R*-*L* load of *R*=40Ω and *L*=50mH. If the thyristor gate delay angle is α = 30°, determine:
i. an expression for the load current
ii. the rms load current
iii. the rms and average current in the thyristors
iv. the power absorbed by the load
v. sketch the load, supply and thyristor voltages and currents.

13.10. A single-phase thyristor ac controller is to delivery 500W to an *R*-*L* load of *R*=25Ω and L=50mH. If the ac supply voltage is 240V ac at 50Hz, determine
i. thyristor rms and average current
ii. maximum voltages across the thyristors.

13.11. The thyristor T_2 in the single-phase controller in figure 13.1a is replaced by a diode. The supply is 240V ac, 50 Hz and the load is 10Ω resistive. Determine the
i. an expression for the rms load voltage in terms of α
ii. the range of rms voltage across the load resistor.

14

DC Choppers

A *dc chopper* is a dc-to-dc voltage converter. It is a static switching electrical appliance that in one electrical conversion, changes an input fixed dc voltage to an adjustable dc output voltage without inductive or capacitive intermediate energy storage. The name *chopper* is connected with the fact that the output voltage is a 'chopped up' quasi-rectangular version of the input dc voltage.
In chapters 12 and 13, thyristor devices were used in conjunction with an ac supply that forces thyristor turn-off at ac supply current reversal. This form of thyristor natural commutation, which is illustrated in figure 14.1a, is termed line or source commutation.
When a dc source is used with a thyristor circuit, energy source facilitated commutation is clearly not possible. If the load is an *R-C* or *L-C* circuit as illustrated in figure 14.1b, the load current falls to zero whence the thyristor in series with the dc supply turns off. Such a natural turn-off process is termed load commutation.
If the supply is dc and the load current has no natural zero current periods, such as with the *R-L* load, dc chopper circuit shown in figure 14.1c, the load current can only be commutated using a self-commutating switch, such as a GTO thyristor, GCT, IGBT or MOSFET. An SCR is not suitable since once the device is latched on in this dc supply application, it remains on.
The dc chopper in figure 14.1c is the simplest of the five dc choppers to be considered in this chapter. This single-ended, grounded-load, dc chopper will be extensively analysed. See example 14.3.

14.1 DC chopper variations

There are five types of dc choppers, of which four are a subset of the fifth - the flexible but basic, four-quadrant H-bridge chopper shown in the centre of figure 14.2. Notice that the circuits in figure 14.2 are highlighted so that the derivation of each dc chopper from the fundamental H-bridge four-quadrant, dc chopper can be seen. Each chopper can be categorized depending on which output I_o-V_o quadrant or quadrants it can operate in, as shown in figure 14.2. The five different choppers in figure 14.2 are classified according to their output I_o - V_o capabilities as follows:

(a)	First quadrant -	I	$+V_o$ $+I_o$	using switch/diode T_1 D_1
(b)	Second quadrant -	II	$+V_o$ $-I_o$	using switch/diode T_2 D_2
(c)	Two quadrant -	I and II	$+V_o$ $\pm I_o$	using switches/diodes T_1 D_1 T_2 D_2
(d)	Two quadrant -	I and IV	$\pm V_o$ $+I_o$	using switches/diodes T_1 D_1 T_4 D_4
(e)	Four quadrant -	I, II, III, and IV	$\pm V_o$ $\pm I_o$	using switches/diodes $T_{1\text{-}4}$ $D_{1\text{-}4}$

In the five choppers in figure 31.2, the subscript of the active switch or switches and diodes specify in which quadrants operation is possible. For example, the chopper in figure 14.2d uses switches T_1 and T_3, (plus diodes D_1 and D_3) so can only operate in the first ($+I_o$,$+V_o$) and third ($-I_o$,$-V_o$) quadrants.
The **first-quadrant chopper** in figure 14.2a (and figure 14.1c) can only produce a positive voltage across the load since the freewheel diode D_1 prevents a negative output voltage. Also, the chopper can only deliver current from the dc source to the load through the unidirectional switch T_1. It is therefore a single quadrant chopper and only operates in the first quadrant ($+I_o$,$+V_o$), where $V_o \le V_s$.
The **second-quadrant chopper,** ($-I_o$,$+V_o$), in figure 14.2b is a voltage boost circuit and current flows from the load to the supply, V_s. The switch T_2 is turned on to build-up the inductive load current. Then when the switch is turned off current is forced to flow through diode D_2 into the dc supply. The two current paths (when the switch is on and when it is off) are shown in figure 14.2b.

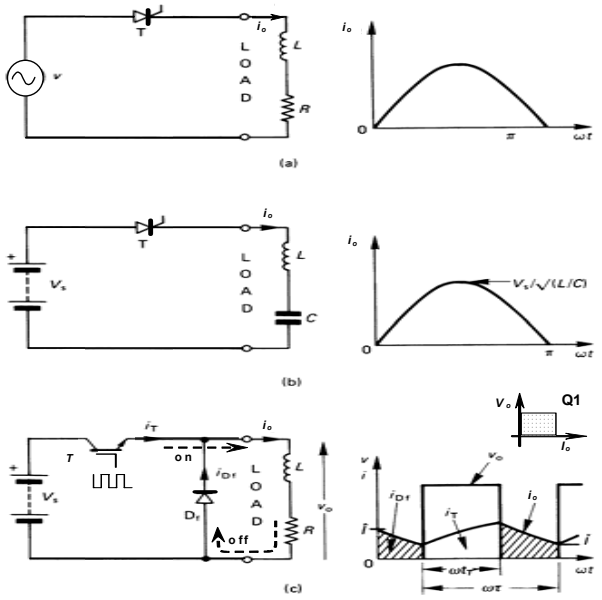


Figure 14.1. *Three basic types of switch commutation techniques: (a) source commutation; (b) load commutation; and (c) switch commutation.*

In the two-quadrant chopper, **quadrants I and II chopper,** ($\pm I_o$,$+V_o$), figure 14.2c, the load voltage is clamped to between 0V and V_s, because of the freewheel diodes D_1 and D_2. Because this chopper is a combination of the first-quadrant chopper in figure 14.2a and the second-quadrant chopper in figure 14.2b, it combines the characteristics of both. Bidirectional load current is possible but the average output voltage is always positive. Energy can be regenerated into the dc supply V_s due to the load inductive stored energy which maintains current flow from the back emf source ($<V_s$) in the load.
The two-quadrant chopper, **quadrants I and IV chopper,** ($+I_o$,$\pm V_o$), figure 14.2d, can produce a positive voltage, negative voltage or zero volts across the load, depending on the duty cycle of the switches and the switching sequence. When both switches are switched simultaneously, an on-state duty cycle of less than 50% ($\delta < ½$) results in a negative average load voltage V_o, while $\delta > ½$ produces a positive average load voltage. Since V_o is reversible, the power flow direction is reversible, for the shown current i_o. Zero voltage loops are created when one of the two switches is turned off.
The **four-quadrant chopper** in the centre of figure 14.2 combines all the properties of the four subclass choppers. It uses four switched and is capable of producing positive or negative voltages across the load, whilst delivering current to the load in either direction, ($\pm I_o$,$\pm V_o$).
The step-up chopper, or boost converter, presented in Chapter 17.4, may be considered a dc chopper variation, which has second quadrant characteristics, while the step-down converter presented in Chapter 17.1 can be considered a first quadrant dc chopper.

14.2 First-Quadrant dc chopper

The basic first-quadrant dc chopper circuit reproduced in figure 14.3a can be used to control a dc load such as a dc motor. As such, the dc load has a back-emf component, $E = k\phi\omega$, the magnitude and polarity of which are dependant on the flux ϕ, (field current i_f) and its direction, and the speed ω and its direction. If the *R-L* load (with time constant $\tau = L/R$) incorporates an opposing back emf, *E*, then when the switch T_1 is off and the diode D_1 is conducting, the load current can be forced towards zero by the opposing back emf. Therefore two output load current modes (continuous and discontinuous load current) can occur depending on the relative magnitude of the back emf, load time constant, and the switch on-state duty cycle. Continuous load current waveforms are shown in figure 14.3b, while waveforms for discontinuous load current, with periods of zero current, are shown in figure 14.3c.

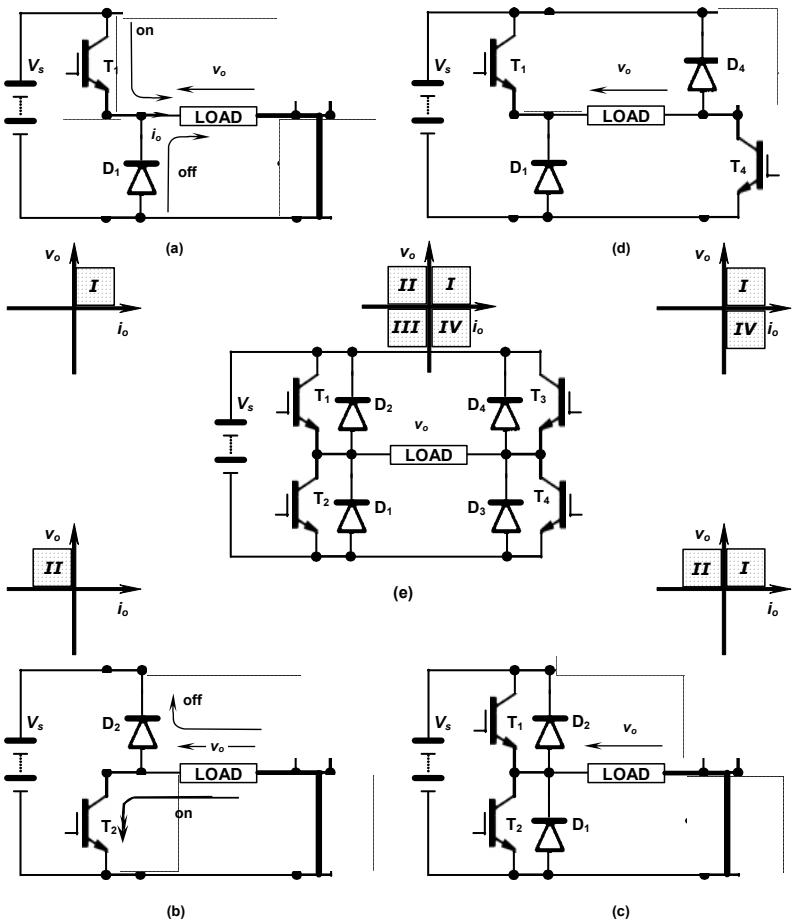


Figure 14.2. *Fundamental four-quadrant chopper (centre) showing derivation of four subclass dc choppers: (a) first-quadrant chopper - I; (b) second-quadrant chopper - II; (c) first and second quadrants chopper – I and II; (d) first and fourth quadrants chopper – I and IV; and (e) four-quadrant chopper.*

In both conduction cases, the average voltage across the load can be controlled by varying the on-to-off time duty cycle of the switch, T_1. The on-state duty cycle, δ, is normally controlled by using pulse-width modulation, frequency modulation, or a combination of both. When the switch is turned off the inductive load current continues and flows through the load freewheel diode, D_1, shown in figure 14.3a

The analysis to follow for all the dc choppers, assumes

- No source impedance
- Constant switch duty cycle
- Steady state conditions have been reached
- Ideal semiconductors and
- No load impedance temperature effects.

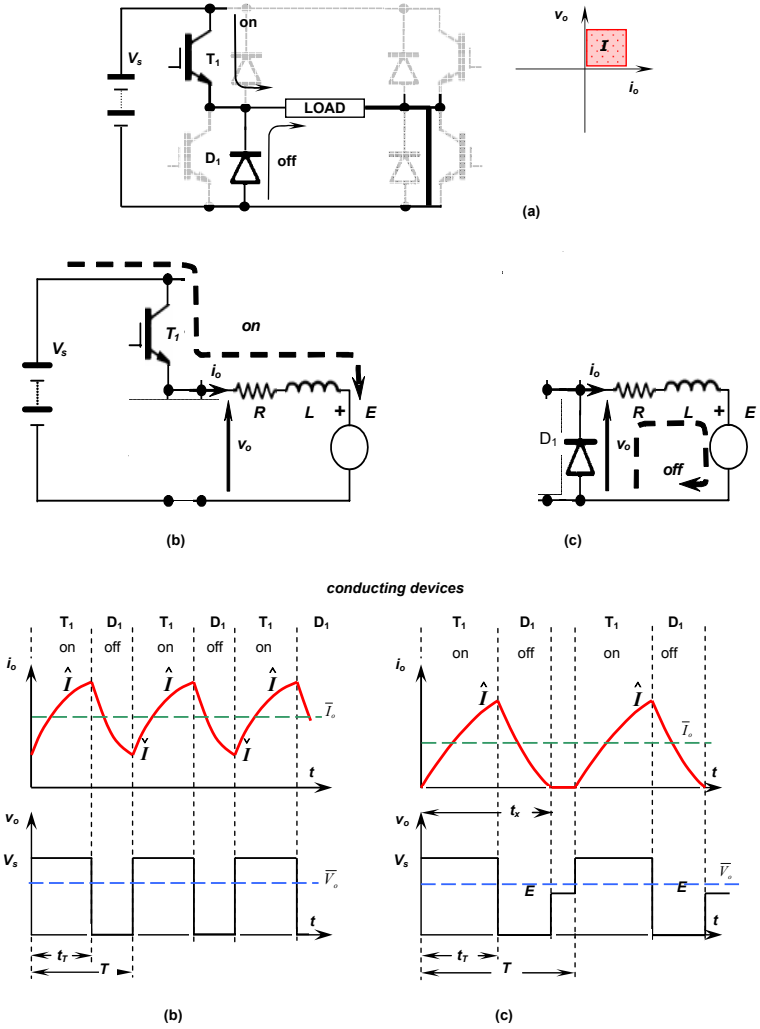


Figure 14.3. *First-quadrant dc chopper and two basic modes of chopper output current operation: (a) basic circuit and current paths; (b) continuous load current; and (c) discontinuous load current after t = t_x.*

14.2.1 Continuous load current

Load waveforms for continuous load current conduction are shown in figure 14.3b.
The output voltage v_o, or load voltage is defined by

$$v_o(t) = \begin{cases} V_s & \text{for } 0 \le t \le t_T \\ 0 & \text{for } t_T \le t \le T \end{cases} \qquad (14.1)$$

The mean load voltage (hence mean load current) is

$$\overline{V}_o = \frac{1}{T}\int_0^{t_T} v_o(t)\,dt = \frac{1}{T}\int_0^{t_T} V_s\,dt$$
$$= \frac{t_T}{T}V_s = \delta V_s \quad \text{whence} \quad \overline{I}_o = V_o - E \Big/ R \tag{14.2}$$

where the switch on-state duty cycle $\delta = t_T/T$ is defined in figure 14.3b.
The rms load voltage is

$$V_{rms} = \left[\frac{1}{T}\int_0^{t_T} v_o^2(t)\,dt\right]^{1/2} = \left[\frac{1}{T}\int_0^{t_T} V_s^2\,dt\right]^{1/2}$$
$$= \sqrt{\frac{t_T}{T}}\,V_s = \sqrt{\delta}\,V_s \tag{14.3}$$

The output ac ripple voltage is

$$V_r = \sqrt{V_{rms}^2 - V_o^2}$$
$$= \sqrt{\left(\sqrt{\delta}\,V_s\right)^2 - \left(\delta V_s\right)^2} = V_s\sqrt{\delta(1-\delta)} \tag{14.4}$$

The maximum rms ripple voltage in the output occurs when $\delta = ½$ giving an rms ripple voltage of $½V_s$.
The output voltage ripple factor is

$$RF = \frac{V_r}{\overline{V}_o} = \sqrt{\left(\frac{V_{rms}}{\overline{V}_o}\right)^2 - 1} = \sqrt{FF^2 - 1}$$
$$= \sqrt{\left(\frac{\sqrt{\delta}\,V_s}{\delta V_s}\right)^2 - 1} = \sqrt{\frac{1}{\delta} - 1} = \sqrt{\frac{1-\delta}{\delta}} \tag{14.5}$$

Thus as the duty cycle $\delta \to 1$, the ripple factor tends to zero, consistent with the output being dc, that is $V_r = 0$.

Steady-state time domain analysis of first-quadrant chopper
- with load back emf and continuous output current

The time domain load current can be derived in a number of ways.

- First, from the Fourier coefficients of the output voltage, the current can be found by dividing by the load impedance at each harmonic frequency.
- Alternatively, the various circuit currents can be found from the time domain load current equations.

i. Fourier coefficients: The Fourier coefficients of the load voltage are independent of the circuit and load parameters and are given by

$$a_n = \frac{V_s}{n\pi}\sin 2\pi n\delta$$
$$b_n = \frac{V_s}{n\pi}(1 - \cos 2\pi n\delta) \qquad \text{for } n \geq 1 \tag{14.6}$$

Thus the peak magnitude and phase of the n^{th} harmonic are given by

$$c_n = \sqrt{a_n^2 + b_n^2}$$
$$\phi_n = \tan^{-1} a_n \Big/ b_n$$

Substituting expressions from equation (14.6) yields

$$c_n = \frac{2V_s}{n\pi}\sin \pi n\delta$$
$$\phi_n = \tan^{-1}\frac{\sin 2\pi n\delta}{1 - \cos 2\pi n\delta} = \tan^{-1}\cot \pi n\delta = ½\pi - \pi n\delta \tag{14.7}$$

where

$$v_n = c_n \sin\left(n\,\omega t + \phi_n\right) \tag{14.8}$$

such that

$$v_o(t) = V_o + \sum_{n=1}^{\infty} c_n \sin\left(n\,\omega t + \phi_n\right)$$
$$= \delta V_s + \sum_{n=1}^{\infty}\frac{2V_s}{n\pi}\sin \pi n\delta \cos\left(n\left(\omega t - \pi\delta\right)\right) \tag{14.9}$$

The load current is given by

$$i_o(t) = \sum_{n=0}^{\infty} i_n = \frac{\overline{V}_o}{R} + \sum_{n=1}^{\infty}\frac{v_n}{Z_n} = \frac{\overline{V}_o}{R} + \sum_{n=1}^{\infty}\frac{c_n \sin\left(n\,\omega t - \phi_n\right)}{Z_n} \tag{14.10}$$

where the load impedance at each harmonic frequency is given by

$$Z_n = \sqrt{R^2 + \left(n\,\omega L\right)^2}$$

ii. Time domain differential equations: By solving the appropriate time domain differential equations, the continuous load current shown in figure 14.3b is defined by

During the **switch on-period**, when $v_o(t) = V_s$

$$L\frac{di_o}{dt} + R\,i_o + E = V_s$$

which yields

$$i_o(t) = \frac{V_s - E}{R}\left(1 - e^{\frac{-t}{\tau}}\right) + \hat{I}\,e^{\frac{-t}{\tau}} \qquad \text{for} \quad 0 \leq t \leq t_T \tag{14.11}$$

During the **switch off-period**, when $v_o(t) = 0$

$$L\frac{di_o}{dt} + R\,i_o + E = 0$$

which, after shifting the zero time reference to t_T, in figure 14.3a, gives

$$i_o(t) = -\frac{E}{R}\left(1 - e^{\frac{-t}{\tau}}\right) + \check{I}\,e^{\frac{-t}{\tau}} \qquad \text{for} \quad 0 \leq t \leq T - t_T \tag{14.12}$$

$$\text{where} \quad \hat{I} = \frac{V_s}{R}\frac{1 - e^{\frac{-t_T}{\tau}}}{1 - e^{\frac{-T}{\tau}}} - \frac{E}{R} \qquad \text{(A)}$$
$$\text{and} \quad \check{I} = \frac{V_s}{R}\frac{e^{\frac{t_T}{\tau}} - 1}{e^{\frac{T}{\tau}} - 1} - \frac{E}{R} \qquad \text{(A)} \tag{14.13}$$

The output ripple current, for continuous conduction, is independent of the back emf E and is given by

$$I_{p-p} = \Delta i_o = \hat{I} - \check{I} = \frac{V_s}{R}\frac{(1 - e^{\frac{-t_T}{\tau}})(1 - e^{\frac{-T+t_T}{\tau}})}{1 - e^{\frac{-T}{\tau}}} \tag{14.14}$$

which in terms of the on-state duty cycle, $\delta = t_T/T$, becomes

$$I_{p-p} = \frac{V_s}{R}\frac{(1 - e^{\frac{-\delta T}{\tau}})(1 - e^{\frac{-(1-\delta)T}{\tau}})}{1 - e^{\frac{-T}{\tau}}} \tag{14.15}$$

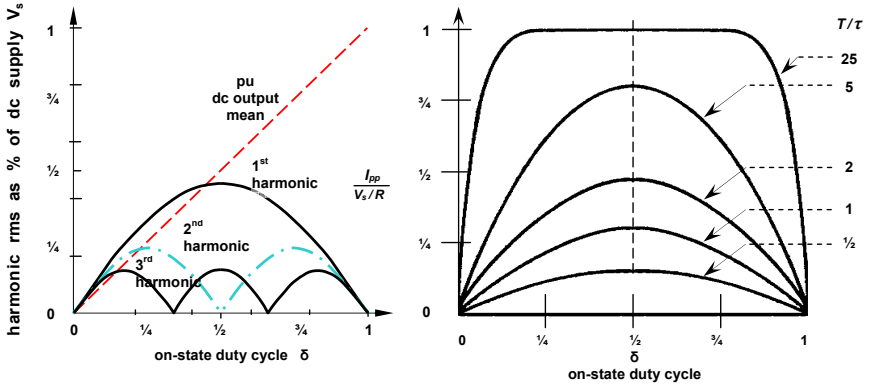


Figure 14.4. *Harmonics in the output voltage and ripple current as a function of duty cycle $\delta = t_T/T$ and ratio of cycle period T (switching frequency, $f_s = 1/T$) to load time constant $\tau = L/R$. Valid only for continuous load current conduction.*

The peak-to-peak ripple current can be extracted from figure 14.4, which shows a family of curves for equation (14.15), normalised with respect to V_s/R. For a given load time constant $\tau = L/R$, switching frequency $f_s = 1/T$, and switch on-state duty cycle δ, the ripple current can be extracted. This figure shows a number of important features of the ripple current.

- The ripple current I_{pp} reduces to zero as $\delta \to 0$ and $\delta \to 1$.
- Differentiation of equation (14.15) reveals that the maximum ripple current $\hat{I}_{p-p}$ occurs at $\delta = ½$.
- The longer the load L/R time constant, τ, the lower the output ripple current $I_{p\text{-}p}$.
- The higher the switching frequency, $1/T$, the lower the output ripple.

If the switch conducts continuously ($\delta = 1$), then substitution of $t_T = T$ into equations (14.11) to (14.13) gives a load voltage V_s and a dc load current is

$$i_o = \hat{I} = \check{I} = \frac{V_s - E}{R} \quad \left(= \frac{V_o - E}{R} = \overline{I}_o \right) \qquad \text{(A)} \tag{14.16}$$

The mean output current with continuous load current is found by integrating the load current over two consecutive periods, the switch conduction given by equation (14.11) and diode conduction given by equation (14.12), which yields

$$\begin{aligned} \overline{I}_o &= \frac{1}{T}\int_0^T i_o(t)\,dt = \left(\overline{V}_o - E\right) \Big/ R \\ &= \left(\delta V_s - E\right) \Big/ R \end{aligned} \qquad \text{(A)} \tag{14.17}$$

The input and output powers are related such that

$$\begin{aligned} P_{in} &= P_{out} \\ P_{in} &= V_s \overline{I}_i = V_S \left(\frac{\delta(V_s - E)}{R} - \frac{\tau}{T}\left(\hat{I} - \check{I}\right) \right) \\ P_{out} &= \frac{1}{T}\int_0^T v_o(t) i_o(t)\,dt \\ &= I_{o\,rms}^2 R + E\overline{I}_o = I_{o\,rms}^2 R + E\left(\frac{\delta V_s - E}{R}\right) \end{aligned} \tag{14.18}$$

from which the average input current can be evaluated.

Alternatively, the average input current, which is the average switch current, $\overline{I}_{switch}$, can be derived by integrating the switch current which is given by equation (14.11), that is

$$\begin{aligned} \overline{I}_i &= \overline{I}_{switch} = \frac{1}{T}\int_0^{t_T} i_o(t)\,dt \\ &= \frac{1}{T}\int_0^{t_T} \left(\frac{V_s - E}{R}\left(1 - e^{\frac{-t}{\tau}}\right) + \check{I} e^{\frac{-t}{\tau}} \right) dt \\ &= \frac{\delta(V_s - E)}{R} - \frac{\tau}{T}\left(\hat{I} - \check{I}\right) \end{aligned} \tag{14.19}$$

The term $\hat{I} - \check{I} = I_{p-p}$ is the peak-to-peak ripple current, which is given by equation (14.15). By Kirchhoff's current law, the average diode current $\overline{I}_{diode}$ is the difference between the average output current $\overline{I}_o$ and the average input current, $\overline{I}_i$, that is

$$\begin{aligned} \overline{I}_{diode} &= \overline{I}_o - \overline{I}_i \\ &= \left(\delta V_s - E\right) \Big/ R - \frac{\delta(V_s - E)}{R} + \frac{\tau}{T}\left(\hat{I} - \check{I}\right) \\ &= \frac{\tau}{T}\left(\hat{I} - \check{I}\right) - \frac{E(1-\delta)}{R} \end{aligned} \tag{14.20}$$

Alternatively, the average diode current can be found by integrating the diode current given in equation (14.12), as follows

$$\begin{aligned} \overline{I}_{diode} &= \frac{1}{T}\int_0^{T - t_T} \left(-\frac{E}{R}\left(1 - e^{\frac{-t}{\tau}}\right) + \hat{I} e^{\frac{-t}{\tau}} \right) dt \\ &= \frac{\tau}{T}\left(\hat{I} - \check{I}\right) - \frac{E(1-\delta)}{R} \end{aligned} \tag{14.21}$$

If E represents motor back emf, then the electromagnetic energy conversion efficiency is given by

$$\eta = \frac{E\overline{I}_o}{P_{in}} = \frac{E\overline{I}_o}{V_s \overline{I}_i} \tag{14.22}$$

The chopper effective (dc) input impedance at the dc source is given by

$$Z_{in} = \frac{V_s}{\overline{I}_i} \tag{14.23}$$

For an *R-L* load without a back emf, set $E = 0$ in the foregoing equations. The discontinuous load current analysis to follow is not valid for an *R-L*, with E=0 load, since the load current never reaches zero, but at best asymptotes towards zero during the off-period of the switch.

14.2.2 Discontinuous load current

With an opposing emf E in the load, the load current can reach zero during the off-time, at a time t_x shown in figure 14.3c. The time t_x can be found by

- deriving an expression for $\hat{\imath}$ from equation (14.11), setting $t = t_T$,
- this equation is substituted into equation (14.12) which is equated to zero, having substituted for $t = t_x$: yielding

$$t_x = t_T + \tau \ln\left(1 + \frac{V_s - E}{E}\left(1 - e^{\frac{-t_T}{\tau}}\right)\right) \qquad \text{(s)} \tag{14.24}$$

This equation shows that $t_x > t_T$.

Alternatively, for a given frequency, $1/T$, discontinuous current will occur as the duty cycle is decreased. Rearranging equation (14.13), after setting $\check{I} = 0$ and extracting the duty cycle $\delta = t_T / T$ yields

$$\delta = \frac{t_T}{T} \le \frac{\tau}{T}\,\ell n\left[1 + \frac{E}{V_s}\left(e^{\frac{T}{\tau}} - 1\right)\right] \tag{14.25}$$

Conversely, discontinuous conduction will occur as the frequency decreases for a given switch on-time, according to

$$T = \frac{1}{f} \ge \tau\,\ell n\left[1 + \frac{V_s}{E}\left(e^{\frac{t_T}{\tau}} - 1\right)\right] \tag{14.26}$$

Figure 14.5 can be used to determine if a particular set of operating conditions involves discontinuous load current.

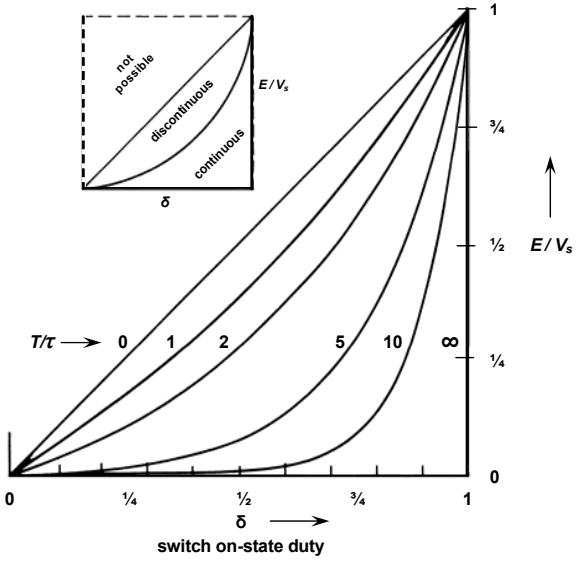


Figure 14.5. *Bounds of discontinuous load current with E>0.*

The load voltage waveform for discontinuous load current conduction shown in figure 14.3c is defined by

$$v_o(t) = \begin{Bmatrix} V_s & \text{for } 0 \le t \le t_T \\ 0 & \text{for } t_T \le t \le t_x \\ E & \text{for } t_x \le t \le T \end{Bmatrix} \tag{14.27}$$

If discontinuous load current exists for a period $T - t_x$, from t_x until T, then the mean output voltage is

$$\overline{V}_o = \frac{1}{T}\left(\int_0^{t_T} V_s\,dt + \int_{t_T}^{t_x} 0\,dt + \int_{t_x}^{T} E\,dt\right) \qquad \left(\text{thence } \overline{I}_o = \overline{V}_o - E \Big/ R\right)$$

$$\overline{V}_o = \delta V_s + \frac{T - t_x}{T} E \qquad \text{(V)} \quad \text{for} \quad t_x \geq t_T \tag{14.28}$$

The rms output voltage with discontinuous load current conduction is given by

$$V_{rms} = \left[\frac{1}{T}\left(\int_0^{t_T} V_s^2\,dt + \int_{t_T}^{t_x} 0^2 d + \int_{t_x}^{T} E^2 dt\right)\right]^{1/2}$$

$$= \sqrt{\delta V_s^2 + \frac{T - t_x}{T} E^2} \qquad \text{(V)} \tag{14.29}$$

The ac ripple voltage and ripple factor can be found by substituting equations (14.28) and (14.29) into

$$V_r = \sqrt{V_{rms}^2 - V_o^2} \tag{14.30}$$

and

$$RF = \frac{V_r}{\overline{V}_o} = \sqrt{\left(\frac{V_{rms}}{\overline{V}_o}\right)^2 - 1} = \sqrt{FF^2 - 1} \tag{14.31}$$

Steady-state time domain analysis of first-quadrant chopper
- with load back emf and discontinuous output current

i. Fourier coefficients: The load current can be derived indirectly by using the output voltage Fourier series. The Fourier coefficients of the load voltage are

$$a_n = \frac{V_s}{n\pi}\sin 2\pi n\delta - \frac{E}{n\pi}\sin 2\pi n \Big/ \frac{t_x}{T}$$

$$b_n = \frac{V_s}{n\pi}(1 - \cos 2\pi n\delta) - \frac{E}{n\pi}\left(1 - \cos 2\pi n \Big/ \frac{t_x}{T}\right) \qquad n \geq 1 \tag{14.32}$$

which using

$$c_n = \sqrt{a_n^2 + b_n^2}$$

$$\phi_n = \tan^{-1} a_n \Big/ b_n$$

give

$$v_o(t) = \overline{V}_o + \sum_{n=1}^{\infty} c_n \sin(n\omega t + \phi_n) \tag{14.33}$$

The appropriate division by $Z_n = \sqrt{R^2 + (n\omega L)^2}$ yields the output current.

ii. Time domain differential equations: For discontinuous load current, $\breve{I} = 0$. Substituting this condition into the time domain equations (14.11) to (14.14) yields equations for discontinuous load current, specifically:

During the **switch on-period**, when $v_o(t) = V_s$,

$$i_o(t) = \frac{V_s - E}{R}\left(1 - e^{\frac{-t}{\tau}}\right) \qquad \text{for} \quad 0 \leq t \leq t_T \tag{14.34}$$

During the **switch off-period**, when $v_o(t) = 0$, after shifting the zero time reference to t_T,

$$i_o(t) = -\frac{E}{R}\left(1 - e^{\frac{-t}{\tau}}\right) + \hat{I}\,e^{\frac{-t}{\tau}} \qquad \text{for} \quad 0 \leq t \leq t_x - t_T \tag{14.35}$$

where from equation (14.34), with $t = t_T$,

$$\hat{I} = \frac{V_s - E}{R}\left(1 - e^{\frac{-t_T}{\tau}}\right) \qquad \text{(A)} \tag{14.36}$$

After t_x, $v_o(t) = E$ and the load current is zero, that is

$$i_o(t) = 0 \qquad \text{for} \quad t_x \leq t \leq T \tag{14.37}$$

The output ripple current, for discontinuous conduction, is dependent of the back emf E and is given by equation (14.36), that is

$$I_{p-p} = \hat{I} = \frac{V_s - E}{R}\left(1 - e^{\frac{-t_T}{\tau}}\right) \tag{14.38}$$

Since $\breve{I} = 0$, the mean output current for discontinuous conduction, is

$$\overline{I}_o = \frac{1}{T}\int_0^{t_x} i_o(t)\,dt = \frac{1}{T}\left[\int_0^{t_T} \frac{V_s - E}{R}\left(1 - e^{\frac{-t}{\tau}}\right) dt + \int_0^{t_x - t_T} \frac{-E}{R}\left(1 - e^{\frac{-t}{\tau}}\right) + \hat{I}\,e^{\frac{-t}{\tau}}\,dt\right]$$

$$= \left(\overline{V}_o - E\right) \Big/ R$$

$$\overline{I}_o = \frac{\delta V_s + \left(1 - \frac{t_x}{T}\right)E}{R} - \frac{E}{R} = \left(\delta V_s - \frac{t_x}{T}E\right) \Big/ R \qquad \text{(A)} \tag{14.39}$$

The input and output powers are related such that

$$P_{in} = V_s \overline{I}_i \qquad P_{out} = I_{o\,rms}^2 R + E\overline{I}_o \qquad P_{in} = P_{out} \tag{14.40}$$

from which the average input current can be evaluated.

Alternatively the average input current, which is the switch average current, is given by

$$\overline{I}_i = \overline{I}_{switch} = \frac{1}{T}\int_0^{t_T} i_o(t)\,dt$$

$$= \frac{1}{T}\int_0^{t_T} \frac{V_s - E}{R}\left(1 - e^{\frac{-t}{\tau}}\right) dt \tag{14.41}$$

$$= \frac{V_s - E}{R}\left(\delta - \frac{\tau}{T}\left(1 - e^{\frac{-t_T}{\tau}}\right)\right) = \frac{V_s - E}{R}\delta - \frac{\tau}{T}\hat{I}$$

The average diode current $\overline{I}_{diode}$ is the difference between the average output current $\overline{I}_o$ and the average input current, $\overline{I}_i$, that is

$$\overline{I}_{diode} = \overline{I}_o - \overline{I}_i$$

$$= \frac{\tau}{T}\hat{I} - \frac{E\left(\frac{t_x}{T} - \delta\right)}{R} \tag{14.42}$$

Alternatively, the average diode current can be found by integrating the diode current given in equation (14.35), as follows

$$\overline{I}_{diode} = \frac{1}{T}\int_0^{t_x - t_T}\left(-\frac{E}{R}\left(1 - e^{\frac{-t}{\tau}}\right) + \hat{I}\,e^{\frac{-t}{\tau}}\right) dt$$

$$= \frac{\tau}{T}\hat{I} - \frac{E\left(\frac{t_x}{T} - \delta\right)}{R} \tag{14.43}$$

If E represents motor back emf, then electromagnetic energy conversion efficiency is given by

$$\eta = \frac{E\overline{I}_o}{P_{in}} = \frac{E\overline{I}_o}{V_s\overline{I}_i} \tag{14.44}$$

The chopper effective input impedance is given by

$$Z_{in} = \frac{V_s}{\overline{I}_i} \tag{14.45}$$

Example 14.1: ***DC chopper (first quadrant) with load back emf***

A first-quadrant dc-to-dc chopper feeds an inductive load of 10 Ω resistance, 50mH inductance, and back emf of 55V dc, from a 340V dc source. If the chopper is operated at 200Hz with a 25% on-state duty cycle, determine, with and without (rotor standstill, $E = 0$) the back emf:

i. the load average and rms voltages;
ii. the rms ripple voltage, hence ripple factor;
iii. the maximum and minimum output current, hence the peak-to-peak output ripple in the current;
iv. the current in the time domain;
v. the average load output current, average switch current, and average diode current;
vi. the input power, hence output power and rms output current;
vii. effective input impedance, (and electromagnetic efficiency for $E > 0$); and
viii. sketch the output current and voltage waveforms.

Solution

The main circuit and operating parameters are

- on-state duty cycle δ = ¼
- period $T = 1/f_s$ = 1/200Hz = 5ms
- on-period of the switch t_T = 1.25ms
- load time constant $\tau = L/R$ = 0.05mH/10Ω = 5ms

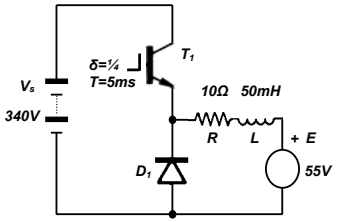


Figure 14.6. Example 14.1. *Circuit diagram.*

i. From equations (14.2) and (14.3), assuming continuous load current, the average and rms output voltages are both independent of the back emf, namely

$$\overline{V}_o = \frac{t_T}{T} V_s = \delta V_s$$
$$= ¼ \times 340\text{V} = 85\text{V}$$
$$V_r = \sqrt{\frac{t_T}{T}}\, V_s = \sqrt{\delta}\, V_s$$
$$= \sqrt{¼} \times 240\text{V} = 120\text{V rms}$$

ii. The rms ripple voltage hence ripple factor are given by equations (14.4) and (14.5), that is

$$V_r = \sqrt{V_{rms}^2 - V_o^2} = V_s\sqrt{\delta(1-\delta)}$$
$$= 340\text{V}\sqrt{¼ \times (1 - ¼)} = 147.2\text{V ac}$$

and

$$RF = \frac{V_r}{V_o} = \sqrt{\frac{1}{\delta} - 1} = \sqrt{FF^2 - 1}$$
$$= \sqrt{\frac{1}{¼} - 1} = \sqrt{3} = 1.732 \qquad FF = 2$$

No back emf, *E* = 0

iii. From equation (14.13), with E = 0, the maximum and minimum currents are

$$\hat{I} = \frac{V_s}{R}\frac{1 - e^{\frac{-t_T}{\tau}}}{1 - e^{\frac{-T}{\tau}}} = \frac{340\text{V}}{10\Omega} \times \frac{1 - e^{\frac{-1.25\text{ms}}{5\text{ms}}}}{1 - e^{\frac{-5\text{ms}}{5\text{ms}}}} = 11.90\text{A}$$
$$\check{I} = \frac{V_s}{R}\frac{e^{\frac{t_T}{\tau}} - 1}{e^{\frac{T}{\tau}} - 1} = \frac{340\text{V}}{10\Omega} \times \frac{e^{¼} - 1}{e^{1} - 1} = 5.62\text{A}$$

The peak-to-peak ripple in the output current is therefore

$$I_{p-p} = \hat{I} - \check{I}$$
$$= 11.90\text{A} - 5.62\text{A} = 6.28\text{A}$$

Alternatively the ripple can be extracted from figure 14.4 using T/τ =1 and δ = ¼.

iv. From equations (14.11) and (14.12), with E = 0, the time domain load current equations are

$$i_o = \frac{V_s}{R}\left(1 - e^{\frac{-t}{\tau}}\right) + \check{I}e^{\frac{-t}{\tau}}$$
$$i_o(t) = 34 \times \left(1 - e^{\frac{-t}{5ms}}\right) + 5.62 \times e^{\frac{-t}{5\text{ms}}}$$
$$= 34 - 28.38 \times e^{\frac{-t}{5\text{ms}}} \quad \text{(A)} \qquad \text{for} \quad 0 \le t \le 1.25\text{ms}$$

$$i_o = \hat{I}e^{\frac{-t}{\tau}}$$
$$i_o(t) = 11.90 \times e^{\frac{-t}{5\text{ms}}} \quad \text{(A)} \qquad \text{for} \quad 0 \le t \le 3.75\text{ms}$$

v. The average load current from equation (14.17), with E = 0, is

$$\overline{I}_o = \overline{V}_o / R = 85\text{V} / 10\Omega = 8.5\text{A}$$

The average switch current, which is the average supply current, is

$$\overline{I}_i = \overline{I}_{switch} = \frac{\delta(V_s - E)}{R} - \frac{\tau}{T}\left(\hat{I} - \check{I}\right)$$
$$= \frac{¼ \times (340\text{V} - 0)}{10\Omega} - \frac{5\text{ms}}{5\text{ms}} \times (11.90\text{A} - 5.62\text{A}) = 2.22\text{A}$$

The average diode current is the difference between the average load current and the average input current, that is

$$\overline{I}_{diode} = \overline{I}_o - \overline{I}_i$$
$$= 8.50\text{A} - 2.22\text{A} = 6.28\text{A}$$

vi. The input power is the dc supply voltage multiplied by the average input current, that is

$$P_{in} = V_s \overline{I}_i = 340\text{V} \times 2.22\text{A} = 754.8\text{W}$$
$$P_{out} = P_{in} = 754.8\text{W}$$

From equation (14.18) the rms load current is given by

$$\overline{I}_{o_{rms}} = \sqrt{\frac{P_{out}}{R}}$$
$$= \sqrt{\frac{754.8\text{W}}{10\Omega}} = 8.7\text{A rms}$$

vii. The chopper effective input impedance is

$$Z_{in} = \frac{V_s}{\overline{I}_i}$$
$$= \frac{340\text{V}}{2.22\text{A}} = 153.2\,\Omega$$

Load back emf, *E* = 55V

i. and *ii.* The average output voltage (85V), rms output voltage (120V rms), ac ripple voltage (147.2V ac), and ripple factor (1.732) are independent of back emf, provided the load current is continuous. The earlier answers for E = 0 are applicable.

iii. From equation (14.13), the maximum and minimum load currents are

$$\hat{I} = \frac{V_s}{R}\frac{1 - e^{\frac{-t_T}{\tau}}}{1 - e^{\frac{-T}{\tau}}} - \frac{E}{R} = \frac{340\text{V}}{10\Omega} \times \frac{1 - e^{\frac{-1.25\text{ms}}{5\text{ms}}}}{1 - e^{\frac{-5\text{ms}}{5\text{ms}}}} - \frac{55\text{V}}{10\Omega} = 6.40\text{A}$$
$$\check{I} = \frac{V_s}{R}\frac{e^{\frac{t_T}{\tau}} - 1}{e^{\frac{T}{\tau}} - 1} - \frac{E}{R} = \frac{340\text{V}}{10\Omega} \times \frac{e^{¼} - 1}{e^{1} - 1} - \frac{55\text{V}}{10\Omega} = 0.12\text{A}$$

The peak-to-peak ripple in the output current is therefore

$$I_{p-p} = \hat{I} - \check{I}$$
$$= 6.4\text{A} - 0.12\text{A} = 6.28\text{A}$$

The ripple value is the same as the E = 0 case, which is as expected since ripple current is independent of back emf with continuous output current.
Alternatively the ripple can be extracted from figure 14.4 using T/τ = 1 and δ = ¼.

iv. The time domain load current is defined by

$$i_o = \frac{V_s - E}{R}\left(1 - e^{\frac{-t}{\tau}}\right) + \check{I}e^{\frac{-t}{\tau}}$$
$$i_o(t) = 28.5 \times \left(1 - e^{\frac{-t}{5ms}}\right) + 0.12e^{\frac{-t}{5\text{ms}}}$$
$$= 28.5 - 28.38e^{\frac{-t}{5\text{ms}}} \quad \text{(A)} \qquad \text{for} \quad 0 \le t \le 1.25\text{ms}$$

$$i_o = -\frac{E}{R}\left(1-e^{\frac{-t}{\tau}}\right)+\hat{I}e^{\frac{-t}{\tau}}$$

$$i_o(t) = -5.5\times\left(1-e^{\frac{-t}{5\text{ms}}}\right)+6.4e^{\frac{-t}{5ms}}$$

$$= -5.5+11.9e^{\frac{-t}{5\text{ms}}} \quad \text{(A)} \qquad \text{for} \quad 0 \le t \le 3.75\text{ms}$$

v. The average load current from equation (14.39) is

$$\bar{I}_o = {}^{V_o - E}\!/_{R}$$

$$= 85\text{V-}55\text{V}/10\Omega = 3\text{A}$$

The average switch current is the average supply current,

$$\bar{I}_i = \bar{I}_{switch} = \frac{\delta(V_s - E)}{R} - \frac{\tau}{T}\left(\hat{I}-\check{I}\right)$$

$$= \frac{¼\times(340\text{V} - 55\text{V})}{10\Omega} - \frac{5\text{ms}}{5\text{ms}}\times(6.40\text{A} - 0.12\text{A}) = 0.845\text{A}$$

The average diode current is the difference between the average load current and the average input current, that is

$$\bar{I}_{diode} = \bar{I}_o - \bar{I}_i$$

$$= 3\text{A} - 0.845\text{A} = 2.155\text{A}$$

vi. The input power is the dc supply voltage multiplied by the average input current, that is

$$P_{in} = V_s\bar{I}_i = 340\text{V}\times 0.845\text{A} = 287.3\text{W}$$

$$P_{out} = P_{in} = 287.3\text{W}$$

From equation (14.18) the rms load current is given by

$$\bar{I}_{o_{rms}} = \sqrt{\frac{P_{out} - E\bar{I}_o}{R}}$$

$$= \sqrt{\frac{287.3\text{W} - 55\text{V}\times 3\text{A}}{10\Omega}} = 3.5\text{A rms}$$

vii. The chopper effective input impedance is

$$Z_{in} = \frac{V_s}{\bar{I}_i}$$

$$= \frac{340\text{V}}{0.845\text{A}} = 402.4\,\Omega$$

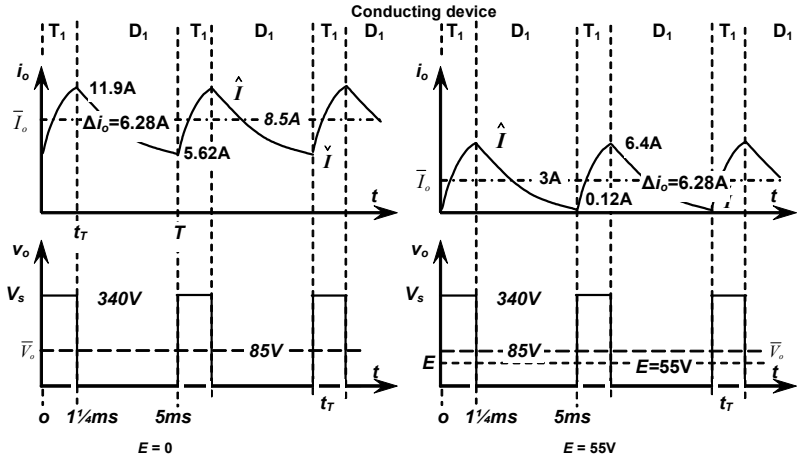


Figure 14.7. Example 14.1. *Circuit waveforms.*

The electromagnetic efficiency is given by equation (14.22), that is

$$\eta = \frac{E\bar{I}_o}{P_{in}}$$

$$= \frac{55\text{V}\times 3\text{A}}{287.3\text{W}} = 57.4\%$$

viii. The output voltage and current waveforms for the first-quadrant chopper, with and without back emf, are shown in figure 14.7.

♣

Example 14.2: ***DC chopper with load back emf - verge of discontinuous conduction***

A first-quadrant dc-to-dc chopper feeds an inductive load of 10 Ω resistance, 50mH inductance, and back emf of 55V dc, from a 340V dc voltage source. If the chopper is operated at 200Hz with a 25% on-state duty cycle, determine:

i. the maximum back emf before discontinuous load current conduction commences with δ=¼;
ii. with 55V back emf, what is the minimum duty cycle before discontinuous load current conduction; and
iii. minimum switching frequency at E = 55V and t_T = 1.25ms before discontinuous conduction.

Solution

The main circuit and operating parameters are

- on-state duty cycle δ = ¼
- period $T = 1/f_s$ = 1/200Hz = 5ms
- on-period of the switch t_T = 1.25ms
- load time constant $\tau = L/R$ = 0.05mH/10Ω = 5ms

First it is necessary to establish whether the given conditions represent continuous or discontinuous load current. The current extinction time t_x for discontinuous conduction is given by equation (14.24), and yields

$$t_x = t_T + \tau\,\ell n\left(1+\frac{V_s - E}{E}\left(1-e^{\frac{-t_T}{\tau}}\right)\right)$$

$$= 1.25\text{ms} + 5\text{ms}\times\ell n\left(1+\frac{340\text{V - }55\text{V}}{55\text{V}}\times\left(1-e^{\frac{-1.25\text{ms}}{5\text{ms}}}\right)\right) = 5.07\text{ms}$$

Since the cycle period is 5ms, which is less than the necessary time for the current to fall to zero (5.07ms), the load current is continuous. From example 14.1 part iv, with E = 55V the load current falls from 6.4A to near zero (0.12A) at the end of the off-time, thus the chopper is operating near the verge of discontinuous conduction. A small increase in E, decrease in the duty cycle δ, or increase in switching period T, would be expected to result in discontinuous load current.

i. $\hat{E}$

The necessary back emf can be determined graphically or analytically.

Graphically:

The bounds of continuous and discontinuous load current for a given duty cycle, switching period, and load time constant can be determined from figure 14.5.
Using δ = ¼, T/τ = 1 with τ = 5ms, and T = 5ms, figure 14.5 gives E / V_s = 0.165. That is, E = 0.165×V_s = 0.165×340V = 56.2V

Analytically:

The chopper is operating too close to the boundary between continuous and discontinuous load current conduction for accurate readings to be obtained from the graphical approach, using figure 14.5. Examination of the expression for minimum current, equation (14.13), gives

$$\check{I} = \frac{V_s}{R}\frac{e^{\frac{t_T}{\tau}}-1}{e^{\frac{T}{\tau}}-1} - \frac{E}{R} = 0$$

Rearranging to give the back emf, *E*, produces

$$E = V_s \frac{e^{\frac{t_T}{\tau}} - 1}{e^{\frac{T}{\tau}} - 1}$$

$$= 340\text{V} \times \frac{e^{\frac{1.25\text{ms}}{5\text{ms}}} - 1}{e^{\frac{5\text{ms}}{5\text{ms}}} - 1} = 56.2\text{V}$$

That is, if the back emf increases from 55V to 56.2V then at and above that voltage, discontinuous load current commences.

ii. $\hat{\delta}$

Again, if equation (14.13) is solved for $\check{I} = 0$ then

$$\check{I} = \frac{V_s}{R} \frac{e^{\frac{t_T}{\tau}} - 1}{e^{\frac{T}{\tau}} - 1} - \frac{E}{R} = 0$$

Rearranging to isolate t_T gives

$$t_T = \tau \, \ell n\left(1 + \frac{E}{V_s}\left(e^{\frac{T}{\tau}} - 1\right)\right)$$

$$= 5\text{ms} \times \ell n\left(1 + \frac{55\text{V}}{340\text{V}}\left(e^{\frac{5\text{ms}}{5\text{ms}}} - 1\right)\right)$$

$$= 1.226\text{ms}$$

If the switch on-state period is reduced by 0.024ms, from 1.250ms to 1.226ms (δ = 24.52%), operation is then on the verge of discontinuous conduction.

iii. $\hat{T}$

If the switching frequency is decreased such that $T = t_x$, then the minimum period for discontinuous load current is given by equation (14.24). That is,

$$t_x = T = t_T + \tau \, \ell n\left(1 + \frac{V_s - E}{E}\left(1 - e^{\frac{-t_T}{\tau}}\right)\right)$$

$$T = 1.25\text{ms} + 5\text{ms} \times \ell n\left(1 + \frac{340\text{V} - 55\text{V}}{55\text{V}} \times \left(1 - e^{\frac{-1.25\text{ms}}{5\text{ms}}}\right)\right) = 5.07\text{ms}$$

Discontinuous conduction operation occurs if the period is increased by more than 0.07ms.

In conclusion, for the given load, for continuous conduction to cease, the following operating conditions can be changed

- increase the back emf *E* from 55V to 56.2V
- decrease the duty cycle δ from 25% to 24.52% (t_T decreased from 1.25ms to 1.226ms)
- increase the switching period *T* by 0.07ms, from 5ms to 5.07ms (from 200Hz to 197.2Hz), with the switch on-time, t_T, unchanged from 1.25ms.

Appropriate simultaneous smaller changes in more than one parameter would suffice.

♣

Example 14.3: ***DC chopper with load back emf – discontinuous conduction***

A first-quadrant dc-to-dc chopper feeds an inductive load of 10 Ω resistance, 50mH inductance, and an opposing back emf of 100V dc, from a 340V dc source. If the chopper is operated at 200Hz with a 25% on-state duty cycle, determine:

i. the load average and rms voltages;
ii. the rms ripple voltage, hence ripple and form factors;
iii. the maximum and minimum output current, hence the peak-to-peak output ripple in the current;
iv. the current in the time domain;
v. the load average current, average switch current and average diode current;
vi. the input power, hence output power and rms output current;
vii. effective input impedance, and electromagnetic efficiency; and
viii. sketch the circuit, load, and output voltage and current waveforms.

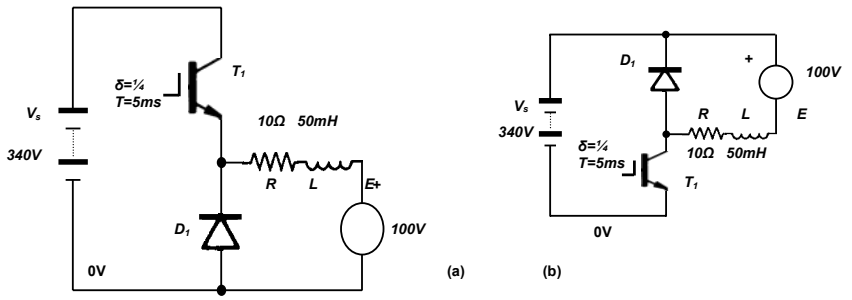


Figure 14.8. Example 14.3. *Circuit diagram: (a) with load connected to ground and (b) load connected so that machine flash-over to ground* (0V), *by-passes the switch T_1.*

Solution

The main circuit and operating parameters are

- on-state duty cycle δ = ¼
- period $T = 1/f_s$ = 1/200Hz = 5ms
- on-period of the switch t_T = 1.25ms
- load time constant $\tau = L/R$ = 0.05mH/10Ω = 5ms

Confirmation of discontinuous load current can be obtained by evaluating the minimum current given by equation (14.13), that is

$$\check{I} = \frac{V_s}{R} \frac{e^{\frac{t_T}{\tau}} - 1}{e^{\frac{T}{\tau}} - 1} - \frac{E}{R}$$

$$\check{I} = \frac{340\text{V}}{10\Omega} \times \frac{e^{\frac{1.25\text{ms}}{5\text{ms}}} - 1}{e^{\frac{5\text{ms}}{5\text{ms}}} - 1} - \frac{100\text{V}}{10\Omega} = 5.62\text{A} - 10\text{A} = -4.38\text{A}$$

The minimum practical current is zero, so clearly discontinuous current periods exist in the load current. The equations applicable to discontinuous load current need to be employed.
The current extinction time is given by equation (14.24), that is

$$t_x = t_T + \tau \, \ell n\left(1 + \frac{V_s - E}{E}\left(1 - e^{\frac{-t_T}{\tau}}\right)\right)$$

$$= 1.25\text{ms} + 5\text{ms} \times \ell n\left(1 + \frac{340\text{V} - 100\text{V}}{100\text{V}} \times \left(1 - e^{\frac{-1.25\text{ms}}{5\text{ms}}}\right)\right)$$

$$= 1.25\text{ms} + 2.13\text{ms} = 3.38\text{ms}$$

i. From equations (14.28) and (14.29) the load average and rms voltages are

$$\overline{V}_o = \delta V_s + \frac{T - t_x}{T} E$$

$$= ¼ \times 340\,\text{V} + \frac{5\text{ms} - 3.38\text{ms}}{5\text{ms}} \times 100\text{V} = 117.4\text{V}$$

$$V_{rms} = \sqrt{\delta V_s^2 + \frac{T - t_x}{T} E^2}$$

$$= \sqrt{¼ \times 340^2 + \frac{5\text{ms} - 3.38\text{ms}}{5\text{ms}} \times 100^2} = 179.3\text{V rms}$$

ii. From equations (14.30) and (14.31) the rms ripple voltage, hence voltage ripple factor, are

$$V_r = \sqrt{V_{rms}^2 - V_o^2}$$

$$= \sqrt{179.3^2 - 117.4^2} = 135.5\text{V ac}$$

$$RF = \frac{V_r}{\overline{V}_o} = \frac{135.5\text{V}}{117.4\text{V}} = 1.15 \qquad FF = \sqrt{RF^2 + 1} = \sqrt{1.15^2 + 1} = 1.52$$

iii. From equation (14.38), the maximum and minimum output current, hence the peak-to-peak output ripple in the current, are

$$\hat{I} = \frac{V_s - E}{R}\left(1 - e^{\frac{-t_T}{\tau}}\right)$$

$$= \frac{340\text{V-}100\text{V}}{10\Omega} \times \left(1 - e^{\frac{-1.25\text{ms}}{5\text{ms}}}\right) = 5.31\text{A}$$

The minimum current is zero so the peak-to-peak ripple current is $\Delta i_o = 5.31\text{A}$.

iv. From equations (14.34) and (14.35), the current in the time domain is

$$i_o(t) = \frac{V_s - E}{R}\left(1 - e^{\frac{-t}{\tau}}\right)$$

$$= \frac{340\text{V - }100\text{V}}{10\Omega} \times \left(1 - e^{\frac{-t}{5\text{ms}}}\right)$$

$$= 24 \times \left(1 - e^{\frac{-t}{5\text{ms}}}\right) \quad \text{(A)} \qquad \text{for} \quad 0 \le t \le 1.25\text{ms}$$

$$i_o(t) = -\frac{E}{R}\left(1 - e^{\frac{-t}{\tau}}\right) + \hat{I}e^{\frac{-t}{\tau}}$$

$$= -\frac{100\text{V}}{10\Omega} \times \left(1 - e^{\frac{-t}{5\text{ms}}}\right) + 5.31e^{\frac{-t}{5\text{ms}}}$$

$$= 15.31 \times e^{\frac{-t}{5\text{ms}}} - 10 \quad \text{(A)} \qquad \text{for} \quad 0 \le t \le 2.13\text{ms}$$

$$i_o(t) = 0 \qquad \text{for} \quad 3.38\text{ms} \le t \le 5\text{ms}$$

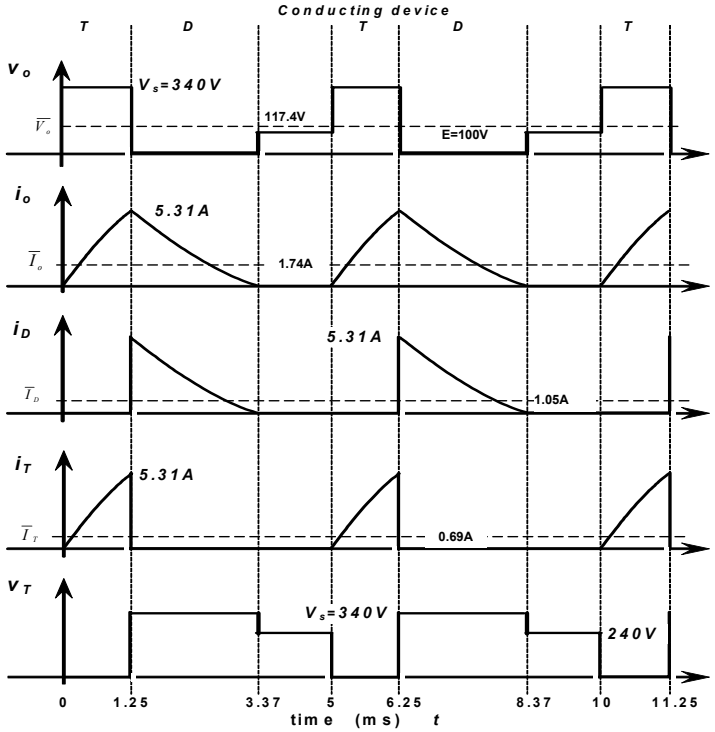


Figure 14.9. Example 14.3. *Chopper circuit waveforms.*

v. From equations (14.39) to (14.42), the average load current, average switch current, and average diode current are

$$\bar{I}_o = \bar{V}_o - E/R$$

$$= 117.4\text{V - }100\text{V}/10\Omega = 1.74\text{A}$$

$$\bar{I}_{diode} = \frac{\tau}{T}\hat{I} - \frac{E\left(\frac{t_x}{T} - \delta\right)}{R}$$

$$= \frac{5\text{ms}}{5\text{ms}} \times 5.31\text{A} - \frac{100\text{V} \times \left(\frac{3.38\text{ms}}{5\text{ms}} - 0.25\right)}{10\Omega} = 1.05\text{A}$$

$$\bar{I}_i = \bar{I}_o - \bar{I}_{diode} = 1.74\text{A} - 1.05\text{A} = 0.69\text{A}$$

vi. From equation (14.40), the input power, hence output power and rms output current are

$$P_{in} = V_s \bar{I}_i = 340\text{V} \times 0.69\text{A} = 234.6\text{W}$$

$$P_{in} = P_{out} = I_{orms}^2 R + E\bar{I}_o$$

Rearranging gives

$$I_{o_{rms}} = \sqrt{P_{in} - E\bar{I}_o}\,/R$$

$$= \sqrt{234.6\text{W - }100\text{V} \times 0.69\text{A}}\,/\,10\Omega = 1.29\text{A}$$

vii. From equations (14.44) and (14.45), the effective input impedance and electromagnetic efficiency, for $E > 0$ are

$$Z_{in} = \frac{V_s}{\bar{I}_i} = \frac{340\text{V}}{0.69\text{A}} = 493\Omega$$

$$\eta = \frac{E\bar{I}_o}{P_{in}} = \frac{E\bar{I}_o}{V_s\bar{I}_i} = \frac{100\text{V} \times 1.74\text{A}}{340\text{V} \times 0.69\text{A}} = 74.2\%$$

viii. The circuit, load, and output voltage and current waveforms are plotted in figure 14.9.

♣

14.3 Second-Quadrant dc chopper

The second-quadrant dc-to-dc chopper shown in figure 14.2b transfers energy from the load, back to the dc energy source V_s, a process called *regeneration*. Its operating principles are the same as those for the boost switch mode power supply analysed in chapter 15.4. The two energy transfer stages are shown in figure 14.10. Controlled energy transfer from the back emf E to the supply V_s, is achieved by varying the switch T_2 on-state duty cycle. Two modes of transfer can occur, as with the first-quadrant chopper already considered. The current in the load inductor can be either continuous or discontinuous, depending on the specific circuit parameters and operating conditions.
In this analysis, and all the choppers analysed, it is assumed that:

- No source impedance;
- Constant switch duty cycle;
- Steady-state conditions have been reached;
- Ideal semiconductors; and
- No load impedance temperature effects.

14.3.1 Continuous load inductor current

Load waveforms for continuous load current conduction are shown in figure 14.11a.
The output voltage v_o, load voltage, or switch voltage, is defined by

$$v_o(t) = \begin{cases} 0 & \text{for } 0 \le t \le t_T \\ V_s & \text{for } t_T \le t \le T \end{cases} \qquad (14.46)$$

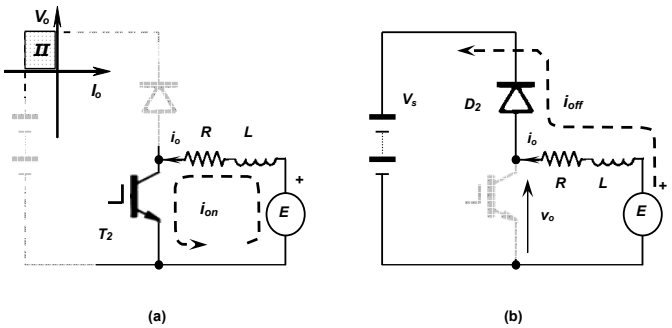


Figure 14.10. *Stages of operation for the second-quadrant chopper: (a) switch-on, boosting current and (b) switch-off, energy into V_s.*

The mean load voltage is

$$\begin{aligned}\overline{V}_o &= \frac{1}{T}\int_0^T v_o(t)\,dt = \frac{1}{T}\int_{t_T}^T V_s\,dt \\ &= \frac{T-t_T}{T}V_s = (1-\delta)V_s\end{aligned} \tag{14.47}$$

where the switch on-state duty cycle $\delta = t_T/T$ is defined in figure 14.11a.

Alternatively the voltage across the dc source V_s is

$$V_s = \frac{1}{1-\delta}\overline{V}_o \tag{14.48}$$

Since $0 \le \delta \le 1$, the step-up voltage ratio, to regenerate into V_s, is continuously adjustable from unity to infinity.
The average output current is

$$\overline{I}_o = \frac{E-\overline{V}_o}{R} = \frac{E-V_s(1-\delta)}{R} \tag{14.49}$$

The average output current can also be found by integration of the time domain output current i_o. By solving the appropriate time domain differential equations, the continuous load current i_o shown in figure 14.11a is defined by

During the **switch on-period**, when $v_o = 0$

$$L\frac{di_o}{dt} + Ri_o = E$$

which yields

$$i_o(t) = \frac{E}{R}\left(1-e^{\frac{-t}{\tau}}\right) + \check{I}e^{\frac{-t}{\tau}} \qquad \text{for} \quad 0 \le t \le t_T \tag{14.50}$$

During the **switch off-period**, when $v_o = V_s$

$$L\frac{di_o}{dt} + Ri_o + V_s = E$$

which, after shifting the zero time reference to t_T, gives

$$i_o(t) = \frac{E-V_s}{R}\left(1-e^{\frac{-t}{\tau}}\right) + \hat{I}e^{\frac{-t}{\tau}} \qquad \text{for} \quad 0 \le t \le T-t_T \tag{14.51}$$

$$\begin{aligned}\text{where} \quad \hat{I} &= \frac{E}{R} - \frac{V_s}{R}\frac{e^{\frac{-t_T}{\tau}} - e^{\frac{-T}{\tau}}}{1-e^{\frac{-T}{\tau}}} \qquad \text{(A)} \\ \text{and} \quad \check{I} &= \frac{E}{R} - \frac{V_s}{R}\frac{1-e^{\frac{-T+t_T}{\tau}}}{1-e^{\frac{-T}{\tau}}} \qquad \text{(A)}\end{aligned} \tag{14.52}$$

The output ripple current, for continuous conduction, is *independent* of the back emf E and is given by

$$I_{p-p} = \hat{I} - \check{I} = \frac{V_s}{R}\frac{(1+e^{\frac{-T}{\tau}}) - (e^{\frac{-t_T}{\tau}} + e^{\frac{-T+t_T}{\tau}})}{1-e^{\frac{-T}{\tau}}} \tag{14.53}$$

which in terms of the on-state duty cycle, $\delta = t_T/T$, becomes

$$I_{p-p} = \frac{V_s}{R}\frac{(1-e^{\frac{-\delta T}{\tau}})(1+e^{\frac{-T}{\tau}})}{1-e^{\frac{-T}{\tau}}} \tag{14.54}$$

This is the same expression derived in 14.2.1 for the first-quadrant chopper. The normalised ripple current design curves in figure 14.3 are valid for the second-quadrant chopper.

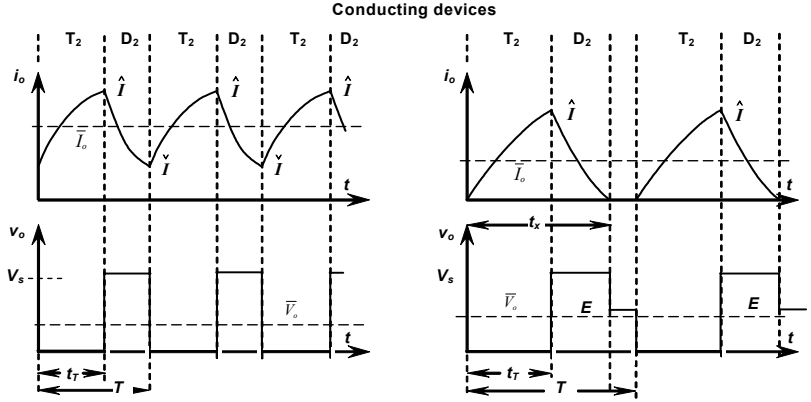


Figure 14.11. *Second-quadrant chopper output modes of current operation: (a) continuous inductor current and (b) discontinuous inductor current.*

The average switch current, $\overline{I}_{switch}$, can be derived by integrating the switch current given by equation (14.50), that is

$$\begin{aligned}\overline{I}_{switch} &= \frac{1}{T}\int_0^{t_T} i_o(t)\,dt \\ &= \frac{1}{T}\int_0^{t_T}\left(\frac{E}{R}\left(1-e^{\frac{-t}{\tau}}\right) + \check{I}e^{\frac{-t}{\tau}}\right)dt \\ &= \frac{\delta E}{R} - \frac{\tau}{T}\left(\hat{I}-\check{I}\right)\end{aligned} \tag{14.55}$$

The term $\hat{I}-\check{I} = I_{p-p}$ is the peak-to-peak ripple current, which is given by equation (14.53). By Kirchhoff's current law, the average diode current $\overline{I}_{diode}$ is the difference between the average output current $\overline{I}_o$ and the average switch current, $\overline{I}_{switch}$, that is

$$\begin{aligned}\overline{I}_{diode} &= \overline{I}_o - \overline{I}_{switch} \\ &= \frac{E-V_s(1-\delta)}{R} - \frac{\delta E}{R} + \frac{\tau}{T}\left(\hat{I}-\check{I}\right) \\ &= \frac{\tau}{T}\left(\hat{I}-\check{I}\right) - \frac{(V_s-E)(1-\delta)}{R}\end{aligned} \tag{14.56}$$

The average diode current can also be found by integrating the diode current given in equation (14.51), as follows

$$\begin{aligned}\overline{I}_{diode} &= \frac{1}{T}\int_0^{T-t_T}\left(\frac{E-V_s}{R}\left(1-e^{\frac{-t}{\tau}}\right) + \hat{I}e^{\frac{-t}{\tau}}\right)dt \\ &= \frac{\tau}{T}\left(\hat{I}-\check{I}\right) - \frac{(V_s-E)(1-\delta)}{R}\end{aligned} \tag{14.57}$$

The power produced (provide) by the back emf source E is

$$P_E = E\overline{I}_o = E\left(\frac{E-V_s(1-\delta)}{R}\right) \tag{14.58}$$

The power delivered to the dc source V_s is

$$P_{V_s} = V_s \bar{I}_{diode} = V_s\left(\frac{\tau}{T}\left(\hat{I} - \check{I}\right) - \frac{(V_s - E)(1-\delta)}{R}\right) \tag{14.59}$$

The difference between the two powers is the power lost in the load resistor, R, that is

$$P_E = P_{V_s} + I_{o_{rms}}^2 R$$

$$I_{o_{rms}} = \sqrt{\frac{E\bar{I}_o - V_s \bar{I}_{diode}}{R}} \tag{14.60}$$

The efficiency of energy transfer between the back emf E and the dc source V_s is

$$\eta = \frac{P_{V_s}}{P_E} = \frac{V_s \bar{I}_{diode}}{E\bar{I}_o} \tag{14.61}$$

14.3.2 Discontinuous load inductor current

With low duty cycles, δ, low inductance, L, or a relatively high dc source voltage, V_s, the minimum output current may reach zero at t_x, before the period T is complete ($t_x < T$), as shown in figure 14.11b. Equation (14.52) gives a boundary identity that must be satisfied for zero current,

$$\check{I} = \frac{E}{R} - \frac{V_s}{R}\frac{1 - e^{\frac{T - t_T}{\tau}}}{1 - e^{\frac{-T}{\tau}}} \quad = \quad 0 \tag{14.62}$$

That is

$$\frac{E}{V_s} = \frac{1 - e^{\frac{-T + t_T}{\tau}}}{1 - e^{\frac{-T}{\tau}}} \tag{14.63}$$

Alternatively, the time domain equations (14.50) and (14.51) can be used, such that $\check{I} = 0$. An expression for the extinction time t_x can be found by substituting $t = t_T$ into equation (14.50). The resulting expression for $\hat{I}$ is then substituted into equation (14.51) which is set to zero. Isolating the time variable, which becomes t_x, yields

$$\hat{I} = \frac{E}{R}\left(1 - e^{\frac{-t_T}{\tau}}\right)$$

$$0 = \frac{E - V_s}{R}\left(1 - e^{\frac{-t_x}{\tau}}\right) + \frac{E}{R}\left(1 - e^{\frac{-t_T}{\tau}}\right)e^{\frac{-t_x}{\tau}}$$

which yields

$$t_x = t_T + \tau\, \ell n\left(1 + \frac{E}{V_s - E}\left(1 - e^{\frac{-t_T}{\tau}}\right)\right) \tag{14.64}$$

This equation shows that $t_x \geq t_T$. Load waveforms for discontinuous load current conduction are shown in figure 14.11b.

The output voltage v_o, load voltage, or switch voltage, is defined by

$$v_o(t) = \begin{cases} 0 & \text{for } 0 \leq t \leq t_T \\ V_s & \text{for } t_T \leq t \leq t_x \\ E & \text{for } t_x \leq t \leq T \end{cases} \tag{14.65}$$

The mean load voltage is

$$\bar{V}_o = \frac{1}{T}\int_0^T v_o(t)\,dt = \frac{1}{T}\left(\int_{t_T}^{t_x} V_s\,dt + \int_{t_x}^{T} E\,dt\right)$$

$$= \frac{t_x - t_T}{T}V_s + \frac{T - t_x}{T}E = \left(\frac{t_x}{T} - \delta\right)V_s + \left(1 - \frac{t_x}{T}\right)E \tag{14.66}$$

$$\bar{V}_o = E - \delta V_s + \frac{t_x}{T}(V_s - E)$$

where the switch on-state duty cycle $\delta = t_T/T$ is defined in figure 14.11b.

The average output current is

$$\bar{I}_o = \frac{E - \bar{V}_o}{R} = \frac{\delta V_s - \frac{t_x}{T}(V_s - E)}{R} \tag{14.67}$$

The average output current can also be found by integration of the time domain output current i_o. By solving the appropriate time domain differential equations, the continuous load current i_o shown in figure 14.11a is defined by

During the **switch on-period**, when $v_o = 0$

$$L\frac{di_o}{dt} + R i_o = E$$

which yields

$$i_o(t) = \frac{E}{R}\left(1 - e^{\frac{-t}{\tau}}\right) \qquad \text{for} \quad 0 \leq t \leq t_T \tag{14.68}$$

During the **switch off-period**, when $v_o = V_s$

$$L\frac{di_o}{dt} + R i_o + V_s = E$$

which, after shifting the zero time reference to t_T, gives

$$i_o(t) = \frac{E - V_s}{R}\left(1 - e^{\frac{-t}{\tau}}\right) + \hat{I}e^{\frac{-t}{\tau}} \qquad \text{for} \quad 0 \leq t \leq t_x - t_T \tag{14.69}$$

$$\text{where} \quad \hat{I} = \frac{E}{R}\left(1 - e^{\frac{-t_T}{\tau}}\right) \qquad \text{(A)}$$

$$\text{and} \quad \check{I} = 0 \qquad \text{(A)} \tag{14.70}$$

After t_x, $v_o(t) = E$ and the load current is zero, that is

$$i_o(t) = 0 \qquad \text{for} \quad t_x \leq t \leq T \tag{14.71}$$

The output ripple current, for discontinuous conduction, is dependent of the back emf E and is given by equation (14.70),

$$I_{p-p} = \hat{I} = \frac{E}{R}\left(1 - e^{\frac{-t_T}{\tau}}\right) \tag{14.72}$$

The average switch current, $\bar{I}_{switch}$, can be derived by integrating the switch current given by equation (14.68), that is

$$\bar{I}_{switch} = \frac{1}{T}\int_0^{t_T} i_o(t)\,dt$$

$$= \frac{1}{T}\int_0^{t_T}\left(\frac{E}{R}\left(1 - e^{\frac{-t}{\tau}}\right)\right)dt \tag{14.73}$$

$$= \frac{\delta E}{R} - \frac{\tau}{T}\hat{I}$$

The term $\hat{I} = I_{p-p}$ is the peak-to-peak ripple current, which is given by equation (14.72). By Kirchhoff's current law, the average diode current $\bar{I}_{diode}$ is the difference between the average output current $\bar{I}_o$ and the average switch current, $\bar{I}_{switch}$, that is

$$\bar{I}_{diode} = \bar{I}_o - \bar{I}_{switch}$$

$$= \frac{\delta V_s - \frac{t_x}{T}(V_s - E)}{R} - \frac{\delta E}{R} + \frac{\tau}{T}\hat{I} \tag{14.74}$$

$$= \frac{\tau}{T}\hat{I} - \frac{\left(\frac{t_x}{T} - \delta\right)(V_s - E)}{R}$$

The average diode current can also be found by integrating the diode current given in equation (14.69), as follows

$$\bar{I}_{diode} = \frac{1}{T}\int_0^{t_x - t_T}\left(\frac{E - V_s}{R}\left(1 - e^{\frac{-t}{\tau}}\right) + \hat{I}e^{\frac{-t}{\tau}}\right)dt$$

$$= \frac{\tau}{T}\hat{I} - \frac{\left(\frac{t_x}{T} - \delta\right)(V_s - E)}{R} \tag{14.75}$$

The power produced by the back emf source E is

$$P_E = E\bar{I}_o \tag{14.76}$$

The power delivered to the dc source V_s is

$$P_{V_s} = V_s \bar{I}_{diode} \tag{14.77}$$

Alternatively, the difference between the two powers is the power lost in the load resistor, R, that is

$$P_E = P_{V_s} + I_{o_{rms}}^2 R$$

$$I_{o_{rms}} = \sqrt{\frac{E\bar{I}_o - V_s\bar{I}_{diode}}{R}} \qquad (14.78)$$

The efficiency of energy transfer between the back emf and the dc source is

$$\eta = \frac{P_{V_s}}{P_E} = \frac{V_s\bar{I}_{diode}}{E\bar{I}_o} \qquad (14.79)$$

Example 14.4: ***Second-quadrant DC chopper – continuous inductor current***

A dc-to-dc chopper capable of second-quadrant operation is used in a 200V dc battery electric vehicle. The machine armature has 1 Ω resistance in series with 1mH inductance.

i. The machine is used for regenerative braking. At a constant speed downhill, the back emf is 150V, which results in a 10A braking current. What is the switch on-state duty cycle if the machine is delivering continuous output current? What is the minimum chopping frequency for these conditions?

ii. At this speed, (that is, E = 150V), determine the minimum duty cycle for continuous inductor current, if the switching frequency is 1kHz. What is the average braking current at the critical duty cycle? What is the regenerating efficiency and the rms machine output current?

iii. If the chopping frequency is increased to 5kHz, at the same speed, (that is, E = 150V), what is the critical duty cycle and the corresponding average dc machine current?

Solution

The main circuit operating parameters are

- V_s = 200V
- E = 150V
- load time constant $\tau = L/R$ = 1mH/1Ω = 1ms

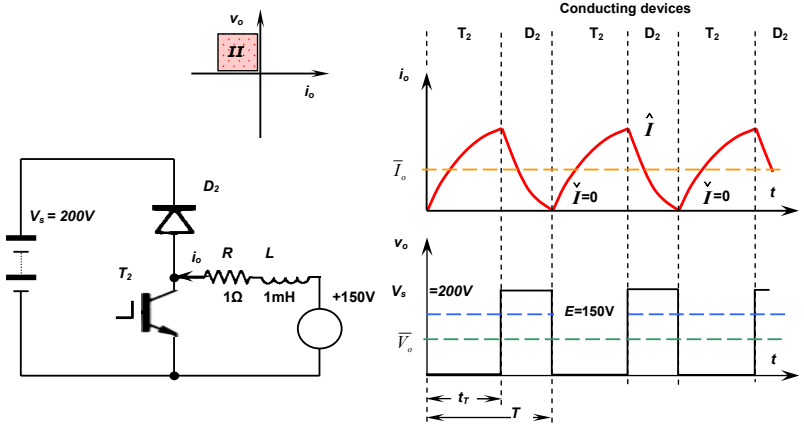


Figure 14.12. Example 14.4. *Circuit diagram and waveforms.*

i. The relationship between the dc supply V_s and the dc machine back emf E is given by equation (14.49), that is

$$\bar{I}_o = \frac{E - \bar{V}_o}{R} = \frac{E - V_s(1-\delta)}{R}$$

$$10\text{A} = \frac{150\text{V} - 200\text{V}\times(1 - \delta)}{1\Omega}$$

that is

$$\delta = 0.3 \equiv 30\% \quad \text{and} \quad \bar{V}_o = 140\text{V}$$

The expression for the average dc machine output current is based on continuous armature inductance current. Therefore the switching period must be shorter than the time t_x predicted by equation (14.64) for the current to reach zero, before the next switch on-period. That is, for $t_x = T$ and $\delta = 0.3$

$$t_x = t_T + \tau\,\ell n\left(1 + \frac{E}{V_s - E}\left(1 - e^{\frac{-t_T}{\tau}}\right)\right)$$

This simplifies to

$$1 = 0.3 + \frac{1\text{ms}}{T}\ell n\left(1 + \frac{150\text{V}}{200\text{V - }150\text{V}}\left(1 - e^{\frac{-0.3T}{1\text{ms}}}\right)\right)$$

$$e^{0.7T} = 4 - 3e^{-0.3T}$$

Iteratively solving this transcendental equation gives T = 0.4945ms. That is the switching frequency must be greater than $f_s = 1/T$ = 2.022kHz, else machine output current discontinuities occur, and equation (14.49) is invalid. The switching frequency can be reduced if the on-state duty cycle is increased as in the next part of this example.

ii. The operational boundary condition giving by equation (14.63), using $T = 1/f_s$ =1/1kHz = 1ms, yields

$$\frac{E}{V_s} = \frac{1 - e^{\frac{-T + t_T}{\tau}}}{1 - e^{\frac{-T}{\tau}}}$$

$$\frac{150\text{V}}{200\text{V}} = \frac{1 - \text{e}^{\frac{(\delta - 1)\times 1\text{ms}}{1\text{ms}}}}{1 - \text{e}^{\frac{-1\text{ms}}{1\text{ms}}}}$$

Solving gives δ = 0.357. That is, the on-state duty cycle must be at least 35.7% for continuous machine output current at a switching frequency of 1kHz.

For continuous inductor current, the average output current is given by equation (14.49), that is

$$\bar{I}_o = \frac{E - \bar{V}_o}{R} = \frac{E - V_s(1-\delta)}{R}$$

$$= \frac{150\text{V - }\bar{V}_o}{1\Omega} = \frac{150\text{V - }200\text{V}\times(1 - 0.357)}{1\Omega} = 21.4\text{A}$$

$$\bar{V}_o = 150\text{V - }21.4\text{A}\times 1\Omega = 128.6\text{V}$$

The average machine output current of 21.4A is split between the switch and the diode (which is in series with V_s).

The diode current is given by equation (14.56)

$$\bar{I}_{diode} = \bar{I}_o - \bar{I}_{switch}$$

$$= \frac{\tau}{T}\left(\hat{I} - \check{I}\right) - \frac{(V_s - E)(1-\delta)}{R}$$

The minimum output current is zero while the maximum is given by equation (14.70).

$$\hat{I} = \frac{E}{R}\left(1 - e^{\frac{-t_T}{\tau}}\right) = \frac{150\text{V}}{1\Omega}\times\left(1 - \text{e}^{\frac{-0.357\times 1\text{ms}}{1\text{ms}}}\right) = 45.0\text{A}$$

Substituting into the equation for the average diode current gives

$$\bar{I}_{diode} = \frac{1\text{ms}}{1\text{ms}}\times(45.0\text{A - }0\text{A}) - \frac{(200\text{V - }150\text{V})\times(1 - 0.357)}{1\Omega} = 12.85\text{A}$$

The power delivered by the dc machine back emf E is

$$P_E = E\bar{I}_o = 150\text{V}\times 21.4\text{A} = 3210\text{W}$$

while the power delivered to the 200V battery source V_s is

$$P_{V_s} = V_s\bar{I}_{diode} = 200\text{V}\times 12.85\text{A} = 2570\text{W}$$

The regeneration transfer efficiency is

$$\eta = \frac{P_{V_s}}{P_E} = \frac{2570\text{W}}{3210\text{W}} = 80.1\%$$

The energy generated deficit, 640W (3210W - 2570W)), is lost in the armature resistance, as I^2R heat dissipation. The output rms current is

$$I_{o_{rms}} = \sqrt{\frac{P}{R}} = \sqrt{\frac{640\text{W}}{1\Omega}} = 25.3\text{A rms}$$

iii. At an increased switching frequency of 5kHz, the duty cycle would be expected to be much lower than the 35.7% as at 1kHz. The operational boundary between continuous and discontinuous armature inductor current is given by equation (14.63), that is

$$\frac{E}{V_s} = \frac{1 - e^{\frac{-T+t_T}{\tau}}}{1 - e^{\frac{-T}{\tau}}}$$

$$\frac{150\text{V}}{200\text{V}} = \frac{1 - \text{e}^{\frac{(-1+\delta)\times 0.2\text{ms}}{1\text{ms}}}}{1 - \text{e}^{\frac{-0.2\text{ms}}{1\text{ms}}}}$$

which yields δ = 26.9% .
The machine average output current is given by equation (14.49)

$$\bar{I}_o = \frac{E - \bar{V}_o}{R} = \frac{E - V_s(1-\delta)}{R}$$

$$= \frac{150\text{V} - \bar{V}_o}{1\Omega} = \frac{150\text{V} - 200\text{V}\times(1 - 0.269)}{1\Omega} = 3.8\text{A}$$

such that the average output voltage $\bar{V}_o$ is 146.2V.

♣

14.4 Two-quadrant dc chopper - Q I and Q II

Figure 14.13 shows the basic two-quadrant dc chopper, which is a reproduction of the circuit in figure 14.2c. Depending on the load and operating conditions, the chopper can seamlessly change between and act in two modes

- Devices T_1 and D_1 form the first-quadrant chopper shown in figure 14.2a, and is analysed in section 14.2. Energy is delivered from the dc source V_s to the *R-L-E* load.
- Devices T_2 and D_2 form the second-quadrant chopper shown in figure 14.2b, which is analysed in section 14.3. Energy is delivered from the generating load dc source *E*, to the dc source V_s.

The two independent choppers can be readily combined as shown in figure 14.13a.
The average output voltage $\bar{V}_o$ and the instantaneous output voltage v_o are never negative, whilst the average source current of V_s can be positive (Quadrant I) or negative (Quadrant II). If the two choppers are controlled to operate independently, with the constraint that T_1 and T_2 do not conduct simultaneously, then the analysis in sections 14.2 and 14.3 are valid. Alternately, it is not uncommon the unify the operation of the two choppers, as follows.

If the chopper is operated such that the switches T_1 and T_2 act in a complementary manner, that is either T_1 or T_2 is on, then some of the independent flexibility offered by each chopper is lost. Essentially the consequence of complementary switch operation is that no extended zero current periods exist in the output, as shown in figures 14.13a and b. Thus the equations describing the features of the first-quadrant chopper in section 14.2.1, for continuous load current, are applicable to this chopper, with slight modification to account for the fact that both the minimum and maximum currents can be negative.

The analysis for continuous inductor current in section 14.2 is valid, but the minimum current is not restricted to zero. Consequently four possible output modes can occur, depending on the relative polarity of the maximum and minimum currents shown in figure 14.13b and c.

i. $\check{I} > 0$, $\hat{I} > 0$ and $\bar{I}_o > 0$
When the minimum current (hence average output current) is greater than zero, the chopper is active in the first-quadrant. Typical output voltage and current waveforms are shown in figure 14.3a. The switch T_2 and diode D_2 do not conduct during any portion of the operating period.

ii. $\check{I} < 0$, $\hat{I} > 0$ and $\bar{I}_o > 0$
When the minimum current is negative but the maximum positive current is larger in absolute magnitude, then for a highly inductive load, the average output current is greater than zero, and the chopper operates in the first-quadrant. If the load is not highly inductive the boundary is determined by the average output current $\bar{I}_o > 0$. The various circuit waveforms are shown in figure 14.13b.

iii. $\check{I} < 0$, $\hat{I} > 0$ and $\bar{I}_o < 0$
For a highly inductive load, if the magnitude of the negative peak is greater than the positive maximum, the average is less than zero and the chopper is operating in the regenerative mode, quadrant II. If the load is not highly inductive the boundary is determined by the average output current $\bar{I}_o < 0$.

iv. $\check{I} < 0$, $\hat{I} < 0$ and $\bar{I}_o < 0$
When the maximum current and the average current are both negative, the chopper is operational in the second-quadrant. Since the load current never goes positive, switch T_1 and diode D_1 never conduct, as shown in figure 14.13c.

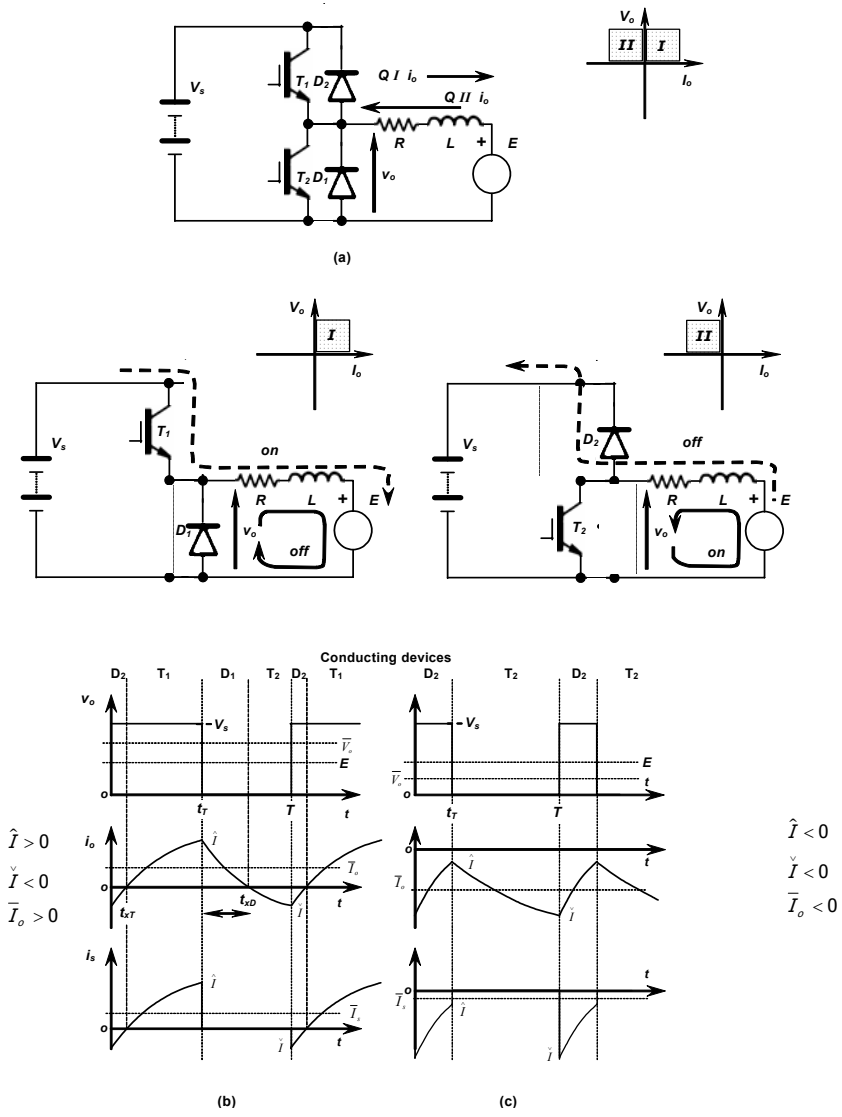


Figure 14.13. *Two-quadrant (I and II) dc chopper circuit where $v_o > 0$: (a) basic two-quadrant dc chopper; (b) operation and waveforms for quadrant I; and (c) operation and waveforms for quadrant II, regeneration into V_s.*

In all cases the average output voltage is solely determined by the switch T_1 on-time duty cycle, since when this switch is turned on the supply V_s is impressed across the load, independent of the direction of the load current. When $i_o > 0$, switch T_1 conducts while if $i_o < 0$, the diode in parallel to switch T_1, namely D_1 conducts, clamping the load to V_s.

The output voltage, which is independent of the load, is described by

$$v_o(t) = \begin{cases} V_s & \text{for } 0 \le t \le t_T \\ 0 & \text{for } t_T \le t \le T \end{cases} \qquad (14.80)$$

Thus

$$\bar{V}_o = \frac{1}{T}\int_0^{t_T} V_s dt = \frac{t_T}{T} V_s = \delta V_s \qquad (14.81)$$

The rms output voltage is also determined solely by the duty cycle,

$$V_{rms} = \left[\frac{1}{T}\int_0^{t_T} V_s^2\, dt\right]^{1/2} = \sqrt{\delta}\, V_s \qquad (14.82)$$

The output ac ripple voltage, hence voltage ripple factor are given by equations (14.3) and (14.5), and are independent of the load:

$$V_r = \sqrt{V_{rms}^2 - V_o^2} = V_s\sqrt{\delta(1-\delta)} \qquad (14.83)$$

and

$$RF = \frac{V_r}{\bar{V}_o} = \sqrt{\frac{1}{\delta} - 1} = \sqrt{\frac{1-\delta}{\delta}} \qquad FF = 1/\sqrt{\delta} \qquad (14.84)$$

The Fourier series for the load voltage can be used to determine the load current at each harmonic frequency as described by equations (14.6) to (14.10).
The time domain differential equations from section 14.2.1 are also valid, where there is no zero restriction on the minimum load current value.

In a **positive voltage loop**, when $v_o(t) = V_s$ and V_s is impressed across the load, the load circuit condition is described by

$$i_o(t) = \frac{V_s - E}{R}\left(1 - e^{\frac{-t}{\tau}}\right) + \check{I}\, e^{\frac{-t}{\tau}} \qquad \text{for} \quad 0 \le t \le t_T \qquad (14.85)$$

During the **switch off-period**, when $v_o = 0$, forming a zero voltage loop

$$i_o(t) = -\frac{E}{R}\left(1 - e^{\frac{-t}{\tau}}\right) + \hat{I}\, e^{\frac{-t}{\tau}} \qquad \text{for} \quad 0 \le t \le T - t_T \qquad (14.86)$$

where

$$\text{where} \quad \hat{I} = \frac{V_s}{R}\frac{1 - e^{\frac{-t_T}{\tau}}}{1 - e^{\frac{-T}{\tau}}} - \frac{E}{R} \quad \text{(A)}$$

$$\text{and} \quad \check{I} = \frac{V_s}{R}\frac{e^{\frac{t_T}{\tau}} - 1}{e^{\frac{T}{\tau}} - 1} - \frac{E}{R} \quad \text{(A)} \qquad (14.87)$$

The peak-to-peak ripple current is independent of E,

$$I_{p-p} = \frac{V_s}{R}\frac{(1 - e^{\frac{-\delta T}{\tau}})(1 - e^{\frac{-(1-\delta)T}{\tau}})}{1 - e^{\frac{-T}{\tau}}} \qquad (14.88)$$

The average output current, $\bar{I}_o$, may be positive or negative and is given by

$$\bar{I}_o = \frac{1}{T}\int_0^T i_o(t)\, dt = \left(\bar{V}_o - E\right)/R = \left(\delta V_s - E\right)/R \quad \text{(A)} \qquad (14.89)$$

The direction of the net power flow between E and V_s determines the chopper operating quadrant. If $\bar{V}_o > E$ then average power flow is to the load, as shown in figure 14.13b, while if $\bar{V}_o < E$, the average power flow is back into the source V_s, as shown in figure 14.13c.

$$V_s \bar{I}_s = \pm I_{o_{rms}}^2 R + E\bar{I}_o \qquad (14.90)$$

Thus the sign of $\bar{I}_o$ determines the direction of net power flow, hence quadrant of operation.

Calculation of individual device average currents in the time domain is complicated by the fact that the energy may flow between the dc source V_s and the load via the switch T_1 (energy to the load) or diode D_2 (energy from the load). It is therefore necessary to ascertain the zero current crossover time, when $\hat{I}$ and $\check{I}$ have opposite signs, which will then specify the necessary bounds of integration.

Equations (14.85) and (14.86) are equated to zero and solved for the time at zero crossover, t_{xT} and t_{xD}, respectively, shown in figure 14.13b.

$$t_{xT} = \tau\, \ell n\left(1 - \frac{\check{I}R}{V_s - E}\right) \quad \text{with respect to } t = 0$$

$$t_{xD} = \tau\, \ell n\left(1 + \frac{\hat{I}R}{E}\right) \quad \text{with respect to } t = t_T \qquad (14.91)$$

The necessary integration for each device can then be determined with the aid of the device conduction information in the parts of figure 14.13 and Table 14.1.

Table 14.1: Device average current ratings

Device and integration bounds, a to b	$\hat{I} > 0, \check{I} > 0$	$\hat{I} > 0, \check{I} < 0$	$\hat{I} < 0, \check{I} < 0$
$\bar{I}_{T1} = \frac{1}{T}\int_a^b \frac{V_s - E}{R}\left(1 - e^{\frac{-t}{\tau}}\right) + \check{I}\, e^{\frac{-t}{\tau}} dt$	0 to t_T	t_{xT} to t_T	0 to 0
$\bar{I}_{D1} = \frac{1}{T}\int_0^b \frac{V_s - E}{R}\left(1 - e^{\frac{-t}{\tau}}\right) + \check{I}\, e^{\frac{-t}{\tau}} dt$	0 to 0	0 to t_{xT}	0 to t_T
$\bar{I}_{T2} = \frac{1}{T}\int_a^b -\frac{E}{R}\left(1 - e^{\frac{-t}{\tau}}\right) + \hat{I}\, e^{\frac{-t}{\tau}} dt$	0 to 0	t_{xD} to $T - t_T$	0 to $T - t_T$
$\bar{I}_{D2} = \frac{1}{T}\int_0^b -\frac{E}{R}\left(1 - e^{\frac{-t}{\tau}}\right) + \hat{I}\, e^{\frac{-t}{\tau}} dt$	0 to $T - t_T$	0 to t_{xD}	0 to 0

The electromagnetic energy transfer efficiency is determined from

$$\eta = \frac{E\bar{I}_o}{V_s \bar{I}_i} \quad \text{for} \quad \bar{I}_o > 0$$

$$\eta = \frac{V_s \bar{I}_i}{E\bar{I}_o} \quad \text{for} \quad \bar{I}_o < 0 \qquad (14.92)$$

Example 14.5: ***Two-quadrant DC chopper with load back emf***

The two-quadrant dc-to-dc chopper in figure 14.13a feeds an inductive load of 10 Ω resistance, 50mH inductance, and back emf of 100V dc, from a 340V dc source. If the chopper is operated at 200Hz with a 25% on-state duty cycle, determine:

i. the load average and rms voltages;
ii. the rms ripple voltage, hence ripple and form factors;
iii. the maximum and minimum output current, hence peak-to-peak output ripple in the current;
iv. the current in the time domain;
v. the current crossover times, if applicable;
vi. the load average current, average switch current and average diode current for all devices;
vii. the input power, hence output power and rms output current;
viii. effective input impedance and electromagnetic efficiency; and
ix. sketch the circuit, load, and output voltage and current waveforms.

Subsequently determine the necessary change in

x. duty cycle δ to result in zero average output current and
xi. back emf E to result in zero average load current.

Solution

The main circuit and operating parameters are

- on-state duty cycle δ = ¼
- period $T = 1/f_s$ = 1/200Hz = 5ms
- on-period of the switch t_T = 1.25ms
- load time constant $\tau = L/R$ = 0.05mH/10Ω = 5ms

i. From equations (14.81) and (14.82) the load average and rms voltages are

$$v_o = \frac{t_T}{T} V_s = \frac{1.25\text{ms}}{5\text{ms}} \times 340\text{V} = ¼ \times 340\text{V} = 85\text{V}$$

$$V_{rms} = \sqrt{\delta}\, V_s = \sqrt{¼} \times 340\text{V} = 170\text{V} \quad \text{rms}$$

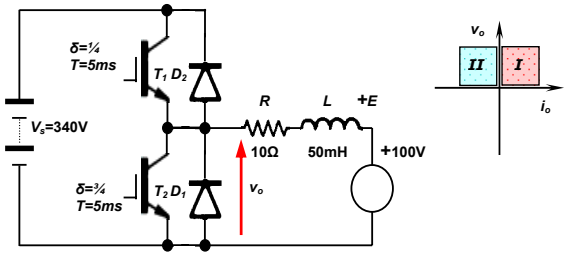


Figure 14.14. Example 14.5. *Circuit diagram.*

ii. The rms ripple voltage, hence voltage ripple factor, from equations (14.83) and (14.84) are

$$V_r = \sqrt{V_{rms}^2 - V_o^2} = V_s\sqrt{\delta(1-\delta)}$$

$$= \sqrt{170^2 - 85^2} = 340\text{V}\sqrt{¼\times(1 - ¼)} = 147.2\text{V}$$

$$RF = \frac{V_r}{\bar{V}_o} = \sqrt{\frac{1}{\delta} - 1} = \sqrt{\frac{1}{¼} - 1} = 1.732 \qquad FF = \frac{1}{\sqrt{\delta}} = \frac{1}{\sqrt{¼}} = 2$$

iii. From equations (14.87) and (14.88), the maximum and minimum output current, hence the peak-to-peak output ripple in the load current are given by

$$\hat{I} = \frac{V_s}{R}\frac{1-e^{\frac{-t_T}{\tau}}}{1-e^{\frac{-T}{\tau}}} - \frac{E}{R} = \frac{340\text{V}}{10\Omega}\times\frac{1 - e^{\frac{-1.25\text{ms}}{5\text{ms}}}}{1 - e^{\frac{-5\text{ms}}{5\text{ms}}}} - \frac{100\text{V}}{10\Omega} = 1.90\text{A}$$

$$\check{I} = \frac{V_s}{R}\frac{e^{\frac{t_T}{\tau}}-1}{e^{\frac{T}{\tau}}-1} - \frac{E}{R} = \frac{340\text{V}}{10\Omega}\times\frac{e^{\frac{1.25\text{ms}}{5\text{ms}}} - 1}{e^{\frac{5\text{ms}}{5\text{ms}}} - 1} - \frac{100\text{V}}{10\Omega} = -4.38\text{A}$$

The peak-to-peak ripple current is therefore $\Delta i_o = 1.90\text{A} - -4.38\text{A} = 6.28\text{A}$ p-p.

iv. The current in the time domain is given by equations (14.85) and (14.86)

$$i_o(t) = \frac{V_s - E}{R}\left(1-e^{\frac{-t}{\tau}}\right) + \check{I}e^{\frac{-t}{\tau}}$$

$$= \frac{340\text{V-}100\text{V}}{10\Omega}\times\left(1-e^{\frac{-t}{5\text{ms}}}\right) - 4.38\times e^{\frac{-t}{5\text{ms}}}$$

$$= 24\times\left(1-e^{\frac{-t}{5\text{ms}}}\right) - 4.38\times e^{\frac{-t}{5\text{ms}}}$$

$$= 24 - 28.38\times e^{\frac{-t}{5\text{ms}}} \qquad \text{for} \quad 0 \le t \le 1.25\text{ms}$$

$$i_o(t) = -\frac{E}{R}\left(1-e^{\frac{-t}{\tau}}\right) + \hat{I}e^{\frac{-t}{\tau}}$$

$$= -\frac{100V}{10\Omega}\times\left(1-e^{\frac{-t}{5\text{ms}}}\right) + 1.90\times e^{\frac{-t}{5\text{ms}}}$$

$$= -10\times\left(1-e^{\frac{-t}{5\text{ms}}}\right) + 1.90\times e^{\frac{-t}{5\text{ms}}}$$

$$= -10 + 11.90\times e^{\frac{-t}{5\text{ms}}} \qquad \text{for} \quad 0 \le t \le 3.75\text{ms}$$

v. Since the maximum current is greater than zero (1.9A) and the minimum is less that zero (- 4.38A), the current crosses zero during the switch on-time and off-time. The time domain equations for the load current are solved for zero to give the cross over times t_{xT} and t_{xD}, as given by equation (14.91), or solved from the time domain output current equations as follows.

During the switch on-time

$$i_o(t) = 24 - 28.38\times e^{\frac{-t}{5\text{ms}}} = 0 \qquad \text{where} \quad 0 \le t = t_{xT} \le 1.25\text{ms}$$

$$t_{xT} = 5\text{ms}\times\ell n\frac{28.38}{24} = 0.838\text{ms}$$

During the switch off-time

$$i_o(t) = -10 + 11.90\times e^{\frac{-t}{5\text{ms}}} = 0 \qquad \text{where} \quad 0 \le t = t_{xD} \le 3.75\text{ms}$$

$$t_{xD} = 5\text{ms}\times\ell n\frac{11.90}{10} = 0.870\text{ms}$$

(1.250ms + 0.870ms = 2.12ms with respect to switch T_1 turn-on)

vi. The load average current, average switch current, and average diode current for all devices;

$$\bar{I}_o = \frac{(\bar{V}_o - E)}{R} = \frac{(\delta V_s - E)}{R}$$

$$\frac{(85\text{V} - 100\text{V})}{10\Omega} = -1.5\text{A}$$

When the output current crosses zero current, the conducting device changes. Table 14.1 gives the necessary current equations and integration bounds for the condition $\hat{I} > 0, \check{I} < 0$. Table 14.1 shows that all four semiconductors are involved in the output current cycle.

$$\bar{I}_{T1} = \frac{1}{T}\int_{t_{xT}}^{t_T}\frac{V_s - E}{R}\left(1-e^{\frac{-t}{\tau}}\right) + \check{I}e^{\frac{-t}{\tau}}dt$$

$$= \frac{1}{5\text{ms}}\int_{0.838\text{ms}}^{1.25\text{ms}} 24 - 28.38\times e^{\frac{-t}{5\text{ms}}}dt = 0.081\text{A}$$

$$\bar{I}_{D1} = \frac{1}{T}\int_{0}^{t_{xT}}\frac{V_s - E}{R}\left(1-e^{\frac{-t}{\tau}}\right) + \check{I}e^{\frac{-t}{\tau}}dt$$

$$= \frac{1}{5\text{ms}}\int_{0}^{0.84\text{ms}} 24 - 28.38\times e^{\frac{-t}{5\text{ms}}}dt = -0.357\text{A}$$

$$\bar{I}_{T2} = \frac{1}{T}\int_{t_{xD}}^{T-t_T} -\frac{E}{R}\left(1-e^{\frac{-t}{\tau}}\right) + \hat{I}e^{\frac{-t}{\tau}}dt$$

$$= \frac{1}{5\text{ms}}\int_{0870\text{ms}}^{3.75\text{ms}} -10 + 11.90\times e^{\frac{-t}{5\text{ms}}}dt = -1.382\text{A}$$

$$\bar{I}_{D2} = \frac{1}{T}\int_{0}^{t_{xD}} -\frac{E}{R}\left(1-e^{\frac{-t}{\tau}}\right) + \hat{I}e^{\frac{-t}{\tau}}dt$$

$$= \frac{1}{5\text{ms}}\int_{0}^{0.870ms} -10 + 11.90\times e^{\frac{-t}{5\text{ms}}}dt = 0.160\text{A}$$

Check $\bar{I}_o + \bar{I}_{T1} + \bar{I}_{D1} + \bar{I}_{T2} + \bar{I}_{D2} = -1.5\text{A} + 0.080\text{A} - 0.357\text{A} - 1.382\text{A} + 0.160\text{A} = 0$

vii. The input power, hence output power and rms output current;

$$P_{in} = P_{V_s} = V_s\bar{I}_i = V_s\left(\bar{I}_{T1} + \bar{I}_{D1}\right)$$

$$= 340\text{V}\times(0.080\text{A} - 0.357\text{A}) = -95.2\text{W}, \text{ (charging } V_s\text{)}$$

$$P_{out} = P_E = E\bar{I}_o = 100\text{V}\times(-1.5\text{A}) = -150\text{W}, \text{ that is generating } 150\text{W}$$

From

$$V_s I_s = I_{o_{rms}}^2 R + E\bar{I}_o$$

$$I_{o_{rms}} = \sqrt{\frac{P_{out} - P_{in}}{R}} = \sqrt{\frac{150\text{W} - 92.5\text{W}}{10\Omega}} = 2.34\text{A rms}$$

viii. Since the average output current is negative, energy is being transferred from the back emf *E* to the dc voltage source V_s, the electromagnetic efficiency of conversion is given by

$$\eta = \frac{V_s\bar{I}_i}{E\bar{I}_o} \text{ for } \bar{I}_o < 0$$

$$= \frac{95.2\text{W}}{150\text{W}} = 63.5\%$$

The effective input impedance is

$$Z_{in} = \frac{V_s}{\overline{I}_i} = \frac{V_s}{\overline{I}_{T1} + \overline{I}_{D1}} = \frac{340\text{V}}{0.080\text{A} - 0.357\text{A}} = -1214\Omega$$

ix. The circuit, load, and output voltage and current waveforms are sketched in the figure 14.15.

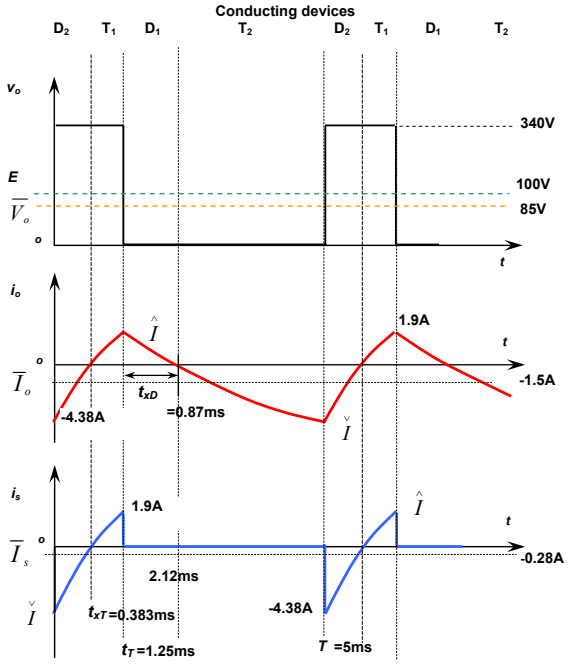


Figure 14.15. Example 14.5. *Circuit waveforms.*

x. Duty cycle δ to result in zero average output current can be determined from the expression for the average output current, equation (14.89), that is

$$\overline{I}_o = \frac{\delta V_s - E}{R} = 0$$

that is

$$\delta = \frac{E}{V_s} = \frac{100\text{V}}{340\text{V}} = 29.4\%$$

xi. As in part x, the average load current equation can be rearranged to give the back emf *E* that results in zero average load current

$$\overline{I}_o = \frac{\delta V_s - E}{R} = 0$$

that is

$$E = \delta V_s = ¼ \times 340\text{V} = 85\text{V}$$

♣

14.5 Two-quadrant dc chopper - Q 1 and Q IV

The unidirectional current, two-quadrant dc chopper, or asymmetrical half H-bridge shown in figure 14.16a incorporates two switches T_1 and T_4 and two complementary diodes D_1 and D_4. In using switches T_1 and T_4 the chopper operates in the first and fourth quadrants, that is, bi-directional voltage output v_o but unidirectional load current, i_o.
The chopper can operate in two quadrants (I and IV), depending on the load and switching sequence. Net power can be delivered to the load, or received from the load provided the polarity of the back emf *E* is reversed. Because of this need to reverse the back emf for regeneration, this chopper is not commonly used in dc machine control. On the other hand, the chopper circuit configuration is commonly used to meet the converter requirements of the switched reluctance machine, which only requires unipolar current to operate. Also see chapter 17.5 for a bidirectional smps variation.
The asymmetrical half H-bridge chopper has three different output voltage states, where one state (the zero output state) has redundancy (two possibilities). Both the output voltage v_o and output current i_o are with reference to the first quadrant arrows in figure 14.16a.

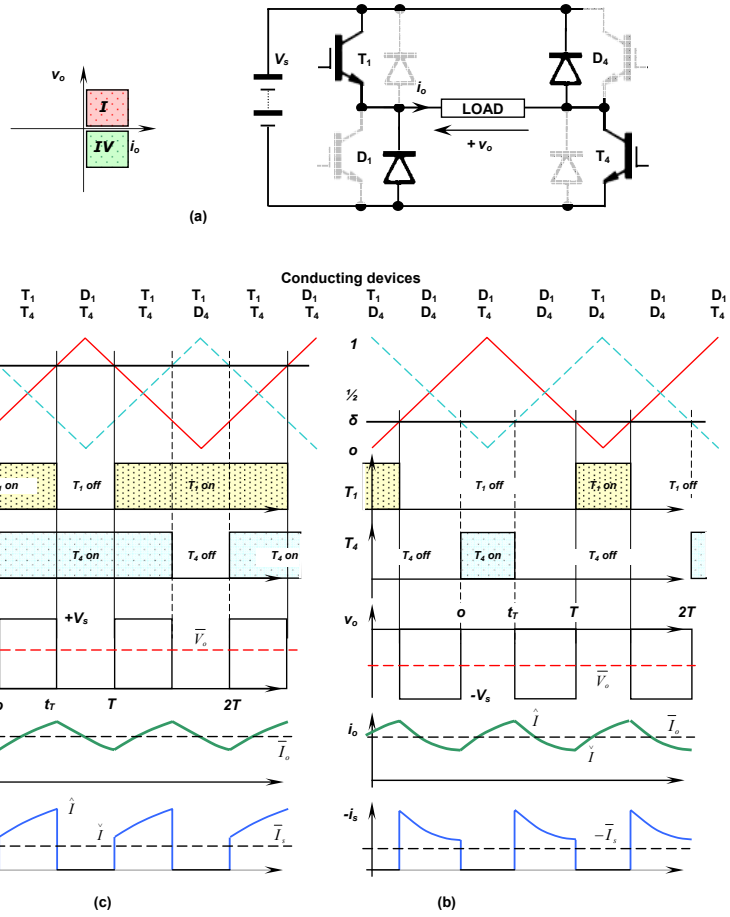


Figure 14.16. *Two-quadrant (I and IV) dc chopper*
(a) circuit where i_o>0: (b) operation in quadrant IV, regeneration into V_s; and (c) operation in quadrant I.

State #1
When both switches T_1 and T_4 conduct, the supply V_s is impressed across the load, as shown in figure 14.17a. Energy is drawn from the dc source V_s.

T_1 and T_4 conducting: $v_o = V_s$

State #2
If only one switch is conducting, and therefore also one diode, the output voltage is zero, as shown in figure 14.17b. Either switch (but only one on at any time) can be the on-switch, hence providing redundancy, that is

T_1 and D_4 conducting: $v_o = 0$
T_4 and D_1 conducting: $v_o = 0$

State #3
When both switches are off, the diodes D_1 and D_4 conduct load energy back into the dc source V_s, as in figure 14.17c. The output voltage is $-V_s$, that is

T_1 and T_4 are not conducting: $v_o = -V_s$

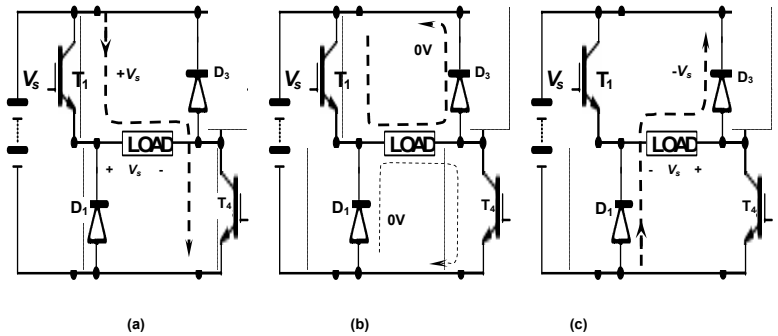


Figure 14.17. *Two-quadrant (I and IV) dc chopper operational current paths: (a) T_1 and T_4 forming a $+V_s$ path; (b) T_1 and D_4 (or T_4 and D_1) forming a zero voltage loop; and (c) D_1 and D_4 creating a $-V_s$ path.*

The two zero output voltage states can most effectively be used if alternated during any switching sequence. In this way, the load switching frequency (load ripple current frequency) is twice the switching frequency of the switches. This reduces the output current ripple for a given switch operating frequency (which minimises the load inductance necessary for continuous load current conduction). Also, by alternating the zero voltage loop, the semiconductor losses are evenly distributed. Specifically, a typical sequence to achieve these features would be

T_1 and T_4	V_s	
T_1 and D_4	0	
T_1 and T_4	V_s	
T_4 and D_1	0	(not T_1 and D_4 again)
T_1 and T_4	V_s	
T_1 and D_4	0, etc.	

The sequence can also be interleaved in the regeneration mode, when only one switch is on at any instant, as follows

D_1 and D_4	$-V_s$	(that is T_1 and T_4 off)
T_1 and D_4	0	
D_1 and D_4	$-V_s$	
T_4 and D_1	0	(not T_1 and D_4 again)
D_1 and D_4	$-V_s$	
T_1 and D_4	0, etc.	

In switched reluctance motor drive application there may be no alternative to using only $\pm V_s$ control loops without the intermediate zero voltage state.

There are two basic modes of chopper switching operation.

- *Multilevel switching* is when both switches are controlled independently to give all three output voltage states (three levels), namely $\pm V_s$ and 0V.

- *Bipolar switching (*or *two level switching)* is when both switches operate in unison, where they turn on together and off together. Only two voltage output states (hence the term bipolar), are possible, $+V_s$ and $-V_s$.

14.5.1 dc chopper:– Q I and Q IV – multilevel output voltage switching (three level)

The interleaved zero voltage states are readily introduced if the control carrier waveforms for the two switches are displaced by 180°, as shown in figure 14.16b and c, for continuous load current. This requirement can be realised if two up-down counters are displaced by 180°, when generating the necessary triangular carriers. As shown in figures 14.16b and c, the switching frequency $1/T_s$ is determined by the triangular wave frequency $1/2T$, whilst advantageously the load experiences twice that frequency, $1/T$, hence the output current has reduced ripple, for a given switch operating frequency.

i. **$0 \le \delta \le ½$**
It can be seen in figure 14.16b that when $\delta \le ½$ both switches never conduct simultaneously hence the output voltage is either 0 or $-V_s$. Operation is in the fourth quadrant. The average output voltage is load independent and for $0 \le \delta \le ½$, using the waveforms in figure 14.16b, is given by

$$\overline{V}_o = \frac{1}{T}\int_{t_T}^{T} -V_s\, dt = \frac{-V_s}{T}\left(T - t_T\right) = -V_s\left(1 - \frac{t_T}{T}\right) \qquad (14.93)$$

Examination of figure 14.16b reveals that the relationship between t_T and δ must produce

when $\delta = 0$: $t_T = T$ and $v_o = -V_s$
when $\delta = ½$: $t_T = 0$ and $v_o = 0$

that is

$$\delta = ½\frac{t_T}{T}$$

(the period of the carrier, $2T$, is twice the switching period, T) which after substituting for t_T/T in equation (14.93) gives

$$\overline{V}_o = -V_s\left(1 - \frac{t_T}{T}\right) \qquad (14.94)$$
$$= -V_s\left(1 - 2\delta\right) = V_s\left(2\delta - 1\right) \quad \text{for} \quad 0 \le \delta \le ½$$

Operational analysis in the fourth quadrant, $\delta \le ½$, is similar to the analysis for the second-quadrant chopper in figure 14.2b and analysed in section 14.3. Operation is characterised by first shorting the output circuit to boost the current, then removing the output short forces current back into the dc supply V_s, via a freewheel diode. The characteristics of this mode of operation are described by the equations (14.50) to (14.79) for the second-quadrant chopper analysed in 14.3, where the output current may again be continuous or discontinuous. The current and voltage references are both reversed in translating equations applicable in quadrants Q II to Q IV.

ii. **$½ \le \delta \le 1$**
As shown in figure 14.16c, when $\delta \ge ½$ and operation is in the first quadrant, at least one switch is conducting hence the output voltage is either $+V_s$ or 0. For continuous load current, the average output voltage is load independent and for $½ \le \delta \le 1$ is given by

$$\overline{V}_o = \frac{1}{T}\int_0^{t_T} V_s dt = \frac{V_s}{T} t_T \qquad (14.95)$$

Examination of figure 14.16c reveals that the relationship between t_T and δ must produce

when $\delta = ½$: $t_T = 0$ and $v_o = 0$
when $\delta = 1$: $t_T = T$ and $v_o = V_s$

that is

$$\delta = ½\left(\frac{t_T}{T} + 1\right)$$

which on substituting for t_T/T in equation (14.95) gives

$$\overline{V}_o = V_s\frac{t_T}{T} = V_s\left(2\delta - 1\right) \quad \text{for} \quad ½ \le \delta \le 1 \qquad (14.96)$$

Since the average output voltage is the same in each case, equations (14.94) and (14.96) for ($0 \le \delta \le 1$), the output current mean is given by the same expression, namely

$$\overline{I}_o = \frac{\overline{V}_o - E}{R} = \frac{V_s(2\delta - 1) - E}{R} \qquad (14.97)$$

Operation in the first quadrant, $\delta \geq ½$, is characterised by the first-quadrant chopper shown in figure 14.2a and considered in section 14.2 along with the equations within that section. The load current can be either continuous, in which case equations (14.6) to (14.23) are valid; or discontinuous in which case equations (14.24) to (14.45) are applicable. Aspects of this mode of switching are extended in 14.5.3.
In applying the equations for the chopper in section 14.2 for the first-quadrant chopper, and the equations in section 14.3 for the second-quadrant chopper, the duty cycle in each case is replaced by

- 2δ -1 in the case of $\delta \geq ½$ for the first-quadrant chopper and
- 2δ in the case of $\delta \leq ½$ for the fourth-quadrant chopper.

This will account for the scaling and offset produced by the triangular carrier signal decoding.

14.5.2 dc chopper: – Q I and Q IV – bipolar voltage switching (two level)

When both switches operate in the same state, that is, both switches are on simultaneously or both are off together, operation is termed bipolar or two level switching.
From figure 14.18 the chopper output states are (assuming continuous load current)

- T_1 and T_4 on $v_o = V_s$
- T_1 and T_4 off $v_o = - V_s$

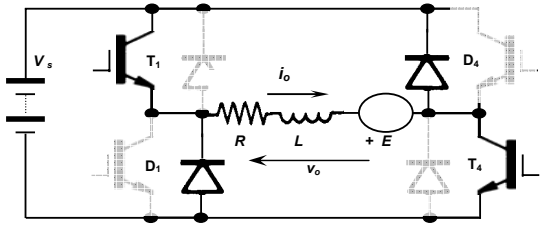
From figure 14.18, the average output voltage is

$$\begin{aligned}\overline{V}_o &= \frac{1}{T}\left(\int_0^{t_T} V_s\,dt + \int_{t_T}^{T} -V_s\,dt\right) \\ &= \frac{V_s}{T}\left(t_T - T + t_T\right) = \left(2\delta - 1\right)V_s\end{aligned} \qquad (14.98)$$

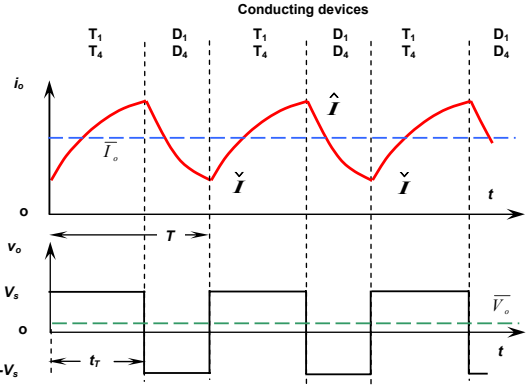
The rms output voltage is independent of the duty cycle and is V_s.
The output ac ripple voltage is

$$\begin{aligned}V_r &= \sqrt{V_{rms}^2 - V_o^2} \\ &= \sqrt{V_s^2 - \left(2\delta - 1\right)^2 V_s^2} = 2V_s\sqrt{\delta\left(1-\delta\right)}\end{aligned} \qquad (14.99)$$

which is a maxima at $\delta = ½$ and a minima for $\delta = 0$ and $\delta = 1$.



(a)



(b)

Figure 14.18. *Two-quadrant (I and IV) dc chopper operation in the bipolar output mode: (a) circuit showing load components and (b) chopper output waveforms.*

The output voltage ripple factor is

$$RF = \frac{V_r}{V_o} = \frac{2V_s\sqrt{\delta\left(1-\delta\right)}}{\left(2\delta-1\right)V_s} = \frac{2\sqrt{\delta\left(1-\delta\right)}}{\left(2\delta-1\right)} \qquad (14.100)$$

Although the average output voltage may reverse, the load current is always positive but can be discontinuous or continuous. Equations describing bipolar output are presented within the next section, 14.5.3, which considers multilevel (two and three level) output voltage switching states.

14.5.3 Multilevel output voltage states, dc chopper

In switched reluctance machine drives it is not uncommon to operate the asymmetrical half H-bridge shown in figure 14.18 such that

- both switches operate in the on-state together to form $+V$ voltage loops;
- switches operate independently to give zero voltage loops; and
- both switches are simultaneously off, forming $–V$ voltage output loops.

The control objective is to generate a current output pulse that tracks a reference shape which starts from zero, rises to maintain a fixed current level, with hysteresis, then the current falls back to zero. The waveform shown in figure 14.19 fulfils this specification.
The switching strategy to produce the current waveform in figure 14.19 aims at:

- For rising current:- use $+V$ loops (and zero volt loops only if necessary)
- For near constant current:- use zero voltage loops (and $\pm V$ loops only if necessary to increase or decrease the current)
- For falling current:- use $– V$ loops (and zero volts loops only if necessary to reduce the fall rate)

Operation is further characterised by continuous load current during the pulse.
Energy is supplied to the load from the dc voltage source during $+V$ loops, and returned to the dc supply during $–V$ loop periods.

The chopper output current during each period is described by equations previously derived in this chapter, but reproduced as follows.

In a **positive voltage loop**, (T_1 and T_4 are both on), when $v_o(t) = V_s$ and V_s is impressed across the load, the load circuit condition is described by

$$L\frac{di_o}{dt} + Ri_o + E = V_s$$

which yields

$$i_o\left(t\right) = \frac{V_s - E}{R}\left(1 - e^{\frac{-t}{\tau}}\right) + \check{I}\,e^{\frac{-t}{\tau}} \qquad \text{for} \quad 0 \leq t \leq t^+ \qquad (14.101)$$

During the first switching cycle the current starts from zero, so $\check{I} = 0$. Otherwise $\check{I}$ is the lower reference, I^-, from the end of the previous cycle.

The current at the end of the positive voltage loop period is the reference level I^+, whilst the time to rise to I^+ is derived by equating equation (14.101) to I^+ and solving for time t^+ at the end of the period. Solving $i_o(t^+) = I^+$ for t^+, gives

$$t^+ = \tau \ln\left(\frac{V_s - E - \check{I}R}{V_s - E - I^+R}\right) \qquad (14.102)$$

In a **zero voltage loop**, when $v_o(t) = 0$, such as circuit loops involving T_1 and D_4 (or T_4 and D_1), the circuit equation is given by

$$L\frac{di_o}{dt} + Ri_o + E = 0$$

which gives

$$i_o\left(t\right) = -\frac{E}{R}\left(1 - e^{\frac{-t}{\tau}}\right) + \hat{I}\,e^{\frac{-t}{\tau}} \qquad \text{for} \quad 0 \leq t \leq t^o \qquad (14.103)$$

where $\hat{I}$ equals the reference current level, I^+ from the previous switching period.
The current at the end of the period is the reference level I^-, whilst the time to fall to I^- is given by equating equation (14.103) to I^- and solving for time, t^o at the end of the period.

$$t^o = \tau\,\ell n\left(\frac{E + \hat{I}R}{E + I^-R}\right) \qquad (14.104)$$

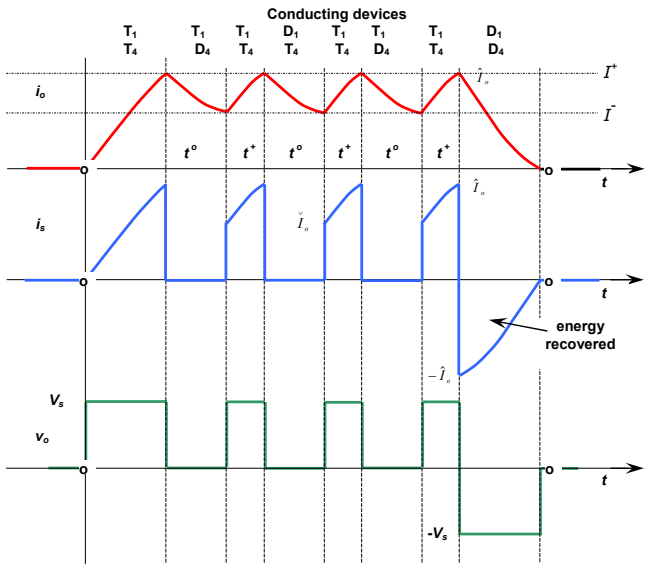


Figure 14.19. *Two-quadrant (I and IV) dc chopper operation in a multilevel output voltage mode.*

In a **negative voltage loop**, when both switches T_1 and T_4 are off, the current falls rapidly and the circuit equation, when $v_o(t) = -V_s$, is

$$L\frac{di_o}{dt} + Ri_o + E = -V_s$$

which gives

$$i_o(t) = \frac{-E - V_s}{R}\left(1 - e^{\frac{-t}{\tau}}\right) + \hat{I}e^{\frac{-t}{\tau}} \qquad \text{for} \quad 0 \le t \le t^- \qquad (14.105)$$

where $\hat{I}$ equals the reference current level, I^+ from the previous switching period.
The current at the end of the period is I^-, whilst the time to reach I^- is given by equating equation (14.103) to I^- and solving for time t^- at the end of the period.

$$t^- = \tau \ell n\left(\frac{V_s + E + \hat{I}R}{V_s + E + I^-R}\right) \qquad (14.106)$$

The same equation is used to determine the time for the final current period when the current decays to zero, whence $I^- = 0$.

The characteristics and features of the three output voltage states are illustrated in the following example, 14.6.

Example 14.6: ***Asymmetrical, half H-bridge, dc chopper***

The asymmetrical half H-bridge, dc-to-dc chopper in figure 14.18 feeds an inductive load of 10 Ω resistance, 50mH inductance, and back emf of 55V dc, from a 340V dc voltage source. The chopper output current is controlled in a hysteresis mode within a current band between limits 5A and 10A. Determine the period of the current shape shown in the figure 14.20:

i. when only $\pm V_s$ loops are used and
ii. when a zero volt loop is used to maintain tracking within the 5A band.

In each case calculate the switching frequency if the current were to be maintained within the hysteresis band for a prolonged period.

How do the on-state losses compare between the two control approaches?

Solution

The main circuit and operating parameters are

- E = 55V and V_s = 340V
- load time constant $\tau = L/R$ = 0.05mH/10Ω = 5ms
- I^+ = 10A and I^- = 5A

Examination of the figure 14.20 shows that only one period of the cycle differs, namely the second period, t_2, where the current is required to fall to the lower hysteresis band level, -5A. The period of the other three regions (t_1, t_3, and t_4) are common and independent of the period of the second region, t_2.

t_1: The first period, the initial rise time, $t^+ = t_1$ is given by equation (14.102), where I^+=10A and $\check{I} = 0$A.

$$t^+ = \tau \ell n\left(\frac{V_s - E - \check{I}R}{V_s - E - I^+R}\right)$$

$$\text{that is} \quad t_1 = 5\text{ms} \times \ell n\left(\frac{340\text{V} - 55\text{V} - 0\text{A}\times 10\Omega}{340\text{V} - 55\text{V} - 10\text{A}\times 10\Omega}\right) = 2.16\text{ms}$$

t_3: In the third period, the current rises from the lower hysteresis band limit of 5A to the upper band limit 10A. The duration of the current increase is given by equation (14.102) again, but with $\check{I} = I^- = 5$A.

$$t^+ = \tau \ell n\left(\frac{V_s - E - \check{I}R}{V_s - E - I^+R}\right)$$

$$\text{that is} \quad t_3 = 5\text{ms} \times \ell n\left(\frac{340\text{V} - 55\text{V} - 5\text{A}\times 10\Omega}{340\text{V} - 55\text{V} - 10\text{A}\times 10\Omega}\right) = 1.20\text{ms}$$

t_4: The fourth and final period is a negative voltage loop where the current falls from the upper band limit of 10A to I^- which equals zero. From equation (14.106) with $\hat{I} = I^+$=10A and I^- = 0A

$$t^- = \tau \ell n\left(\frac{V_s + E + \hat{I}R}{V_s + E + I^-R}\right)$$

$$\text{that is} \quad t_4 = 5\text{ms} \times \ell n\left(\frac{340\text{V} + 55\text{V} + 10\text{A}\times 10\Omega}{340\text{V} + 55\text{V} + 0\text{A}\times 10\Omega}\right) = 1.13\text{ms}$$

The current pulse period is given by

$$T_p = t_1 + t_2 + t_3 + t_4$$
$$= 2.16\text{ms} + t_2 + 1.20\text{ms} + 1.13\text{ms}$$
$$= 4.49\text{ms} + t_2$$

i. **t_2:** When only -V_s paths are used to decrease the current, the time t_2 is given by equation (14.106), with I^-=5A and $\hat{I}$ =10A,

$$t^- = \tau \ell n\left(\frac{V_s + E + \hat{I}R}{V_s + E + I^-R}\right)$$

$$\text{that is} \quad t_2 = 5\text{ms} \times \ell n\left(\frac{340\text{V} + 55\text{V} + 10\text{A}\times 10\Omega}{340\text{V} + 55\text{V} + 5\text{A}\times 10\Omega}\right) = 0.53\text{ms}$$

The total period, T_p, of the chopped current pulse when a 0V loop is not used, is

$$T_p = t_1 + t_2 + t_3 + t_4$$
$$= 2.16\text{ms} + 0.53\text{ms} + 1.20\text{ms} + 1.13\text{ms} = 5.02\text{ms}$$

ii. **t_2:** When a zero voltage loop is used to maintain the current within the hysteresis band, the current decays slowly, and the period time t_2 is given by equation (14.104), with I^- = 5A and $\hat{I}$ =10A,

$$t^o = \tau \ell n\left(\frac{E + \hat{I}R}{E + I^-R}\right)$$

$$\text{that is} \quad t_2 = 5\text{ms} \times \ell n\left(\frac{55\text{V} + 10\text{A}\times 10\Omega}{55\text{V} + 5\text{A}\times 10\Omega}\right) = 1.95\text{ms}$$

The total period, T_p, of the chopped current pulse when a 0V loop is used, is

$$T_p = t_1 + t_2 + t_3 + t_4$$
$$= 2.16\text{ms} + 1.95\text{ms} + 1.20\text{ms} + 1.13\text{ms} = 6.44\text{ms}$$

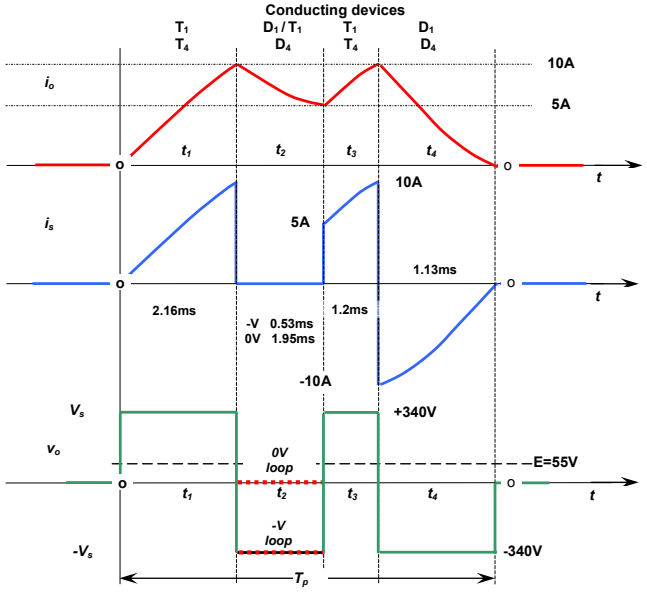


Figure 14.20. Example 14.6. *Circuit waveforms.*

The current falls significantly faster within the hysteresis band if negative voltage loops are employed rather that zero voltage loops, 0.53ms versus 1.95ms.
The switching frequency within the current bounds has a period $t_2 + t_3$, and each case is summarized in the following table. For longer current chopping, t_2 and t_3 dominate the switching frequency.
Using zero voltage current loops (alternated) reduces the switching frequency of the H-bridge switches by a factor of over three, for a given peak-to-peak ripple current.
If the on-state voltage drop of the switches and the diodes are similar for the same current level, then the on-state losses are similar, and evenly distributed for both control methods. The on-state losses are similar because each of the three states always involves the same current variation flowing through two semiconductors. The principal difference is in the significant increase in switching losses when only $\pm V$ loops are used (1:3.42).

Table Example 14.6. Switching losses.

Voltage loops	$t_2 + t_3$	Current ripple frequency	Switch frequency	Switch loss ratio
±V	0.53ms+1.20ms =1.73ms	578Hz	578Hz	$\frac{578}{169} = 3.42$
+V and zero	1.95ms+1.20ms = 3.15ms	317Hz	169Hz	1

♣

14.6 Four-quadrant dc chopper

The four-quadrant H-bridge dc chopper is shown in figure 14.21 where the load current and voltage are referenced with respect to T_1, so that the quadrant of operation with respect to the switch number is persevered.
The H-bridge is a flexible basic configuration where its use to produce single-phase ac is considered in chapter 15.1.1, while its use in smps applications is considered in chapter 17.8.2. It can also be used as a dc chopper for the four-quadrant control of a dc machine.

With the flexibility of four switches, a number of different control methods can be used to produce four-quadrant output voltage and current (bidirectional voltage and current). All practical methods should employ complementary device switching in each leg (either T_1 or T_4 on but not both and either T_2 or T_3 on, but not both) so as to minimise distortion by ensuring current continuity around zero current output.
One control method involves controlling the H-bridge as two virtually independent two-quadrant choppers, with the over-riding restriction that no two switches in the same leg conduct simultaneously. One chopper is formed with T_1 and T_4 grouped with D_1 and D_4, which gives positive current i_o but bidirectional voltage $\pm v_o$ (QI and QIV operation). The second chopper is formed by grouping T_2 and T_3 with D_2 and D_3, which gives negative output current $-i_o$, but bi-direction voltage $\pm v_o$ (QII and QIII operation).

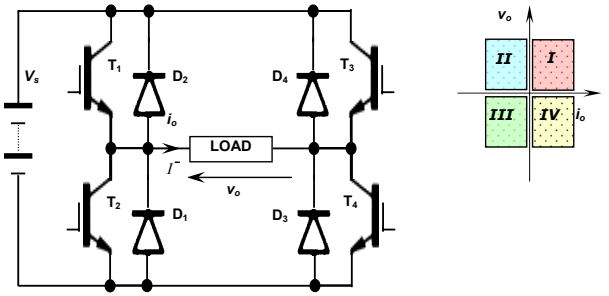


Figure 14.21. *Four-quadrant dc chopper circuit, showing first quadrant i_o and v_o references.*

The second control method is to unify the operation of all four switches within a generalised control algorithm.
With both control methods, the chopper output voltage can be either multilevel or bipolar, depending on whether zero output voltage loops are employed or not. Bipolar output states increase the ripple current magnitude, but do facilitate faster current reversal, without crossover distortion. Operation is independent of the direction of the output current i_o.

Since the output voltage is reversible for each control method, a triangular based modulation control method, as used with the asymmetrical H-bridge dc chopper in figure 14.16, is applicable in each case. Two generalised unified H-bridge control approaches are considered – bipolar and three-level output.

14.6.1 *Unified four-quadrant dc chopper - bipolar voltage output switching*

The simpler output to generate is bipolar output voltages, which use one reference carrier triangle as shown in figure 14.22 parts (c) and (d). The output voltage switches between + V_s and – V_s and the relative duration of each state depends on the magnitude of the modulation index δ.

If δ = 0 then T_1 and T_4 never turn-on since T_2 and T_3 conduct continuously which impresses – V_s across the load.
At the other extreme, **if δ = 1** then T_1 and T_4 are on continuously and $+V_s$ is impressed across the load.
If δ = ½ then T_1 and T_4 are turned on for half of the period T, while T_2 and T_3 are on for the remaining half of the period. The output voltage is – V_s for half of the time and + V_s for the remaining half of any period. The average output voltage is therefore zero, but disadvantageously, the output current needlessly ripples about zero (with an average value of zero).

The chopper output voltage is defined in terms of the triangle voltage reference level v_Δ by

- $v_\Delta > \delta$, $v_o = -V_s$
- $v_\Delta < \delta$, $v_o = +V_s$

From figure 14.22c and d, the average output voltage varies linearly with δ such that

$$\overline{V}_o = \frac{1}{T}\left(\int_0^{t_T} +V_s\,dt + \int_{t_T}^{T} -V_s\,dt\right) = \frac{1}{T}(2t_T - T)V_s = \left(2\frac{t_T}{T} - 1\right)V_s \qquad (14.107)$$

Examination of figures 14.22c and d reveals that the relationship between t_T and δ must produce

when $\delta = 0$: $t_T = 0$ and $v_o = -V_s$

when $\delta = ½$: $t_T = ½T$ and $v_o = 0$

when $\delta = 1$: $t_T = T$ and $v_o = +V_s$

that is

$$\delta = \frac{t_T}{T}$$

which on substituting for t_T/T in equation (14.107) gives

$$\begin{aligned} \overline{V}_o &= \left(2\frac{t_T}{T} - 1\right) V_s \\ &= (2\delta - 1)\, V_s \quad \text{for} \quad 0 \le \delta \le 1 \end{aligned} \tag{14.108}$$

The average output voltage can be positive or negative, depending solely on δ. No current discontinuity occurs since the output voltage is never actually zero. Even when the average voltage is zero, ripple current flows though the load, with an average value of zero amps.

The rms output voltage is independent of the duty cycle and is V_s.
The output ac ripple voltage is

$$\begin{aligned} V_r &= \sqrt{V_{rms}^2 - V_o^2} \\ &= \sqrt{V_s^2 - (2\delta - 1)^2 V_s^2} = 2V_s\sqrt{\delta(1-\delta)} \end{aligned} \tag{14.109}$$

The ac ripple voltage is zero at $\delta = 0$ and $\delta = 1$, when the output voltage is pure dc, namely - V_s or V_s, respectively. The maximum ripple voltage occurs at $\delta = ½$, when $V_r = V_s$.
The output voltage ripple factor is

$$\begin{aligned} RF &= \frac{V_r}{V_o} = \frac{2V_s\sqrt{\delta(1-\delta)}}{(2\delta-1)V_s} \\ &= \frac{2\sqrt{\delta(1-\delta)}}{(2\delta-1)} = \sqrt{\left(\frac{1}{2\delta-1}\right)^2 - 1} = \sqrt{FF^2 - 1} \end{aligned} \tag{14.110}$$

Circuit operation is characterized by two time domain equations:

During the **on-period for T1 and T4**, when $v_o(t) = V_s$

$$L\frac{di_o}{dt} + Ri_o + E = V_s$$

which yields

$$i_o(t) = \frac{V_s - E}{R}\left(1 - e^{\frac{-t}{\tau}}\right) + \check{I}e^{\frac{-t}{\tau}} \qquad \text{for} \quad 0 \le t \le t_T \tag{14.111}$$

During the **on-period for T2 and T3**, when $v_o(t) = -V_s$

$$L\frac{di_o}{dt} + Ri_o + E = -V_s$$

which, after shifting the zero time reference to t_T, gives

$$i_o(t) = -\frac{V_s + E}{R}\left(1 - e^{\frac{-t}{\tau}}\right) + \hat{I}e^{\frac{-t}{\tau}} \qquad \text{for} \quad 0 \le t \le T - t_T \tag{14.112}$$

The initial conditions $\hat{I}$ and $\check{I}$ are determined by using the steady-state boundary conditions:

$$\begin{aligned} \text{where} \quad \hat{I} &= \frac{V_s}{R}\frac{1 - 2e^{\frac{-t_T}{\tau}} + e^{\frac{-T}{\tau}}}{1 - e^{\frac{-T}{\tau}}} - \frac{E}{R} \qquad \text{(A)} \\ \text{and} \quad \check{I} &= \frac{V_s}{R}\frac{2e^{\frac{t_T}{\tau}} - 1 + e^{\frac{-T}{\tau}}}{1 - e^{\frac{-T}{\tau}}} - \frac{E}{R} \qquad \text{(A)} \end{aligned} \tag{14.113}$$

The peak-to-peak ripple current is independent of load emf, E, and twice that given by equation (14.15).
The mean output current is given by

$$\overline{I}_o = \frac{(\overline{V}_o - E)}{R} = \frac{((1-2\delta)V_s - E)}{R} \qquad \text{(A)} \tag{14.114}$$

which can be positive or negative, as seen in figure 14.22c and d.

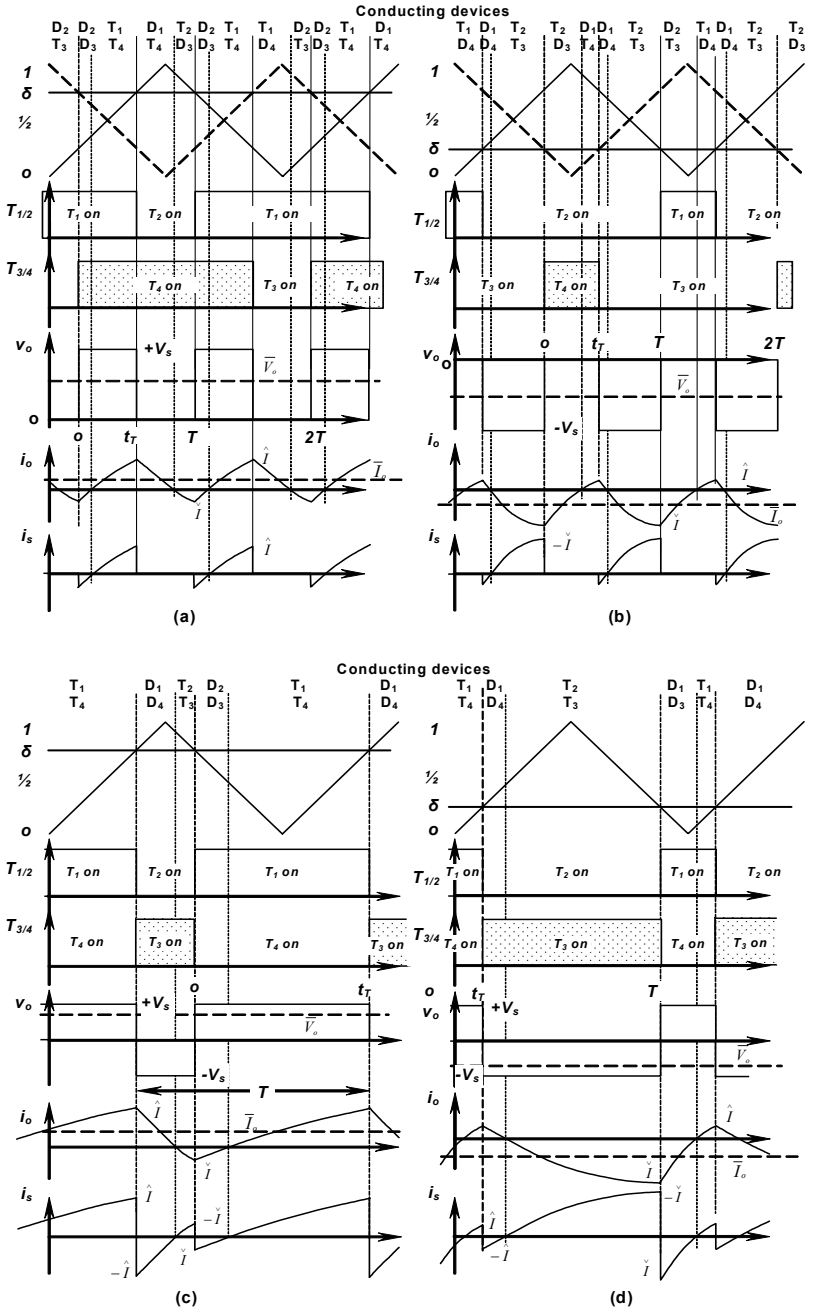


Figure 14.22. *Four-quadrant dc chopper circuit waveforms: multilevel (three-level) output voltage (a) with $\overline{V}_o > 0$ and $\overline{I}_o > 0$; (b) with $\overline{V}_o < 0$ and $\overline{I}_o < 0$; bipolar (two-level) output voltage (c) with $\overline{V}_o > 0$ and $\overline{I}_o > 0$; (d) with $\overline{V}_o < 0$ and $\overline{I}_o < 0$.*

Figures 14.22c and d show chopper output voltage and current waveforms for conditions of positive average voltage and current in part (c) and negative average voltage and current in part (d). Each part is shown with the current having a positive maximum value and a negative minimum value. Such a load current condition involves activation of all possible chopper conducting paths (sequences) as shown at the top of each part in figure 14.22 and transposed to table 14.3A. The table shows how the conducting device possibilities (states) decrease if the minimum value is positive or the maximum value is negative.

Table 14.3A: Four-quadrant chopper bipolar (two-level) output voltage states

Conducting devices sequences								
$\overline{V}<0$					$\overline{V}>0$			
T_1	D_1			$\check{I}>0$	T_1	D_1		
T_4	D_4				T_4	D_4		
$\overline{V}<0$					$\overline{V}>0$			
T_1	D_1	T_2	D_2	$\hat{I} > 0$	T_1	D_1	T_2	D_2
T_4	D_4	T_3	D_3	$\check{I} < 0$	T_4	D_4	T_3	D_3
$\overline{V}<0$					$\overline{V}>0$			
		T_2	D_2	$\hat{I}<0$			T_2	D_2
		T_3	D_3				T_3	D_3

If the minimum output current is positive, that is, $\check{I}$ is positive, then only components for a first and fourth quadrant chopper conduct. Specifically T_2, T_3, D_2, and D_3 do not conduct. Examination of figure 12.14c shows that the output current conduction states are as shown in table 14.3A for $\check{I}>0$.

If the output current never goes positive, that is $\hat{I}$ is negative, then T_1, T_4, D_1, and D_4 do not conduct. The conducting sequence becomes as shown in table 14.3A for $\hat{I}<0$. Because the output is bipolar ($\pm V_s$), the average chopper output voltage, $\overline{V}_o$ does not affect the three possible steady state sequences. Table 14.3A shows that the conducting devices are independent of the average output voltage polarity. That is, the switching states are the same on the left and right sides of table 14.3A.

The transition between these three possible sequences, due to a current level polarity change, is seamless. The only restriction is that both switches in any leg do not conduct simultaneously. This is ensured by inserting a brief dead-time between a switch turning off and its leg complement being turned on. That is, dead-time between the switching of the complementary pair (T_1-T_2), and in the other leg the complementary pair is (T_3-T_4).

14.6.2 Unified four-quadrant dc chopper - multilevel voltage output switching

In order to generate three output states, specifically $\pm V_s$ and 0V, two triangular references are used which are displaced by 180° from one another as shown in figure 14.22a and b. One carrier triangle is used to specify the state of the leg formed by T_1 and T_2 (the complement of T_1), while the other carrier triangle specifies the state of leg formed by switches T_3 and T_4, (the complement of T_3). The output voltage level switches between $+V_s$, 0V, and $-V_s$ depending on the modulation index δ, such that $0 \le \delta \le 1$. A characteristic of the output voltage is that, depending on δ, only a maximum of two of the three states appear in the output, in steady-state. The 0V state is always one of the two alternating states. An alternative method to generate the same switching waveforms, is to us one triangular carrier and two references, δ and 1-δ.

If $\boldsymbol{\delta = 0}$ then T_1 and T_4 never turn-on since T_2 and T_3 conduct continuously which impresses $-V_s$ across the load. As δ increases from zero, the 0V state appears as well as the $-V_s$ state, the later of which decreases in duration as δ increases.

At $\boldsymbol{\delta = ½}$ the output is zero since T_2 and T_3 (or T_1 and T_4) are never on simultaneously to provide a path involving the dc source. The output voltage is formed by alternating 0V loops (T_1 and T_3 on, alternating to T_2 and T_4 on, etc.). The average output voltage is therefore zero.

At the extreme $\boldsymbol{\delta = 1}$, T_1 and T_4 are on continuously and V_s is impressed across the load. As δ is reduced from one, the 0V state is introduced, progressively lengthening to all of the period as δ reduces to ½.

The voltage output in terms of the triangular level v_Δ reference is defined by

For $0 \le \delta < ½$
- $v_\Delta > \delta$, $v_o = -V_s$
- $v_\Delta < \delta$, $v_o = 0$

For $\delta = ½$
- $v_\Delta > \delta$, $v_o = 0$
- $v_\Delta < \delta$, $v_o = 0$

For $½ > \delta \ge 1$
- $v_\Delta > \delta$, $v_o = 0$
- $v_\Delta < \delta$, $v_o = V_s$

From figure 14.22b **for $\boldsymbol{\delta < ½}$**, the average output voltage varies linearly with δ such that

$$\overline{V}_o = \frac{1}{T}\left(\int_0^{t_T} 0\,dt + \int_{t_T}^{T} -V_s\,dt\right) = \frac{1}{T}(t_T - T)V_s = \left(\frac{t_T}{T} - 1\right)V_s \qquad (14.115)$$

Examination of figure 14.22b reveals that the relationship between t_T and δ must produce

when $\delta = 0$: $t_T = 0$ and $v_o = -V_s$

when $\delta = ½$: $t_T = T$ and $v_o = 0$

that is

$$\delta = ½\frac{t_T}{T}$$

which on substituting for t_T/T in equation (14.115) gives

$$\overline{V}_o = \left(\frac{t_T}{T} - 1\right)V_s = (2\delta - 1)\,V_s \qquad (14.116)$$

From figure 14.22a **for $\boldsymbol{\delta > ½}$**, the average output voltage varies linearly with δ such that

$$\overline{V}_o = \frac{1}{T}\left(\int_0^{t_T} V_s\,dt + \int_{t_T}^{T} 0\,dt\right) = V_s\frac{t_T}{T} \qquad (14.117)$$

Examination of figure 14.22a reveals that the relationship between t_T and δ must produce

when $\delta = ½$: $t_T = 0$ and $v_o = 0$

when $\delta = 1$: $t_T = T$ and $v_o = V_s$

that is

$$\delta = ½\left(\frac{t_T}{T} + 1\right)$$

which on substituting for t_T/T in equation (14.117) gives

$$\overline{V}_o = (2\delta - 1)V_s \qquad (14.118)$$

Alternately, if one pole produces δ then the other pole produces the complement ,1 - δ, such that the output is the difference, $\delta - (1 - \delta) = 2\delta - 1$.
Since the same expression results for $\delta \le ½$ with bipolar switching, the average output current is the same for the range $0 \le \delta \le 1$, that is

$$\overline{I}_o = \left(\overline{V}_o - E\right)\Big/R = \left((2\delta - 1)V_s - E\right)\Big/R \qquad \text{(A)} \qquad (14.119)$$

which can be positive or negative, depending on δ and the load emf, E.

Although the average voltage equations of the multilevel and bipolar controlled dc choppers are the same, the rms voltage and ripple voltage differ, as does the peak-to-peak output ripple current. Unlike the bipolar controlled chopper, the rms voltage for the multilevel controlled chopper is not a single continuous function.

 # For $\boldsymbol{\delta \le ½}$ the rms load voltage is

$$V_{rms} = \left[\frac{1}{T}\int_{t_T}^{T}(V_s)^2\,dt\right]^{½} = \sqrt{1 - 2\delta}\;V_s \qquad (14.120)$$

The output ac ripple voltage is

$$\begin{aligned} V_r &= \sqrt{V_{rms}^2 - V_o^2} \\ &= \sqrt{\left(\sqrt{1-2\delta}\,V_s\right)^2 - \left((2\delta-1)V_s\right)^2} \\ &= \sqrt{2}\,V_s\sqrt{\delta(1-2\delta)} \end{aligned} \qquad (14.121)$$

The output voltage ripple factor is

$$\begin{aligned} RF &= \frac{V_r}{\overline{V}_o} = \sqrt{\left(\frac{V_{rms}}{\overline{V}_o}\right)^2 - 1} = \sqrt{FF^2 - 1} \\ &= \sqrt{2 \times \frac{\delta}{1-2\delta}} = \sqrt{\left(\frac{1}{1-2\delta}\right)^2 - 1} \end{aligned} \qquad (14.122)$$

Thus as the duty cycle $\delta \to 0$, the ripple factor tends to zero, consistent with dc output voltage, that is V_r = 0. The ripple factor is undefined when the average output voltage is zero, at δ = ½.

The minimum rms ripple voltage in the output occurs when δ=½ or 0 giving an rms ripple voltage of zero, since the average is a dc value at the extremes (0V and -V_s respectively). The maximum ripple occurs at δ = ¼, when V_r = ½V_s, which is the same as when δ = ¾, (but half that obtained with the bipolar output control method, V_s).

 # For $\delta \geq$ ½ the rms load voltage is

$$\begin{aligned} V_{rms} &= \left[\frac{1}{T}\int_{t_T}^{T}(-V_s)^2\,dt\right]^{1/2} \\ &= \sqrt{2\delta - 1}\;\; V_s \end{aligned} \qquad (14.123)$$

The output ac ripple voltage is

$$\begin{aligned} V_r &= \sqrt{V_{rms}^2 - V_o^2} \\ &= \sqrt{\left(\sqrt{2\delta-1}\,V_s\right)^2 - \left((2\delta-1)V_s\right)^2} = \sqrt{2}\,V_s\sqrt{(2\delta-1)(1-\delta)} \end{aligned} \qquad (14.124)$$

The minimum rms ripple voltage in the output occurs when δ = ½ or 1 giving an rms ripple voltage of zero, since the average is a dc value at the extremes (0V and V_s respectively). The maximum ripple occurs at δ = ¾, when V_r = ½V_s, which is half that obtained with the bipolar output control method.

The output voltage ripple factor is

$$\begin{aligned} RF &= \frac{V_r}{\overline{V}_o} = \sqrt{FF^2 - 1} = \sqrt{\left(\frac{V_{rms}}{\overline{V}_o}\right)^2 - 1} \\ &= \sqrt{2 \times \frac{1-\delta}{2\delta-1}} = \sqrt{\left(\frac{1}{2\delta-1}\right)^2 - 1} \end{aligned} \qquad (14.125)$$

Thus as the duty cycle $\delta \to 1$, the ripple factor tends to zero, consistent with the output being dc, that is V_r = 0. The ripple factor is undefined when the average output voltage is zero, at δ = ½.
Circuit operation is characterized by three time domain equations.

During the **on-period for T1 and T4**, when $v_o(t) = V_s$

$$L\frac{di_o}{dt} + Ri_o + E = V_s$$

which yields

$$i_o(t) = \frac{V_s - E}{R}\left(1 - e^{\frac{-t}{\tau}}\right) + \check{I}\,e^{\frac{-t}{\tau}} \qquad \text{for} \quad 0 \leq t \leq t_T \text{ and } \delta \geq ½ \qquad (14.126)$$

During the **on-period for T2 and T3**, when $v_o(t) = -V_s$

$$L\frac{di_o}{dt} + Ri_o + E = -V_s$$

which, after shifting the zero time reference to t_T, gives

$$i_o(t) = -\frac{V_s + E}{R}\left(1 - e^{\frac{-t}{\tau}}\right) + \hat{I}\,e^{\frac{-t}{\tau}} \qquad \text{for} \quad 0 \leq t \leq T - t_T \text{ and } \delta \leq ½ \qquad (14.127)$$

The third equation is for a zero voltage loop.
During the **switch off-period**, when $v_o(t)$ = 0

$$L\frac{di_o}{dt} + Ri_o + E = 0$$

which, after shifting the zero time reference, in figure 14.22a or b, gives

$$\begin{aligned} &i_o(t) = -\frac{E}{R}\left(1 - e^{\frac{-t}{\tau}}\right) + \hat{I}\,e^{\frac{-t}{\tau}} \\ &0 \leq t \leq t_T \text{ and } \delta \leq ½ \\ &0 \leq t \leq T - t_T \text{ and } \delta \geq ½ \end{aligned} \qquad (14.128)$$

The initial conditions $\hat{I}$ and $\check{I}$ are determined by using the usual steady-state boundary condition method and are dependent on the transition states. For example, for continuous steady-state transitions between +V_s loops and 0V loops, the boundary conditions are given by

$$\begin{aligned} \text{where} \quad \hat{I} &= \frac{V_s}{R}\frac{1 - e^{\frac{-t_T}{\tau}}}{1 - e^{\frac{-T}{\tau}}} - \frac{E}{R} \qquad \text{(A)} \\ \text{and} \quad \check{I} &= \frac{V_s}{R}\frac{e^{\frac{t_T}{\tau}} - 1}{e^{\frac{T}{\tau}} - 1} - \frac{E}{R} \qquad \text{(A)} \end{aligned} \qquad (14.129)$$

Figures 14.22a and b show output voltage and current waveforms for conditions of positive average voltage and current in part (a) and negative average voltage and current in part (b). Each part is shown with the current having a positive maximum value and a negative minimum value. Such a load current condition involves the activation of all possible chopper conducting paths, which are shown at the top of each part in figure 14.22 and transposed to table 14.3B. The conducting device possibilities decrease if the minimum value is positive or the maximum value is negative.

Table 14.3B: A Four-quadrant chopper multilevel (three-level) output voltage states

Conducting devices sequences																
$\overline{V} > 0$									$\check{I} > 0$	$\overline{V} < 0$							
		T_1	D_1			T_1	T_1			T_1	D_1			D_1	D_1		
		T_4	T_4			T_4	D_4			D_4	D_4			T_4	D_4		
$\overline{V} > 0$										$\overline{V} < 0$							
D_2	D_2	T_1	D_1	T_2	D_2	T_1	T_1		$\check{I} < 0$	T_1	D_1	T_2	T_2	D_1	D_1	T_2	D_2
T_3	D_3	T_4	T_4	D_3	D_3	T_4	D_4		$\hat{I} > 0$	D_4	D_4	T_3	D_3	T_4	D_4	T_3	T_3
$\overline{V} > 0$										$\overline{V} < 0$							
D_2	D_2			T_2	D_2				$\hat{I} < 0$			T_2	T_2			T_2	D_2
T_3	D_3			D_3	D_3							T_3	D_3			T_3	T_3

If the minimum output current is positive, that is, $\check{I}$ is positive, then only components for a first and fourth quadrant chopper conduct. Specifically T_2, T_3, D_2, and D_3 do not conduct, thus do not appear in the output sequence. Examination of figure 12.14c shows that the output current conduction states are as shown in table 14.3B for $\check{I} > 0$.
If the output current never goes positive, that is $\hat{I}$ is negative, then T_1, T_4, D_1, and D_4 do not conduct, thus do not appear in the output device sequence. The conducting sequence is as shown in table 14.3B for $\hat{I} < 0$.
Unlike the bipolar control method, the output sequence is affected by the average output voltage level, as well as the polarity of the output current swing. The transition between the six possible sequences due to load voltage and current polarity changes, is seamless. The only restriction is that both switching devices in any leg do not conduct simultaneously. This is ensured by inserting a brief dead-time between a switch turning off and its leg complement being turned on.

Example 14.7: ***Four-quadrant dc chopper***

The H-bridge, dc-to-dc chopper in figure 14.21 feeds an inductive load of 10 Ω resistance, 50mH inductance, and back emf of 55V dc, from a 340V dc source. If the chopper is operated with a 200Hz multilevel carrier as in figure 14.22 a and b, with a modulation depth of δ = ¼, determine:

i. the average output voltage and switch T_1 on-time
ii. the rms output voltage and ac ripple voltage, hence voltage ripple and form factors
iii. the average output current, hence quadrant of operation
iv. the electromagnetic power being extracted from the back emf *E*.

If the mean load current is to be halved, what is

v. the modulation depth, δ, requirement
vi. the average output voltage and the corresponding switch T_1 on-time
vii. the electromagnetic power being extracted from the back emf E?

Solution

The main circuit and operating parameters are

- modulation depth δ = ¼
- period $T_{carrier} = 1/f_{carrier}$ = 1/200Hz = 5ms
- E = 55V and V_s = 340V dc
- load time constant $\tau = L/R$ = 0.05mH/10Ω = 5ms

i. The average output voltage is given by equation (14.116), and for δ < ½,

$$\overline{V}_o = \left(\frac{t_T}{T} - 1\right)V_s = (2\delta - 1)V_s$$
$$= 340\text{V}\times(2\times¼ - 1) = -170\text{V}$$

where

$$t_T = 2\delta T = 2\times¼\times(½\times 5\text{ms}) = 1.25\text{ms}$$

Figure 14.22 reveals that the carrier frequency is half the switching frequency, thus the 5ms in the above equation has been halved. The switches T_1 and T_4 are turned on for 1.25ms, while T_2 and T_3 are subsequently turned on for 3.75ms.

ii. The rms load voltage, from equation (14.120), is

$$V_{rms} = \sqrt{1-2\delta}\ V_s$$
$$= 340\text{V}\times\sqrt{1 - 2\times¼} = 240\text{V rms}$$

From equation (14.121), the output ac ripple voltage, hence voltage ripple factor, are

$$V_r = \sqrt{2}\,V_s\sqrt{\delta(1-2\delta)}$$
$$= \sqrt{2}\times 340\text{V}\times\sqrt{¼\times(1 - 2\times¼)} = 170\text{V ac}$$

$$RF = \frac{V_r}{\overline{V}_o} = \frac{170\text{V}}{|-170\text{V}|} = 1 \qquad FF = \sqrt{RF^2 + 1} = \sqrt{2} = 1.41$$

iii. The average output current is given by equation (14.119)

$$\overline{I}_o = \frac{\overline{V}_o - E}{R} = \frac{(2\delta - 1)V_s - E}{R}$$
$$= \frac{340\text{V}\times(2\times¼ - 1) - 55\text{V}}{10\Omega} = -22.5\text{A}$$

Since both the average output current and voltage are negative (-170V and -22.5A) the chopper with a modulation depth of δ = ¼, is operating in the third quadrant.

iv. The electromagnetic power developed by the back emf E is given by

$$P_E = E\overline{I}_o = 55\text{V}\times(-22.5\text{A}) = -1237.5\text{W}$$

v. The average output current is given by

$$\overline{I}_o = \left(\overline{V}_o - E\right)\Big/R = \left((2\delta - 1)V_s - E\right)\Big/R$$

when the mean current is -11.25A, δ = 0.415, as derived in part vi.

vi. Then, if the average current is halved to -11.25A

$$\overline{V}_o = E + \overline{I}_o R$$
$$= 55\text{V} - 11.25\text{A}\times 10\Omega = -57.5\text{V}$$

The average output voltage rearranged in terms of the modulation depth δ gives

$$\delta = ½\left(1 + \frac{\overline{V}_o}{V_s}\right)$$
$$= ½\times\left(1 + \frac{-57.5\text{V}}{340\text{V}}\right) = 0.415$$

The switch on-time when δ < ½ is given by

$$t_T = 2\delta T = 2\times 0.415\times(½\times 5\text{ms}) = 2.07\text{ms}$$

From figure 14.22b both T_1 and T_4 are turned on for 2.07ms, although, from table 14.3B, for negative load current, $\overline{I}_o$ = -11.25A, the parallel connected freewheel diodes D_2 and D_3 conduct alternately, rather than the switches (assuming $\overline{I}_o < 0$). The switches T_1 and T_4 are turned on for 1.25ms, while T_2 and T_3 are subsequently turned on for 2.93ms.

vii. The electromagnetic power developed by the back emf E is halved and is given by

$$P_E = E\overline{I}_o = 55\text{V}\times(-11.25\text{A}) = -618.75\text{W}$$

♣

Reading list

Dewan, S. B. and Straughen, A., *Power Semiconductor Circuits,* John Wiley & Sons, New York, 1975.

Dubey, G.K., *Power Semiconductor Controlled Drives,* Prentice-Hall International, 1989.

Mohan, N., Undeland, T. M., & Robbins, W.P., *Power Electronics: Converters, Applications & Design,* John Wiley & Sons, New York, 2003.

Problems

14.1. The dc GTO thyristor chopper shown in figure 14.1c operates at 1 kHz and supplies a series 5Ω and 10mH load from an 84V dc battery source. Derive general expressions for the mean load voltage and current, and the load rms voltage at an on-time duty cycle of δ. Evaluate these parameters for δ = 0.25.
[21 V, 4.2 A; 42 V]

14.2. The dc chopper in figure 14.1c controls a load of R = 10 Ω, L = 10mH and 40V battery. The supply is 340V dc and the chopping frequency is 5 kHz. Calculate (a) the peak-to-peak load ripple current, (b) the average load current, (c) the rms load current, (d) the effective input resistance, and (e) the rms switch current.

15

DC to AC Inverters – Switched Mode

Inversion is the conversion of dc power to ac power at a desired output voltage or current and frequency. A static semiconductor inverter circuit performs this electrical energy inverting transformation. The terms voltage-fed and current-fed are used in connection with the output from inverter circuits.
A *voltage-source inverter (VSI)* is one in which the dc input voltage is essentially constant and independent of the load current drawn. The inverter specifies the load voltage while the drawn current shape is dictated by the load. Being a voltage source, an open circuit output is allowable.
A *current-source inverter (CSI)* is one in which the source, hence the load current is predetermined and the load impedance determines the output voltage. The supply current cannot change quickly. This current is controlled by series dc supply inductance which prevents sudden changes in current. The load current magnitude is controlled by varying the input dc voltage to the large inductance, hence inverter response to load changes is slow. Being a current source, the inverter can survive an output short circuit thereby offering fault ride-through properties.
Voltage control may be required to maintain a fixed output voltage when the dc input voltage regulation is poor, or to control power to a load. The inverter and its output can be single-phase, three-phase or multi-phase. Variable output frequency may be required for ac motor speed control where, in conjunction with voltage or current control, constant motor flux can be maintained.
Inverter output waveforms (either voltage or current) are usually rectilinear in nature and as such contain harmonics which may lead to reduced load efficiency and performance. Load harmonic reduction can be achieved by either filtering, selected harmonic-reduction chopping or pulse-width modulation.

The quality of an inverter output is normally evaluated in terms of its *harmonic factor*, ρ, *distortion factor*, μ, and *total harmonic distortion*, *thd*. In section 13.7.2 these first two factors were defined in terms of the supply current. For VSI inverters the factors are redefined in terms of the output voltage harmonics as follows

$$\rho_n = \left|\frac{V_n}{V_1}\right| = n\mu_n \qquad n>1 \tag{15.1}$$

The distortion factor for an individual harmonic is

$$\mu_n = \left|\frac{V_n}{nV_1}\right| = \frac{\rho_n}{n} \tag{15.2}$$

$$thd = \sqrt{\left[\sum_{n\geq 2}^{\infty}\left(\frac{V_n}{n}\right)^2\right]}\ /V_1 = \sqrt{\sum_{n\geq 2}^{\infty}\mu_n^2} = \sqrt{\sum_{n\geq 2}^{\infty}\left(\frac{\rho_n}{n}\right)^2} \tag{15.3}$$

The factor V_n/n is used since the harmonic currents produced in an inductive load attenuate with frequency. The harmonic currents produce unwanted heating and torque oscillations in ac motors, although such harmonic currents are not a drawback to the power delivered to a resistive heating load or an incandescent lighting load. Harmonics reflected back into the input may be problematic.

15.1 dc-to-ac voltage-source inverter bridge topologies

15.1.1 Single-phase voltage-source inverter bridge

Figure 15.1a shows an H-bridge inverter (VSI) for producing an ac voltage and employing switches which may be transistors (MOSFET or IGBT), or at high powers, thyristors (GTO or GCT). Device

conduction patterns are also shown in figures 15.1b and c. With inductive loads (not purely resistive), stored energy at turn-off is fed through the bridge reactive feedback or freewheel diodes D_1 to D_4. These four diodes clamp the load voltage to within the dc supply voltage rails (0 to V_s).

15.1.1i - *Square-wave (bipolar) output*
Figure 15.1b shows waveforms for a square-wave output ($2t_1 = t_2$) where each device is turned on as appropriate for 180°, (that is π) of the output voltage cycle (state sequence 10, 01, 10, ..).
The load current i_L grows exponentially through T_1 and T_2 (state 10) according to

$$V_s = L\frac{di_L}{dt} + i_L R \qquad \text{(V)} \tag{15.4}$$

When T_1 and T_2 are turned off, T_3 and T_4 are turned on (state 01), thereby reversing the load voltage polarity. Because of the inductive nature of the load, the load current cannot reverse instantaneously and load reactive energy flows back into the supply via diodes D_3 and D_4 (which are in parallel with T_3 and T_4 respectively) according to

$$-V_s = L\frac{di_L}{dt} + i_L R \qquad \text{(V)} \tag{15.5}$$

The load current falls exponentially and at zero, T_3 and T_4 become forward-biased and conduct load current, thereby feeding power to the load.

The output voltage is a square wave of magnitude ± V_s, figure 15.1b, and has an rms value of V_s.
For a simple *R-L* load, with time constant $\tau = L/R$, during the first cycle with no initial load current, solving equation (15.4) yields a load current

$$i_L(t) = \frac{V_s}{R}\left(1 - e^{\frac{-t}{\tau}}\right) \qquad \text{(A)} \tag{15.6}$$

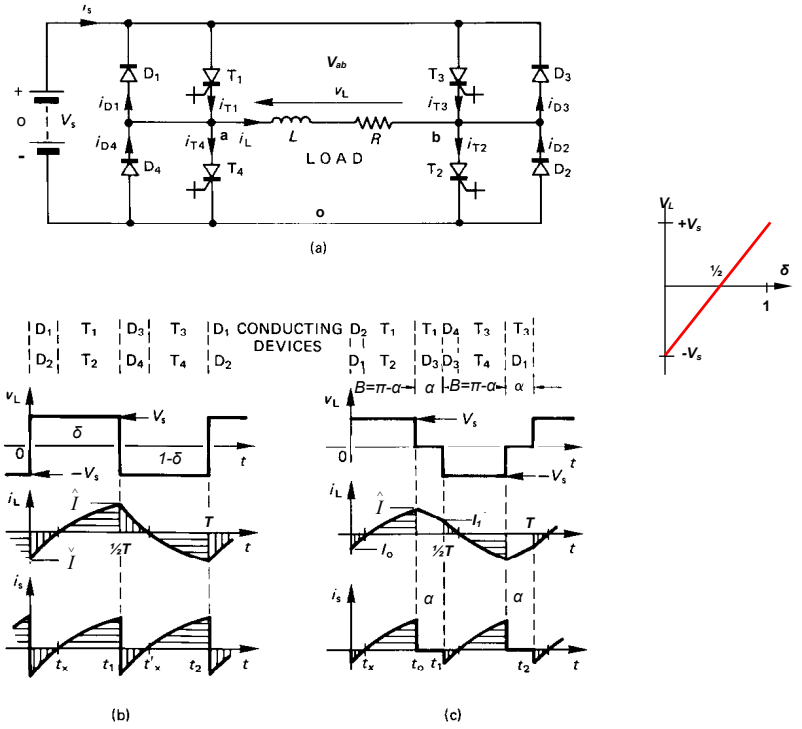


Figure 15.1. *GCT thyristor single-phase bridge inverter:*
(a) circuit diagram; (b) square-wave output voltage; and (c) quasi-square-wave output voltage.

Under steady-state load conditions, the initial current is $\check{I}$ as shown in figure 15.1b, and equation (15.4) yields

$$i_L(t) = \frac{V_s}{R} - \left(\frac{V_s}{R} - \check{I}\right)e^{\frac{-t}{\tau}} \quad \text{(A)} \qquad 0 \le t \le t_1 = \tfrac{1}{2}T \quad \text{(s)} \tag{15.7}$$

for $v_L = V_s$ (V)

$\check{I} \le 0$ (A)

During the second half-cycle ($t_1 \le t \le t_2$) when the supply is effectively reversed across the load, equation (15.5) yields

$$i_L(t) = -\frac{V_s}{R} + \left(\frac{V_s}{R} + \hat{I}\right)e^{\frac{-t}{\tau}} = -\frac{V_s}{R}\left(1 - \left(1 + \tanh\left(\frac{t_1}{2\tau}\right)\right)e^{\frac{-t}{\tau}}\right) \quad \text{(A)} \qquad 0 \le t \le t_2 - t_1 = \tfrac{1}{2}T \quad \text{(s)} \tag{15.8}$$

for $v_L = -V_s$ (V)

$\hat{I} \ge 0$ (A)

A new time axis has been used in equation (15.8) starting at $t = t_1$ in figure 15.1b. Since in steady-state by symmetry, $\hat{I} = -\check{I}$, the initial steady-state current $\hat{I}$ can be found from equation (15.7) when, at $t = t_1$, $i_L = \hat{I}$ yielding

$$\hat{I} = -\check{I} = \frac{V_s}{R}\frac{1 - e^{\frac{-t_1}{\tau}}}{1 + e^{\frac{-t_1}{\tau}}} = \frac{V_s}{R}\tanh\left(\frac{t_1}{2\tau}\right) \quad \text{(A)} \tag{15.9}$$

The zero current cross-over point t_x, shown on figure 15.1b, can be found by solving equation (15.7) for $t = t_x$ when $i_L = 0$, which yields

$$t_x = \tau\, \ell n\left(1 - \frac{\check{I}R}{V_s}\right) = \tau\, \ell n\left(1 + \frac{\hat{I}R}{V_s}\right) \quad \text{(s)} \tag{15.10}$$

The average thyristor current, $\bar{I}_T$, average diode current, $\bar{I}_D$, and mean source current, $\bar{I}_s$ can be found by integration of the load current over the appropriated bounds shown in the following integrals.

$$\bar{I}_T = \frac{1}{t_2}\int_{t_x}^{t_1} i_L(t)\,dt = \frac{1}{t_2}\left[\frac{V_s}{R}(t_1 - t_o) + \tau\left(\frac{V_s}{R} + \hat{I}\right)\left(e^{\frac{-t_x}{\tau}} - e^{\frac{-t_1}{\tau}}\right)\right] \tag{15.11}$$

where i_L is given by equation (15.7) and

$$\bar{I}_D = \frac{1}{t_2}\int_0^{t_x} -i_L(t)\,dt = \frac{1}{t_2}\left[-\frac{V_s}{R}t_x - \tau\left(\frac{V_s}{R} + \hat{I}\right)\left(e^{\frac{-t_x}{\tau}} - 1\right)\right] \tag{15.12}$$

where i_L is given by equation (15.8).

Inspection of the source current waveform in figure 15.1b shows that the average dc voltage source current is related to the average semiconductor device currents by

$$\bar{I}_s = 2\left(\bar{I}_T - \bar{I}_D\right) = \frac{1}{t_2}\left[\frac{V_s}{R}t_1 + \tau\left(\frac{V_s}{R} + \hat{I}\right)\left(e^{\frac{-t_1}{\tau}} - 1\right)\right] \tag{15.13}$$

The steady-state mean power delivered by the dc supply and absorbed by the resistive load component R is given by

$$P_L = \frac{1}{t_1}\int_0^{t_1} V_s\, i_L(t)\;\; dt \;\; = \;\; V_s\bar{I}_s \;\; \left(= I_{Lrms}^2 R\right) \quad \text{(W)} \tag{15.14}$$

where $i_L(t)$ is given by equation (15.7). Rather than integration involving equations (15.7) and (15.8), the mean load power can be used to determine the rms load current:

$$i_{Lrms} = \sqrt{P_L/R} = \sqrt{V_s\bar{I}_s/R} \quad \text{(A)} \tag{15.15}$$

The rms output voltage is V_s and the output fundamental frequency f_o is $f_o = 1/T = 1/2t_1 = 1/t_2$.

The instantaneous output voltage expressed as a Fourier series is given by

$$V_L = \frac{4}{\pi}V_s\sum_{n\,odd}^{\infty}\frac{1}{n}\sin n\omega_o t \quad \text{(V)} \tag{15.16}$$

where $\omega_o = 2\pi f_o = 2\pi/t_2$ and for n = 1 the magnitude of the fundament frequency f_o is $\frac{4}{\pi}V_s$ which is an output rms fundamental voltage v_{o1} of

$$v_{o1} = \frac{2\sqrt{2}}{\pi}V_s = 0.90V_s \quad \text{(V)} \tag{15.17}$$

The load current can be expressed in terms of the Fourier voltage waveform series, that is

$$i_L(\omega t) = \frac{4}{\pi}V_s\sum_{n=1,3,5}^{\infty}\frac{1}{nZ_n}\sin\left(n\omega_o t - \phi_n\right) = \sum_{n=1,3,5}^{\infty} I_n\sin\left(n\omega_o t - \phi_n\right) \tag{15.18}$$

where $I_n = \frac{4}{\pi}\frac{V_s}{nZ_n}$ whence $I_{n\,\text{rms}} = \frac{I_n}{\sqrt{2}}$

$Z_n = \sqrt{R^2 + (n\omega_o L)^2}$ $\phi_n = \tan^{-1} n\omega_o L/R$ such that $\cos\phi_1 = R/Z_1$

The fundamental output power is

$$P_1 = I_1^2 R = \left(\frac{v_{o1}}{Z_1}\right)^2 R = \frac{V_s^2}{R}\left(\frac{2\sqrt{2}}{\pi}\right)^2\cos^2\phi_1 \tag{15.19}$$

The load power is given by the sum of each harmonic I^2R power component, that is

$$P_L = \sum_{n=1,3,5}^{\infty}\left(I_n/\sqrt{2}\right)^2 R = \sum_{n=1,3,5}^{\infty} I_{n\,rms}^2 R \qquad \left(= V_s\bar{I}_s\right) \tag{15.20}$$

Alternately, after integrating equation (15.14), with the load current from equation (15.8)

$$P_L = \frac{V_s^2}{R}\left(1 - \frac{2\tau}{t_1}\frac{1 - e^{-\frac{t_1}{\tau}}}{1 + e^{-\frac{t_1}{\tau}}}\right) = \frac{V_s^2}{R}\left(1 - \frac{2\tau}{t_1}\tanh\left(\frac{t_1}{2\tau}\right)\right) \tag{15.21}$$

From $P_L = I_{rms}^2 R$ the rms loads current is

$$I_{Lrms} = \frac{V_s}{R}\sqrt{1 - \frac{2\tau}{t_1}\tanh\left(\frac{t_1}{2\tau}\right)} \tag{15.22}$$

The load power factor is given by

$$pf = \frac{P}{S} = \frac{I_{Lrsm}^2 R}{I_{Lrms}V_{rms}} = \sqrt{1 - \frac{2\tau}{t_1}\tanh\left(\frac{t_1}{2\tau}\right)} \tag{15.23}$$

15.1.1ii - *Quasi-square-wave (multilevel) output*

The rms output voltage form a H-bridge can be varied by producing a quasi-square output voltage ($2t_1 = t_2$, $t_0 < t_1$) as shown in figure 15.1c. After T_1 and T_2 have been turned on (state 10), at the angle α one device is turned off. If T_1 is turned off (and T_4 is turned on after a short delay), the load current slowly freewheels through T_2 and D_4 (state 00) in a zero voltage loop according to

$$0 = L\frac{di_L}{dt} + i_L R \quad \text{(V)} \tag{15.24}$$

When T_2 is turned off and T_3 turned on (state 01), the remaining load current rapidly reduces to zero back into the dc supply V_s, through diodes D_3 and D_4. When the load current reaches zero, T_3 and T_4 become forward biased and the output current reverses, through T_3 and T_4.

The output voltage shown in figure 15.1c consists of a sequence of non-zero voltages $\pm V_s$, alternated with zero output voltage periods. During the zero output voltage period a diode and switch conduct, firstly T_1 and D_3 in the first period, and T_3 and D_1 in the second zero output period. In each case, a *zero voltage loop* is formed by a switch, diode, and the load. The next two zero output sequences would be T_2 and D_4 then T_4 and D_2, forming *alternating zero voltage loops* (sequence 10, 00, 01, 11, 10, ..) rather than repeating a continuous T_1 and D_3 then T_3 and D_1 sequence of zero voltage loops (sequence 10, 11,

01, 11, 10, .. or sequence 10, 00, 01, 00, 10, ..). By alternating the zero voltage loops (between states 00 and 11), losses are uniformly distributed between the semiconductors, device switching frequency is half that experienced by the load, and a finer output voltage resolution is achievable.
With reference to figure 15.1c, the load current i_L for an applied quasi square-wave voltage is defined as follows.

(i) $v_L > 0$

$$i_{L_I}(t) = \frac{V_s}{R} - \left(\frac{V_s}{R} - I_o\right) e^{\frac{-t}{\tau}} \qquad 0 \le t \le t_o \qquad (15.25)$$

for $I_o \le 0$ (A)

(ii) $v_L = 0$

$$i_{L_{II}}(t) = \hat{I}\, e^{\frac{-t}{\tau}} \qquad 0 \le t \le t_1 - t_o \qquad (15.26)$$

for $\hat{I} \ge 0$ (A)

(iii) $v_L < 0$

$$i_L(t) = -\frac{V_s}{R} + \left(\frac{V_s}{R} + I_1\right) e^{\frac{-t}{\tau}} = -i_{L_I}(t) \qquad 0 \le t \le t_o \qquad (15.27)$$

for $I_1 \ge 0$ (A)

The currents I_o, $\hat{I}$, and I_1 are given by

$$I_o = -\frac{V_s}{R}\,\frac{e^{\frac{-t_1+t_o}{\tau}} - e^{\frac{-t_1}{\tau}}}{1+e^{\frac{-t_1}{\tau}}} \quad \text{(A)} \qquad (15.28)$$

$$\hat{I} = \frac{V_s}{R}\,\frac{1-e^{\frac{-t_o}{\tau}}}{1+e^{\frac{-t_1}{\tau}}} \quad \text{(A)} \qquad (15.29)$$

$$I_1 = -I_o \quad \text{(A)} \qquad (15.30)$$

The zero current cross-over instant, t_x, shown in figure 15.1c, is found by solving equation (15.25) for t when i_L equals zero current.

$$t_x = \tau\, \ell n\left(1 - \frac{I_o R}{V_s}\right) = \tau\, \ell n\left(1 + \frac{I_1 R}{V_s}\right) \qquad (15.31)$$

The average thyristor current, $\bar{I}_T$, average diode current, $\bar{I}_D$, and mean source current, $\bar{I}_s$ can be found by integration of the load current over the appropriated bounds (assuming alternating zero volt loops).

$$\bar{I}_T = \frac{1}{t_2}\int_{t_x}^{t_1} i_{L_I}(t)\,dt + \frac{1}{2t_2}\int_0^{t_1-t_o} i_{L_{II}}(t)\,dt \qquad (15.32)$$

where i_L is given by equations (15.25) and (15.26) for the respective integrals, and

$$\bar{I}_D = \frac{1}{t_2}\int_0^{t_x} -i_{L_I}(t)\,dt + \frac{1}{2t_2}\int_0^{t_1-t_o} i_{L_{II}}(t)\,dt \qquad (15.33)$$

where i_L is given by equations (15.25) and (15.26) for the respective integrals.

Inspection of the source current waveform in figure 15.1c shows that the average source current is related to the average semiconductor device currents by

$$\bar{I}_s = \frac{1}{t_1}\int_0^{t_o} i_{L_I}(t)dt = 2\left(\bar{I}_T - \bar{I}_D\right) \qquad (15.34)$$

The steady-state mean load and dc source powers are

$$P_L = \frac{1}{t_1}\int_0^{t_o} V_s\, i_L(t)\;\; dt \;\; = \;\; V_s \bar{I}_s \qquad \left(= I_{Lrms}^2 R\right) \qquad \text{(W)} \qquad (15.35)$$

where $i_L(t)$ is given by equation (15.25). The mean load power can be used to determine the rms load current:

$$I_{Lrms} = \sqrt{P_L / R} = \sqrt{V_s \bar{I}_s / R} \quad \text{(A)} \qquad (15.36)$$

The output fundamental frequency f_o is $f_o = 1/2t_1 = 1/t_2$.

The variable rms output voltage, for $0 \le \alpha \le \pi$, is

$$v_{rms} = \sqrt{\frac{1}{t_1}\int_0^{t_o} V_s^2\, dt} = \sqrt{1-\alpha/\pi}\; V_s \qquad (15.37)$$

and the output fundamental frequency f_o is $f_o = 1/t_2$. This equation for rms output voltage shows that only the n^{th} harmonic can be eliminated when $\cos \tfrac{1}{2} n\alpha = 0$, that is for $\alpha = \pi / n$. In so eliminating the n^{th} harmonic, from equation (15.38), the magnitude of the fundamental is reduced to $\tfrac{1}{\pi} V_s \cos \alpha/n$.
The output voltage V_L in its Fourier coefficient series form is given by

$$V_L = \frac{4}{\pi} V_s \sum_{n\ odd}^{\infty} \frac{\cos \tfrac{1}{2} n\alpha}{n} \sin n\omega_o t \qquad \text{(V)} \qquad (15.38)$$

and for n = 1, the rms fundamental of the output voltage v_{o1} is given by

$$v_{o1} = \frac{2\sqrt{2}}{\pi} V_s \cos \tfrac{1}{2}\alpha = 0.90 \times V_s \times \cos \tfrac{1}{2}\alpha \qquad \text{(V)} \qquad (15.39)$$

The characteristics of these load voltage harmonics are shown in figure 15.2.

An alternative approach is to consider the control of one leg phase shifted by β radians with respect of the other leg. The phase output voltage for each leg, with respect to the supply (artificial) mid point, o, is

$$V_{ao} = \sum_{n=1,3,5,\ldots}^{\infty} \frac{2V_s}{n\pi} \sin n\omega_o t$$

$$V_{bo} = \sum_{n=1,3,5,\ldots}^{\infty} \frac{2V_s}{n\pi} \sin\left(n\left(\omega_o t - \beta\right)\right)$$

The output V_{ab} is then given by the difference, that is

$$V_{ao} - V_{bo} = V_{ab} = \sum_{n=1,3,5,\ldots}^{\infty} \frac{2V_s}{n\pi} \sin n\omega_o t - \sum_{n=1,3,5,\ldots}^{\infty} \frac{2V_s}{n\pi} \sin\left(n\left(\omega_o t - \beta\right)\right)$$

$$= \sum_{n=1,3,5,\ldots}^{\infty} \frac{4V_s}{n\pi} \sin\frac{n\beta}{2} \cos n\omega_o t$$

Expressing the phase shift angle β in terms of the delay angle α, where $\beta = \pi - \alpha$, yields equation (15.38).

The load current can be expressed in terms of the Fourier voltage waveform series, that is

$$i_L(\omega t) = \frac{V_L}{Z_L} = \frac{4}{\pi} V_s \sum_{n=1,3,..}^{\infty} \frac{\cos \tfrac{1}{2} n\alpha}{n Z_n} \sin\left(n\omega_o t - \phi_n\right) = \sum_{n=1,3,5,..}^{\infty} I_n \sin\left(n\omega_o t - \phi_n\right) \qquad (15.40)$$

where $I_n = \frac{4}{\pi}\frac{V_s}{n Z_n} \cos \tfrac{1}{2} n\alpha$ whence $I_{n\,\mathrm{rms}} = \frac{I_n}{\sqrt{2}}$

$$Z_n = \sqrt{R^2 + (n\omega_o L)^2} \qquad \phi_n = \tan^{-1} n\omega_o L / R$$

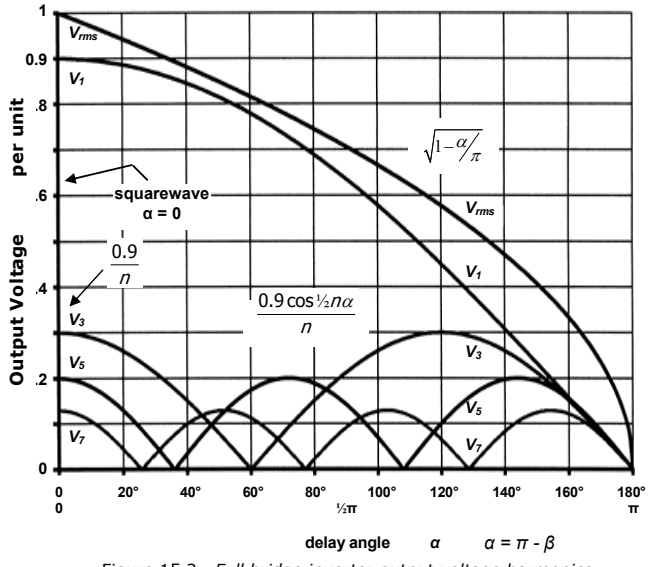


Figure 15.2. *Full bridge inverter output voltage harmonics normalised with respect to square wave rms output voltage, V_{rms}=V_s.*

The load power is given by the sum of each harmonic i^2R power component, that is

$$P_L = \sum_{n=1,3,5}^{\infty} \left(I_n/\sqrt{2}\right)^2 R = \sum_{n=1,3,5,\ldots}^{\infty} I_{n\,rms}^2 R \qquad \left(= V_s \bar{I}_s\right) \tag{15.41}$$

The load power and rms current can be evaluated from equations (15.21) and (15.22) provided the rms voltage given by equation (15.37) replaces V_s. That is

$$P_L = \frac{V_s^2}{R}\left(1 - \alpha/\pi\right)\left(1 - \frac{2\tau}{t_1}\tanh\left(\frac{t_1}{2\tau}\right)\right) = i_{Lrms}^2 R \tag{15.42}$$

$$i_{Lrms} = \frac{V_s}{R}\sqrt{1-\alpha/\pi}\sqrt{1 - \frac{2\tau}{t_1}\tanh\left(\frac{t_1}{2\tau}\right)} \tag{15.43}$$

The load power factor is independent of α and is given by equation (15.23), that is

$$pf = \frac{P}{S} = \frac{i_{Lrsm}^2 R}{i_{Lrms} v_{rms}} = \sqrt{1 - \frac{2\tau}{t_1}\tanh\left(\frac{t_1}{2\tau}\right)} \tag{15.44}$$

A variation of the basic four-switch dc to ac single-phase H-bridge is the half-bridge version where two series switches (one pole or leg) and diodes are replaced by a split two-capacitor voltage source, as shown in figure 15.3. This reduces the number of semiconductors and gate circuit requirements, but at the expense of halving the maximum output voltage. Example 15.3 illustrates the half-bridge and its essential features. Behaviour characteristics are as for the full-bridge, square-wave, single-phase inverter but V_s is replaced by ½V_s in the appropriate equations. Only a rectangular-wave bipolar output voltage can be obtained. Since zero volt loops cannot be created, no rms voltage control is possible. The rms output voltage is ½V_s, while the output power is a quarter that of the full H-bridge.

Example 15.1: ***Single-phase H-bridge with an L-R load***

A single-phase H-bridge inverter, as shown in figure 15.1a, supplies a 10 Ω resistance with inductance 50 mH, from a 340 V dc source. If the bridge is operating at 50 Hz (output), determine the average supply current and the load rms voltage and current and steady-state current waveforms with

i. a square-wave output
ii. a symmetrical quasi-square-wave output with a 50 per cent on-time.

Solution

The time constant of the load, τ = 0.05mH/10Ω = 5 ms, t_1 = 10ms and t_2 = 20ms.

i. The output voltage rms value is 340 V ac.
Equation (15.9) gives the load current at the time when the supply polarity is reversed across the load, as shown in figure 15.1b, that is

$$\hat{I} = -\check{I} = \frac{V_s}{R}\frac{1-e^{\frac{-t_2}{\tau}}}{1+e^{\frac{-t_1}{\tau}}} \qquad \text{(A)}$$

where t_1 = 10 ms. Therefore

$$\hat{I} = -\check{I} = \frac{340\text{V}}{10\Omega}\times\frac{1-e^{-2}}{1+e^{-2}} \qquad \text{(A)}$$
$$= 25.9\text{A}$$

When v_L = +340 V, from equation (15.7) the load current is given by

$$i_L = 34 - (34+25.9)\times e^{-200t} = 34 - 59.9e^{-200t} \qquad 0 \le t \le 10\text{ ms}$$

From equation (15.10) the zero current cross-over time, t_x, occurs $5\text{ms}\times\ell n\left(1 + 25.9\text{A}\times 10\Omega/340\text{V}\right)$ = 2.83ms after load voltage reversal.
When v_L = -340 V, from equation (15.8) the load current is given by

$$i_L = -34 + (34+25.9)\times e^{-200t} = -34 + 59.9e^{-200t} \qquad 0 \le t \le 10\text{ ms}$$

The mean power delivered to the load is given by equation (15.14), that is

$$P_L = \frac{1}{10\text{ms}}\int_0^{10\text{ms}} 340\text{V}\times\{34 - 59.9\times e^{-200t}\}\,dt$$
$$= 2755\text{ W}$$

From $P = i^2R$, the load rms current is

$$i_{Lrms} = \sqrt{P_L/R} = \sqrt{2755\text{W}/10\Omega} = 16.60\text{A} \quad \text{and} \quad \bar{I}_s = P_L/V_s = 2755\text{W}/340\text{V} = 8.1\text{A}$$

These power and rms current results can be confirmed with equations (15.21) and (15.22).

ii. The quasi-square output voltage has a 5 ms on-time, t_o, and a 5 ms period of zero volts.
From equation (15.37) the rms output voltage is
$V_s\sqrt{1-5\text{ms}/10\text{ms}} = V_s/\sqrt{2} = 240\text{V rms}$.
The current during the different intervals is specified by equations (15.25) to (15.30). Alternately, the steady-state load current equations can be specified by determining the load current equations for the first few cycles at start-up until steady-state conditions are attained.
First 5 ms on-period when v_L = 340 V and initially i_L = 0 A

$$i_L = 34 - 34\,e^{-200t}$$

and at 5ms, i_L = 21.5A
First 5 ms zero-period when v_L = 0 V

$$i_L = 21.5\,e^{-200t}$$

and at 5ms, i_L =7.9A
Second 5 ms on-period when v_L = -340 V

$$i_L = -34 + (34+7.9)\times e^{-200t}$$

with i_L = 0 at 1 ms and ending with i_L = -18.6 A
Second 5 ms zero-period when v_L = 0 V

$$i_L = -18.6\,e^{-200t}$$

ending with i_L = -6.8A
Third 5 ms on-period when v_L = 340 V

$$i_L = 34 - (34+6.8)\times e^{-200t}$$

with i_L = 0 at 0.9 ms and ending with i_L = 19.0 A
Third 5 ms zero-period when v_L = 0 V

$$i_L = 19.0\,e^{-200t}$$

ending with i_L = 7.0A
Fourth 5 ms on-period when v_L = -340 V

$$i_L = -34 + (34+7.0)\times e^{-200t}$$

with i_L = 0 at 0.93 ms and ending with i_L = -18.9 A
Fourth 5 ms zero-period when v_L = 0 V

$$i_L = -18.9\,e^{-200t}$$

ending with i_L = -7.0A

Steady-state load current conditions have been reached and the load current waveform is as shown in figure 15.1c. Convergence of an iterative solution is more rapid if the periods considered are much longer than the load time constant (and vice versa).
The mean load power for the quasi-square wave is given by

$$P_L = \frac{1}{10\text{ms}}\int_0^{5\text{ms}} 340\text{V}\times\{34 - 41\times e^{-200t}\}\,dt$$
$$= 1378\text{ W}$$

The load rms and supply currents are

$$i_{Lrms} = \sqrt{P_L/R} = \sqrt{1378\text{W}/10\Omega} = 11.74\text{A} \qquad \bar{I}_s = P_L/V_s = 1378\text{W}/340\text{V} = 4.05\text{A}$$

♣

Example 15.2: ***H-bridge inverter ac output factors***

In each waveform case (square and quasi-square) of example 15.1a calculate

i. the average and peak current in the switches
ii. the average and peak current in the diodes
iii. the peak blocking voltage of each semiconductor type
iv. the average source current
v. the harmonic factor and distortion factor of the lowest order harmonic
vi. the total harmonic distortion

Solution

Square-wave

i. The peak current in the switch is $\hat{I}$ = 25.9 A and the current zero cross-over occurs at t_x =2.83ms. The average switch current, from equation (15.11) is

$$\bar{I}_T = \frac{1}{20\text{ms}}\int_{2.83\text{ms}}^{10\text{ms}} (34 - 59.9\, e^{-200t})\, dt$$
$$= 5.71 \text{ A}$$

ii. The peak diode current is 25.9 A. The average diode current from equation (15.12) is

$$\bar{I}_D = \frac{1}{20\text{ms}}\int_{0}^{2.83\text{ms}} (34 - 59.9\, e^{-200t})\, dt$$
$$= 1.66 \text{ A}$$

iii. The maximum blocking voltage of each device is 340 V dc.

iv. The average supply current is

$$\bar{I}_s = 2\left(\bar{I}_T - \bar{I}_D\right) = 2\times(5.71\text{A} - 1.66\text{A}) = 8.10\text{A}$$

This results in the supply delivery power of 340Vdc × 8.10A = 2754W

v. From equation (15.16), with the third as the lowest harmonic, the distortion factors are

$$hf = \rho_3 = \left|\frac{V_3}{V_1}\right| = \tfrac{1}{3}, \text{ that is, } 33\tfrac{1}{3} \text{ per cent}$$

$$df = \mu_3 = \left|\frac{V_3}{3V_1}\right| = \tfrac{1}{9}, \text{ that is, } 11.11 \text{ per cent}$$

vi. From equation (15.16)

$$thd = \sqrt{\sum\left(\frac{V_n}{n}\right)^2}\Big/ V_1$$
$$= \sqrt{\left(\tfrac{1}{3}\right)^2 + \left(\tfrac{1}{5}\right)^2 + \left(\tfrac{1}{7}\right)^2 + \ldots..}$$
$$= 46.2 \text{ per cent}$$

Quasi-square-wave, $\alpha = \tfrac{1}{2}\pi$ (5 ms) and from equation (15.31) t_x = 0.93ms

i. The peak switch current is 18.9 A.
From equation (15.32) the average switch current, using alternating zero volt loops, is

$$\bar{I}_T = \frac{1}{20\text{ms}}\int_{0.93\text{ms}}^{5\text{ms}} (34 - 41e^{-200t})\, dt + \frac{1}{40\text{ms}}\int_{0}^{5\text{ms}} 19e^{-200t} dt$$
$$= 2.18 + 1.50 = 3.68 \text{ A}$$

ii. The peak diode current (and peak switch current) is 18.9 A. The average diode current, from equation (15.33), when using alternating zero volt loops, is given by

$$\bar{I}_D = \frac{1}{20\text{ms}}\int_{0}^{0.93\text{ms}} \left(-34 + 41e^{-200t}\right) dt + \frac{1}{40\text{ms}}\int_{0}^{5\text{ms}} 19e^{-200t} dt$$
$$= 0.16 + 1.50 = 1.66 \text{ A}$$

iii. The maximum blocking voltage of each device type is 340 V.

iv. The average supply current is

$$\bar{I}_s = 2\left(\bar{I}_T - \bar{I}_D\right) = 2\times(3.68\text{A} - 1.66\text{A}) = 4.04\text{A}$$

This results in the supply delivery power of 340Vdc × 4.04A = 1374W

v. The harmonics are given by equations (15.1) to (15.3)

$$hf = \rho_3 = \left|\frac{V_3}{V_1}\right| = \tfrac{1}{3}\sqrt{2}\Big/\tfrac{1}{\sqrt{2}} = \tfrac{1}{3}, \text{ that is, } 33\tfrac{1}{3} \text{ per cent}$$

$$df = \mu_3 = \left|\frac{V_3}{nV_1}\right| = \rho_3/n = \tfrac{1}{9}, \text{ that is, } 11.11 \text{ per cent}$$

vi.

$$thd = \sqrt{\left[\sum_{n\geq 2}^{\infty}\left(\frac{V_n}{n}\right)^2\right]}\Big/ V_1$$
$$= \sqrt{\left(\frac{1}{3}\right)^2 + \left(\frac{-1}{5}\right)^2 + \left(\frac{1}{7}\right)^2 + \left(\frac{1}{9}\right)^2 + \ldots} \quad = 46.2 \text{ per cent}$$

♣

Example 15.3: ***Harmonic analysis of H-bridge inverter with an L-R load***

For each delay case (α = 0° and α = 90°) in example 15.1, using Fourier voltage analysis, determine (ignore harmonics above the 10th):

i. the magnitude of the fundamental and first four harmonics
ii. the load rms voltage and current
iii. load power
iv. load power factor

Solution

The appropriate harmonic analysis is outline in the following table, for α = 0° and α = 90°.

n	Z_n	V_n (α=0)	I_n (α=0)	V_n (α=90°)	I_n (α=90°)
harmonic	$\sqrt{R^2 + (2\pi 50nL)^2}$	$\frac{0.9V_s}{n}$	V_n/Z_n	$\frac{0.9V_s}{n}\cos(\frac{1}{2}n\alpha)$	V_n/Z_n
	Ω	V	A	V	A
1	18.62	306	16.43	216.37	11.62
3	48.17	102	2.12	-72.12	-1.50
5	79.17	61.2	0.77	-43.28	-0.55
7	110.41	43.71	0.40	30.91	0.28
9	141.72	34	0.24	24.04	0.17
		332.95V	16.59A	235.43V	11.73A

i. The magnitude of the fundamental voltage is 306V for the square wave and is reduced to 216V when a phase delay angle of 90° is introduced. The table shows that the harmonics magnitudes reduce $(1/n)$ as the harmonic order increases.

ii. The rms load current and voltage can be derived by the square root of the sum of the squares of the fundamental and harmonic components, that is, for the current

$$i_{rms} = \sqrt{I_1^2 + I_3^2 + I_5^2 + \ldots..}$$

The load rms currents, from the table, are 16.59A and 11.73A, which agree with the values obtained in example 15.1a. Notice that the predicted rms voltages of 333V and 235V differ significantly from the values in example 15.1a, given by $V_s\sqrt{1 - 9/\pi}$, namely 340V and 240.4V respectively. This is because the magnitude of the harmonics higher in order than 10 are not insignificant. The error introduced into the rms current value by ignoring these higher order voltages is insignificant because the impedance increases approximately proportionally with harmonic number, hence the resultant current becomes much smaller (insignificant) as the order increases.

iii. The load power is the load i^2R loss, that is

$$P_L = i_{rms}^2 R = 16.59^2 \times 10\Omega = 2752\text{W} \text{ for } \alpha = 0$$
$$P_L = i_{rms}^2 R = 11.73^2 \times 10\Omega = 1376\text{W} \text{ for } \alpha = 90°$$

iv. The load power factor is the ratio of real power dissipated to apparent power, that is

$$pf = \frac{P}{S} = \frac{i_{rms}^2 R}{i_{rms} v_{rms}} = \frac{2752\text{W}}{16.59\text{A}\times 340\text{V}} = 0.488 \text{ for } \alpha = 0$$

$$pf = \frac{P}{S} = \frac{i_{rms}^2 R}{i_{rms} v_{rms}} = \frac{1376\text{W}}{11.79\text{A}\times 240.4\text{V}} = 0.486 \text{ for } \alpha = 90°$$

Equations (15.23) and (15.44) confirm the load power factor is 0.488, independent of α.

♣

Example 15.4: ***Single-phase half-bridge inverter with an L-R load***

A single-phase half-bridge inverter as shown in the figure 15.3, supplies a 10 Ω resistance with inductance 50 mH from a 340 V dc source. If the bridge is operating at 50 Hz, determine for the square-wave output

i. steady-state current waveforms
ii. the load rms voltage
iii. the peak load current and its time domain solution, $i_L(t)$
iv. the average and peak current in the switches
v. the average and peak current in the diodes
vi. the peak blocking voltage of each semiconductor type
vii. the power delivered to the load, rms load current, and average supply current

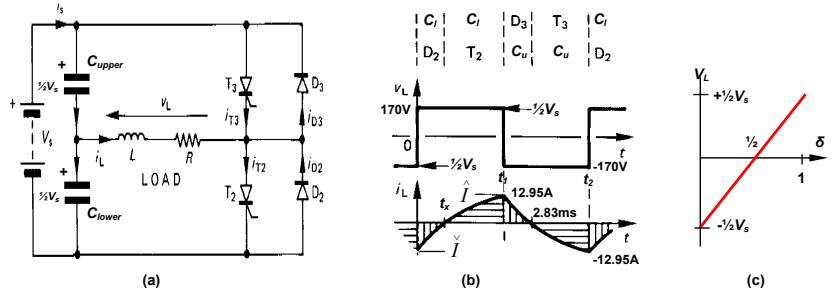


Figure 15.3. *GCT thyristor single-phase half-bridge inverter:*
(a) circuit diagram; (b) square-wave output voltage; and (c) output voltage transfer function.

Solution

From examples 15.1 and 15.2, τ = 5ms.

i. Figure 15.3 shows the output voltage and current waveforms, with various circuit component current waveforms superimposed. Note that no zero voltage loops can be created with the half-bridge. Only load voltages ±½V_s , that is ±170V dc, are possible.

ii. The output voltage swing is ±½V_s, ±170V, thus the rms output voltage is ½V_s, 170V. This is, half that of the full-bridge inverter using the same magnitude source voltage V_s, 340V dc.

iii. The peak load current is half that given by equation (15.9), that is

$$\hat{I} = \frac{½V_s}{R}\frac{1-e^{\frac{-t_1}{\tau}}}{1+e^{\frac{-t_1}{\tau}}} = \frac{½V_s}{R}\tanh\left(\frac{t_1}{2\tau}\right)$$

$$= \frac{½\times 340\text{V}}{10\Omega}\times\tanh\left(\frac{10\text{ms}}{2\times 5\text{ms}}\right) = 12.95\text{A}$$

The load current waveform is defined by equations (15.7) and (15.8), specifically

$$i_{L_I}(t) = \frac{½V_s}{R} - \left(\frac{½V_s}{R} - \check{I}\right)\times e^{\frac{-t}{\tau}}$$

$$= \frac{½\times 340\text{V}}{10\Omega} - \left(\frac{½\times 340\text{V}}{10\Omega} + 12.95\text{A}\right)\times e^{\frac{-t}{5\text{ms}}}$$

$$= 17 - 29.95\, e^{\frac{-t}{5\text{ms}}} \qquad \text{for} \qquad 0 \le t \le 10\text{ms}$$

and

$$i_{L_{II}}(t) = -\frac{½V_s}{R} + \left(\frac{½V_s}{R} + \hat{I}\right)\times e^{\frac{-t}{\tau}}$$

$$= -\frac{½\times 340\text{V}}{10\Omega} + \left(\frac{½\times 340\text{V}}{10\Omega} + 12.95\right)\times e^{\frac{-t}{5\text{ms}}}$$

$$= -17 + 29.95\, e^{\frac{-t}{5\text{ms}}} \qquad \text{for} \qquad 0 \le t \le 10\text{ms}$$

By halving the effective supply voltage, the current swing is also halved.

iv. The peak switch current is $\hat{I} = 12.95\text{A}$.
The average switch current is given by

$$\bar{I}_T = \frac{1}{20\text{ms}}\int_{2.83\text{ms}}^{10\text{ms}} (17 - 29.95e^{\frac{-t}{5\text{ms}}})\, dt$$

$$= 2.86\text{ A}$$

v. The peak diode current is $\hat{I} = 12.95\text{A}$.
The average diode current is given by

$$\bar{I}_D = \frac{1}{20\text{ms}}\int_{0}^{2.83\text{ms}} \left(17 - 29.95e^{\frac{-t}{5\text{ms}}}\right) dt$$

$$= 0.83\text{ A}$$

vi. When a switch or diode of a parallel pair conduct, the complementary pair of devices experience a voltage V_s, 340V dc. Thus although the load experiences half the supply voltage, the semiconductors experience twice that voltage, the same voltage experienced by the switches in the full bridge inverter.

vii. The load power (whence various currents) is found by averaging the instantaneous load power

$$P_L = \frac{1}{10\text{ms}}\int_0^{10ms} 170\text{V}\times\left(17 - 29.95\times e^{-200t}\right) dt \qquad i_{rms} = \sqrt{P_L/R} \qquad \bar{I}_s = P_L/V_s$$

$$= 638.5\text{ W} \qquad = \sqrt{638.5\text{W}/10\Omega} = 8\text{A} \qquad = 638.5\text{W}/340\text{V} = 1.88\text{A}$$

♣

15.1.1iii - PWM-wave output

The output voltage and frequency of a single-phase voltage-source inverter bridge can be control using one of two forms of pulse-width modulation, termed:

- bipolar
- multi-level, usually (meaninglessly) called unipolar

Both pwm techniques have been analysed extensively for dc voltage outputs when applied to the two quadrant and four quadrant dc choppers considered in Chapter 14, sections 14.5 and 14.6. It will be seen that the same triangular modulation principles can be applied and extended, when producing low-harmonic single-phase ac output voltages and currents. The main voltage output difference between the two methods is the harmonic content near the carrier frequency and its harmonics. Three-phase pwm is a naturally extension to the single-phase case, except single-phase pwm offers more degrees of flexibility than its application to three phase inverters, although three-phase pwm does have the attribute of triplen harmonic cancellation, due to the use of one (co-phasal) triangular carrier.

Bipolar pulse width modulation

Bipolar modulation is the simplest pwm method and involves comparing a fixed frequency and magnitude triangular carrier with the ac waveform desired, called the modulation waveform. The modulation waveform is usually a sinusoid of magnitude (modulation index) M such that $0 \le M \le 1$.
The waveforms in figure 15.4 shown that the load voltage V_L swings between the two voltage levels, $+V_s$ and $-V_s$, (hence the term bipolar output voltage), according to

- T_1 and T_2 are on when $v_{ref} > v_\Delta$ (T_3 and T_4 are off) such that $V_L = +V_s$
- T_3 and T_4 are on when $v_{ref} < v_\Delta$ (T_1 and T_2 are off) such that $V_L = -V_s$

Multi-level pulse width modulation

Two multilevel output voltage techniques can be use with single-phase voltage fed ac bridges. In both case, two triangular carries displaced by 180° give the same output for the same switching frequency.

i. The waveforms in figure 15.5 show that the load voltage V_L swings between the two voltage levels, $+V_s$ and $-V_s$, with interspaced zero periods (hence the term multilevel, specifically three-level in this case, 0V and ±V_s), according to

- T_1 is on when $v_{ref} > v_\Delta$ such that $V_{ao} = +V_s$
- T_4 is on when $v_{ref} < v_\Delta$ such that $V_{ao} = 0V$

- T_3 is on when $v_{ref} < -v_\Delta$ such that $V_{bo} = V_s$
- T_2 is on when $v_{ref} > -v_\Delta$ such that $V_{bo} = 0V$

The multilevel load output voltage is the difference between the two leg voltage waveforms and can be defines as follows:

- T_1 and T_2 are on such that $V_{ao} = +V_s$, $V_{bo} = 0V$, $V_{ab} = +V_s$
- T_2 and T_3 are on such that $V_{ao} = 0V$, $V_{bo} = +V_s$, $V_{ab} = -V_s$

- T_1 and T_3 are on such that $V_{ao} = +V_s$, $V_{bo} = +V_s$, $V_{ab} = 0V$
- T_2 and T_4 are on such that $V_{ao} = 0V$, $V_{bo} = 0V$, $V_{ab} = 0V$

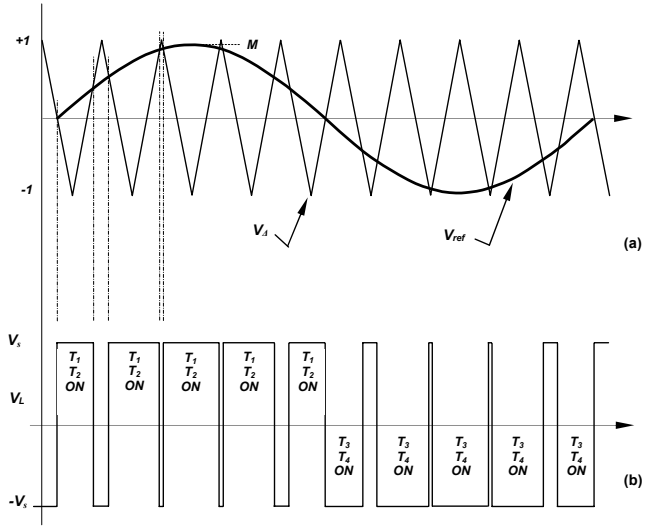


Figure 15.4. *Bipolar pulse width modulation:*
(a) carrier and modulation waveforms and (b) resultant load pwm waveform.

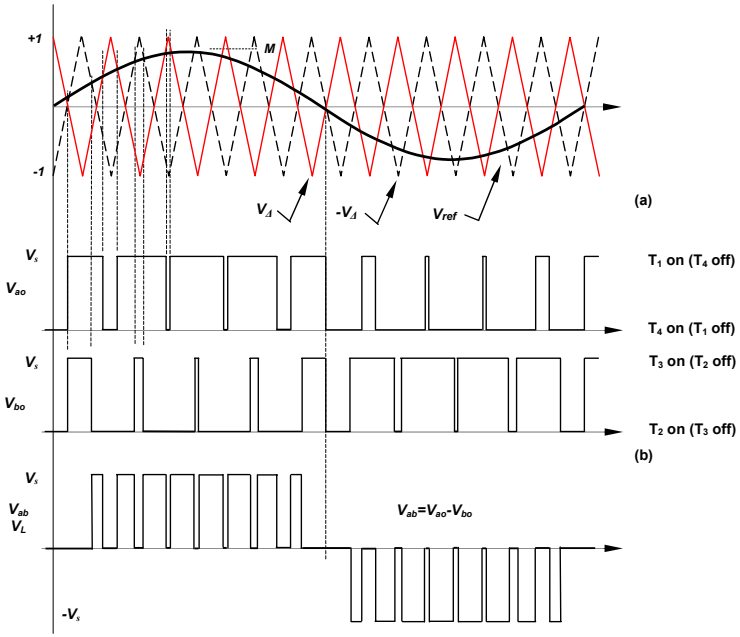


Figure 15.5. *Multilevel (3 level) pulse width modulation:*
(a) carriers and modulation waveforms and (b) resultant load pwm waveforms.

The two zero output states are interleaved to balance switching losses between all four bridge switches. Device switching is at the carrier frequency, but the bridge load voltage (hence load current) experiences twice the leg switching frequency since the two carriers are displaced by 180°.

ii. A second multilevel output voltage approach is shown in figure 15.16, where the triangular carriers are not only displaced by 180° in time, but are vertically displaced, as for multilevel inverter pwm generation, which is considered in section 15.4 (half the magnitude and twice the frequency as in figure 15.6). The upper triangle modulates reference values greater than zero, while the lower triangle modulates when the reference is less than zero.

Spectral comparison between bipolar and multilevel pwm waveforms

The key features of the H-bridge inverter output voltage with bipolar pwm are (fig 15.6a):

- a triangular carrier has only odd Fourier components, so the output spectrum only has carrier components at odd harmonics of the carrier frequency
- the first carrier components occur at the carrier frequency, $\boldsymbol{f_c}$
- side-band components occur spaced by $\boldsymbol{2f_o}$ from other components, around all multiples of the carrier frequency $\boldsymbol{f_c}$

From figure 15.6b, the key features of the H-bridge inverter output voltage with multilevel pwm are:

- the output switching frequency is double $\boldsymbol{2f_c}$ each leg switching frequency $\boldsymbol{f_c}$, since the switching of each leg is time shifted (by 180°), hence the first carrier related components in the output occur at $\boldsymbol{2f_c}$ and then at multiples of $\boldsymbol{2f_c}$ (effectively the carrier is $2f_o$)
- no triangular carrier Fourier components exist in the output voltage since the two carriers are in anti-phase (180° apart), effectively cancelling one another in spectrum terms
- side-band components occur spaced by $\boldsymbol{2f_o}$ from other components, around each multiple of the carrier frequency $\boldsymbol{2f_c}$

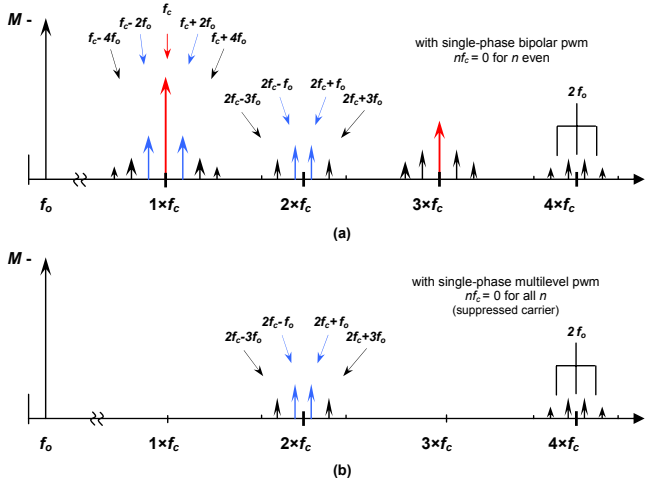


Figure 15.6. *Typical phase output frequency spectrum, at a give switch commutation frequency, for:*
(a) bipolar pwm and (b) multilevel pwm.

15.1.2 Three-phase voltage-source inverter bridge

The basic dc to three-phase voltage-source inverter (VSI) bridge is shown in figure 15.7. It comprises six power switches together with six associated reactive energy feedback diodes. Each of the three inverter legs operates at a relative time displacement (phase) of ⅔π, 120°.

Table 15.1: Quasi-square-wave six conduction states - *180° conduction*

Interval	***Three** conducting switches*								*leg state*	*voltage vector*
1	T_1	T_2	T_3						101	v_5
2		T_2	T_3	T_4					001	v_1
3			T_3	T_4	T_5				011	v_3
4				T_4	T_5	T_6			010	v_2
5					T_5	T_6	T_1		110	v_6
6						T_6	T_1	T_2	100	v_4

15.1.2i - 180° (π) conduction

Figure 15.8 shows inverter bridge quasi-square output voltage waveforms for a 180° switch conduction pattern. Each switch conducts for 180°, such that no two series connected (leg or arm) semiconductor switches across the voltage rail conduct simultaneously. Six patterns exist for one output cycle and the rate of sequencing these patterns, $6f_o$, specifies the bridge output frequency, f_o. The conducting switches during the six distinct intervals are shown and can be summarised as in Table 15.1.

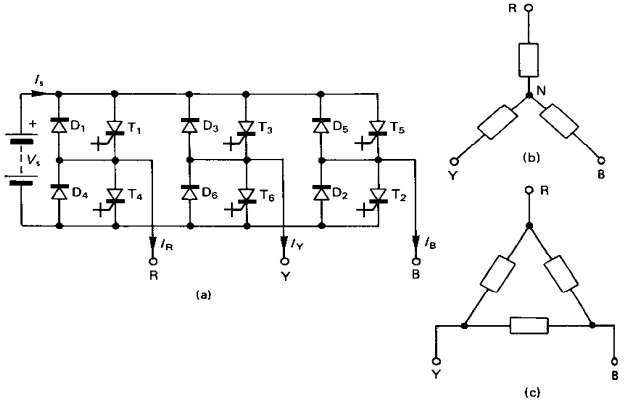


Figure 15.7. *Three-phase VSI inverter circuit:*
(a) GCT thyristor bridge inverter; (b) star-type load; and (c) delta-type load.

The three output voltage waveforms can be derived by analysing a balanced resistive star load and considering each of the six connection patterns, as shown in figure 15.9, using the maxtrix in figure 15.8c. Effectively the resistors representing the three-phase load are sequentially cycled anticlockwise one at a time, being alternately connected to each supply rail. The output voltage is independent of the load, as it is for all voltage source inverters.

Alternatively, the generation of the three-phase voltages can be analysed analytically by using the rotating *voltage space vector* technique. With this approach, the output voltage state from each of the three inverter legs (or poles) is encoded as summarised in table 15.1, where a '1' signifies the upper switch in the leg is on, while a '0' means the lower switch is on in that leg. The resultant binary number (one bit for each of the three inverter legs), represents the output voltage vector number (when converted to decimal). The six voltage vectors are shown in figure 15.10 forming sextant boundaries, where the quasi-square output waveform in figure 15.8b is generated by stepping instantaneously from one vector position to another in an anticlockwise direction. Note that the rotational stepping sequence is arranged such that when rotating in either direction, only one leg changes state, that is, one device turns off and then the complementary switch of that leg turns on, at each step. This minimises the inverter switching losses. The dwell time of the created rotating vector at each of the six vector positions, is ⅓π ($\frac{1}{6}T$) of the cycle period (T). Note that the line-to-line zero voltage states 000 and 111 are not used. These represent the condition when ***either*** all the upper switches (T_1, T_3, T_5) are on ***or*** all the lower switches (T_2, T_4, T_5) are switched on (represented as the origin in figure 15.10). Phase reversal can be obtained by interchanging two phase outputs, or as is the preferred method, the direction of the rotating vector sequence is reversed. Reversing is therefore effectively achieved by back-tracking along each output waveform.

With reference to figure 15.8b, the line-to-load neutral voltage Fourier coefficients are given by

$$V_{n_{L-N}} = \frac{2}{3\pi} V_s \frac{\left(2 + \cos\frac{n\pi}{3} - \cos\frac{2n\pi}{3}\right)}{n} \qquad (15.45)$$

The line-to-load neutral voltage is therefore

$$V_{n_{L-N}} = \frac{2}{\pi} V_s \sum_{n=1,6r\pm1}^{\infty} \frac{\sin n\,\omega t}{n} \qquad r = 1, 2, 3, .. \qquad (15.46)$$

that is

$$v_{RN} = \frac{2}{\pi} V_s \left[\sin\omega t + \tfrac{1}{5}\sin 5\omega t + \tfrac{1}{7}\sin 7\omega t + \tfrac{1}{11}\sin 11\omega t + \ldots\right] \quad \text{(V)} \qquad (15.47)$$

similarly for v_{YN} and v_{BN}, where ωt is substituted by ωt+⅔π and ωt-⅔π respectively.

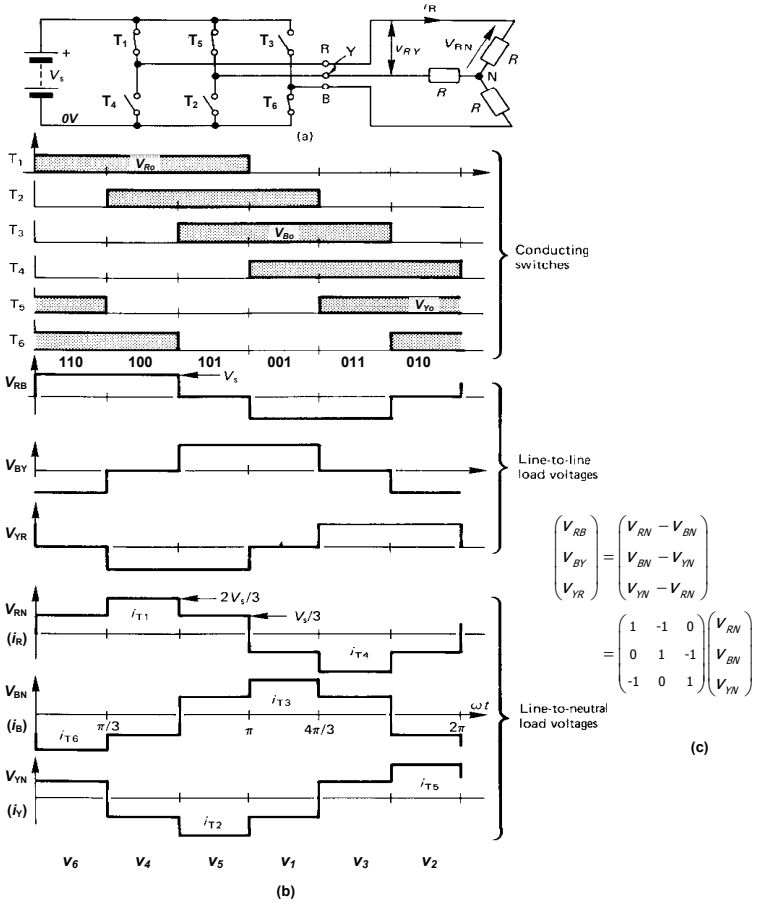


Figure 15.8. *A three-phase bridge inverter employing 180° switch conduction with a resistive load: (a) the bridge circuit showing T_1, T_5, and T_6 conducting (leg state **v_6 :– 110**); (b) circuit voltage and current waveforms with each of six sequential output voltage vectors identified; and (c) phase voltage to line voltage conversion matrix.*

The line-to-line voltage, from equation (15.38) with $\alpha = \frac{1}{3}\pi$, gives Fourier coefficients defined by

$$V_{n_{L-L}} = \frac{4}{\pi} V_s \frac{\left(\cos\frac{n\pi}{6}\right)}{n} \qquad (15.48)$$

The line-to-line voltage is thus

$$V_{n_{L-L}} = \frac{2\sqrt{3}}{\pi} V_s \sum_{n=1,6r\pm1}^{\infty} \left\|\cos\frac{n\pi}{6}\right\| \frac{\sin n\omega t}{n} \qquad r = 1, 2, 3, . \qquad (15.49)$$

(the $\| \ \|$ symbol provides the sign), that is

$$v_{RB} = \frac{2\sqrt{3}}{\pi} V_s \left[\sin\omega t - \tfrac{1}{5}\sin 5\omega t - \tfrac{1}{7}\sin 7\omega t + \tfrac{1}{11}\sin 11\omega t + \ldots\right] \quad (\text{V}) \qquad (15.50)$$

and similarly for v_{BY} and v_{YR}. Figure 15.8b shows that v_{RB} is shifted $\frac{1}{6}\pi$ with respect to v_{RN}, hence to obtain the three line voltages while maintaining a v_{RN} reference, ωt should be substituted with $\omega t + \frac{1}{6}\pi$, $\omega t - \frac{1}{2}\pi$ and $\omega t + \frac{5}{6}\pi$, respectively.
Since the interphase voltages consist of two square waves displaced by $\frac{2}{3}\pi$, no triplen harmonics (3, 6, 9, . . .) exist. The outputs comprise harmonics given by the series $n = 6r \pm 1$ where $r \geq 0$ and is an integer. The nth harmonic has a magnitude of $1/n$ relative to the fundamental.

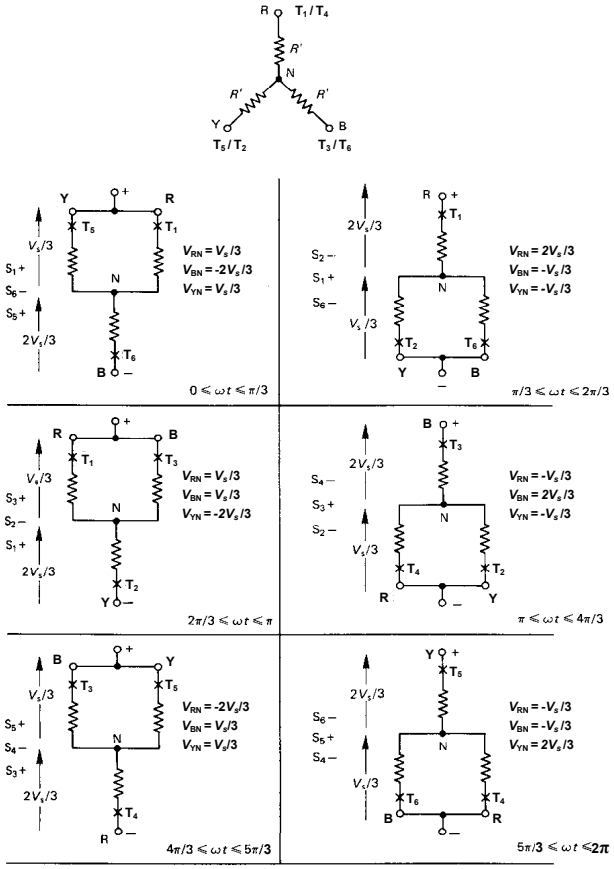


Figure 15.9. *Determination of the line-to-neutral voltage waveforms for a balanced resistive load and 180° conduction as illustrated in figure 15.8.*

By examination of the interphase output voltages in figure 15.8 it can be established that the mean half-cycle voltage is $\frac{2}{3}V_s$ and the rms value is $\sqrt{\frac{2}{3}}\,V_s$, namely 0.816 V_s. From equation (15.50) the rms value of the fundamental is $\sqrt{6}\,V_s/\pi$, namely 0.78 V_s, that is 3/π times the total rms voltage value.
The three-phase inverter output voltage properties are summarised in Table 15.2.

15.1.2ii - 120° (⅔π) conduction

The basic three-phase inverter bridge in figure 15.7 can be controlled with each switch conducting for 120°. As a result, at any instant only two switches (one upper and one non-complementary lower) conduct and the resultant quasi-square output voltage waveforms are shown in figure 15.11. A 60° (⅓π), dead time exists between two series switches conducting, thereby providing a safety margin against *simultaneous conduction* of the two series devices (for example T_1 and T_4) across the dc supply rail. This safety margin is obtained at the expense of a lower semi-conductor device utilisation and rms output voltage than with 180° device conduction. The device conduction pattern is summarised in Table 15.3. A feature with ⅔π conduction is that the phase currents can be measured from the dc link current.

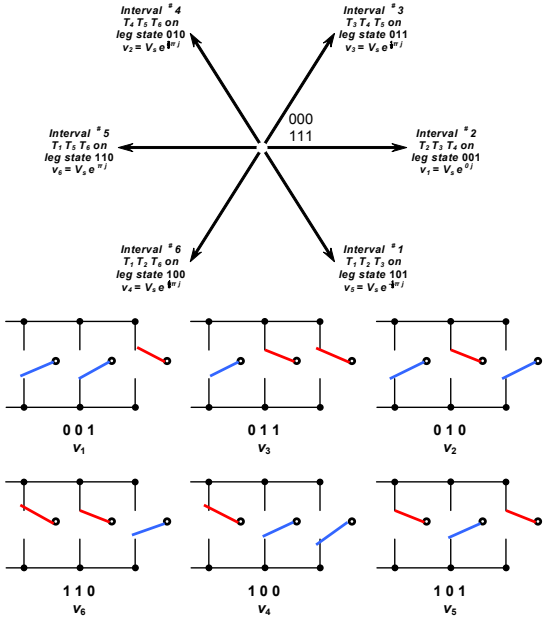


Figure 15.10. *Generation and arrangement of the six quasi-square inverter output voltage states.*

Figure 15.8b for 180° conduction and 15.11b for 120° conduction show that the line to neutral voltage of one conduction pattern is proportional to the line-to-line voltage of the other. That is, from equation (15.38) with $\alpha = \frac{1}{3}\pi$

$$\begin{aligned} v_{RN}\left(\tfrac{1}{3}\pi\right) &= \tfrac{1}{2}v_{RY}\left(\pi\right) = \sum_{n=1,3,5}^{\infty} \frac{2}{\pi n} V_s \cos\frac{n\pi}{6}\sin n\omega t \\ &= \frac{\sqrt{3}}{\pi} V_s \left[\sin\omega t - \tfrac{1}{5}\sin 5\omega t - \tfrac{1}{7}\sin 7\omega t + \tfrac{1}{11}\sin 11\omega t + \ldots\right] \quad (\text{V}) \end{aligned} \qquad (15.51)$$

and

$$\begin{aligned} v_{RY}\left(\tfrac{2}{3}\pi\right) &= \tfrac{3}{2}v_{RN}\left(\pi\right) = \sum_{n=1,3,5}^{\infty} \frac{2\sqrt{3}}{\pi n} V_s \cos\frac{n\pi}{6}\sin n\omega t \\ &= \frac{3}{\pi} V_s \left[\sin\omega t + \tfrac{1}{5}\sin 5\omega t + \tfrac{1}{7}\sin 7\omega t + \tfrac{1}{11}\sin 11\omega t + \ldots\right] \quad (\text{V}) \end{aligned} \qquad (15.52)$$

Also $v_{RY} = \sqrt{3}\,v_{RN}$ and the phase relationship between these line and phase voltages, of $\frac{1}{6}\pi$, has not been retained. That is, with respect to figure 15.11b, substitute ωt with $\omega t + \frac{1}{6}\pi$ in equation (15.51) and $\omega t + \frac{1}{3}\pi$ in equation (15.52).

The output voltage properties for both 120° and 180° conduction are summarised in the Table 15.2.

Independent of the conduction angle (120°, 180° or even 150°), quasi-square 180° conduction occurs with inductive loads, producing the six hexagon states shown in the upper part of figure 15.10. The resistive load assumptions made in this section for explanation purposes can be misleading.

Table 15.2: Quasi-squarewave voltage properties for a resistive load

Conduction period	Fundamental voltage		Characteristic		
	peak	rms	Total rms	Distortion Factor	THD
	$\hat{V}_1$	V_1	V_{rms}	μ	thd
180°	**(V)**	**(V)**	**(V)**		
Phase Voltage $V_{L\text{-}N}$	$\frac{2}{\pi}V_s = 0.637V_s$	$\frac{\sqrt{2}}{\pi}V_s = 0.450V_s$	$\frac{\sqrt{2}}{3}V_s = 0.471V_s$	$\frac{3}{\pi} = 0.955$	$\sqrt{\frac{\pi^2}{9}-1} = 0.311$
Line Voltage $V_{L\text{-}L}$	$\frac{2\sqrt{3}}{\pi}V_s = 1.10V_s$	$\frac{\sqrt{6}}{\pi}V_s = 0.78V_s$	$\sqrt{\frac{2}{3}}V_s = 0.816V_s$	$\frac{3}{\pi} = 0.955$	$\sqrt{\frac{\pi^2}{9}-1} = 0.311$
120°	**(V)**	**(V)**	**(V)**		
Phase Voltage $V_{L\text{-}N}$	$\frac{\sqrt{3}}{\pi}V_s = 0.551V_s$	$\frac{\sqrt{6}}{2\pi}V_s = 0.390V_s$	$\frac{1}{\sqrt{6}}V_s = 0.408V_s$	$\frac{3}{\pi} = 0.955$	$\sqrt{\frac{\pi^2}{9}-1} = 0.311$
Line Voltage $V_{L\text{-}L}$	$\frac{3}{\pi}V_s = 0.955V_s$	$\frac{3}{\sqrt{2}\pi}V_s = 0.673V_s$	$\frac{1}{\sqrt{2}}V_s = 0.707V_s$	$\frac{3}{\pi} = 0.955$	$\sqrt{\frac{\pi^2}{9}-1} = 0.311$

Table 15.3: Quasi-squarewave conduction states - *120° conduction*

Interval	***Two** conducting devices*						
1	T_1	T_2					
2		T_2	T_3				
3			T_3	T_4			
4				T_4	T_5		
5					T_5	T_6	
6						T_6	T_1

15.1.3 Inverter ac output voltage and frequency control techniques

It is a common requirement that the output voltage and/or frequency of an inverter be varied in order to control the load power or, in the case of an induction motor, to control the shaft speed and torque by maintaining a constant *V/f* ratio. The six VSI modulation control techniques to be considered are:

- Variable voltage dc link
- Single-pulse width modulation
- Multi-pulse width modulation
- Multi-pulse, selected notching modulation
- Sinusoidal pulse width modulation
- Triplen injection
  - Triplens injected into the modulation waveform
  - Voltage space vector modulation
- Selected harmonic elimination

15.1.3i - Variable voltage dc link

The rms voltage of a square-wave can be changed and controlled by varying the dc link source voltage. A *variable dc link* voltage can be achieved with a dc chopper as considered in chapter 14 or an ac phase-controlled thyristor bridge as considered in sections 12.2 and 12.4. A dc link *L-C* smoothing filter may be necessary.

15.1.3ii - Single-pulse width modulation

Simple pulse-width control can be employed as considered in section 15.1.1b, where a single-phase bridge is used to produce a quasi-square-wave output voltage as shown in figure 15.1c.
An alternative method of producing a quasi-square wave of controllable pulse width is to transformer-add the square-wave outputs from two push-pull bridge inverters as shown in figure 15.12a. By phase-shifting the output by α, a quasi-square sum results as shown in figure 15.12b.

The output voltage can be described by

$$V_o = \sum_{n\ odd}^{\infty} v_{an} \sin n\omega t \qquad \text{(V)} \tag{15.53}$$

where

$$v_{an} = \frac{2}{\pi}\int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi} V_s \cos n\alpha\, d\alpha = \frac{4}{n\pi}V_s \cos(\tfrac{1}{2}n\alpha) \qquad \text{(V)} \tag{15.54}$$

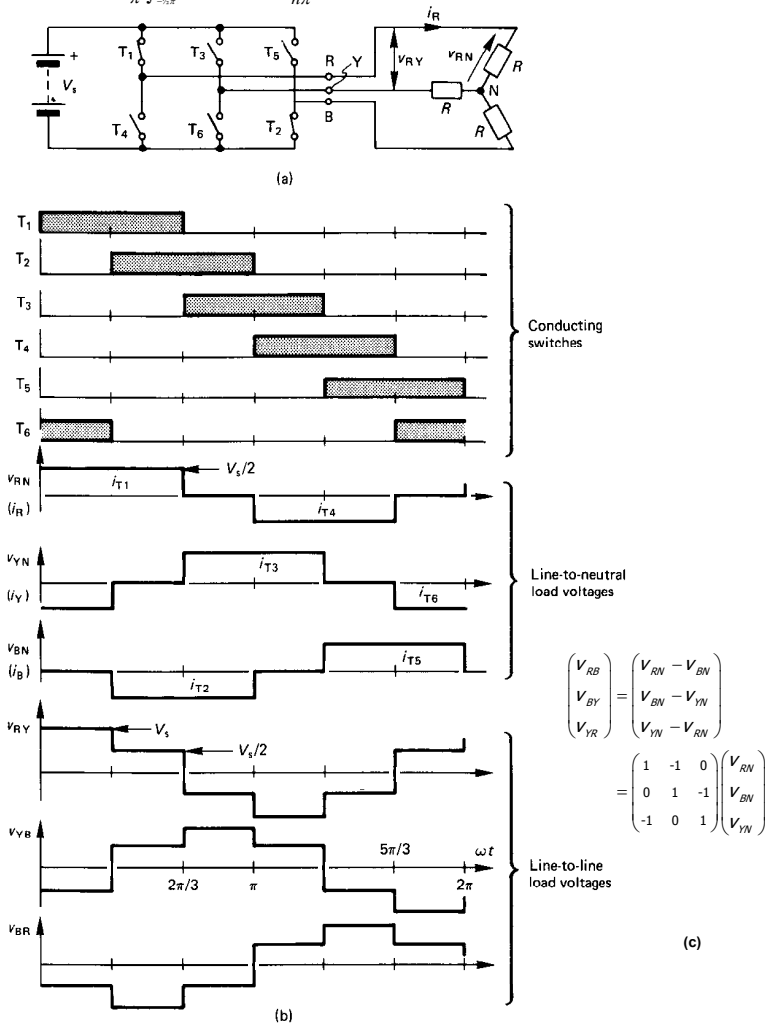


Figure 15.11. *A three-phase bridge inverter employing 120° switch conduction with a resistive star load: (a) the bridge circuit showing T_1 and T_2 conducting; (b) circuit voltage and current waveforms; and (c) phase voltage to line voltage conversion matrix.*

The rms output voltage is

$$V_r = V_s\sqrt{1-\alpha/\pi} \quad \text{(V)} \qquad (15.55)$$

and the rms value of the fundamental is

$$V_1 = \frac{2\sqrt{2}}{\pi} V_s \cos\tfrac{1}{2}\alpha \quad \text{(V)} \qquad (15.56)$$

As α increases, the magnitude of the harmonics, particularly the third, becomes significant compared with the fundamental magnitude. This type of control may be used in high power applications.

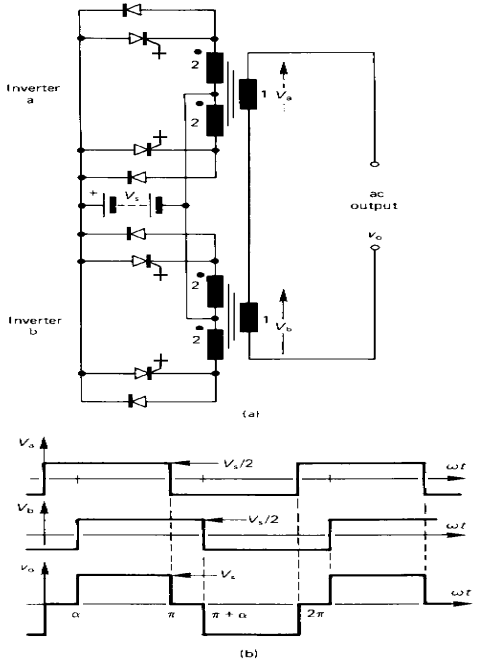


Figure 15.12. *Voltage control by combining phase-shifted push-pull inverters: (a) two inverters with two transformers for summing and (b) circuit voltage waveforms for a phase displacement of α.*

Example 15.5: ***Single-pulse width modulation***

Two single-phase H-bridge inverter outputs are transformer added, as shown in figure 15.12. Each inverter operates at 50Hz but phase shifted so as to produce 240V rms fundamental output when the rail voltage of each inverter is 340V dc and the transformers turns ratios are 2:2:1.
Determine

i. the phase shift between the two single phase inverters
ii. the rms output voltage
iii. the frequency and magnitude of the first 4 harmonics of 50Hz and their rms ac contribution to the rms output
iv. rms voltage of higher order harmonics (higher frequencies than those in part iii.)
v. the total harmonic distortion of the output voltage.

Solution

i. The output is a quasi-square waveform of magnitude ±340V dc. The magnitude of the 50Hz fundamental is given by equation (15.54), for n =1:

$$v_{a1} = \frac{4}{\pi} V_s \cos(\tfrac{1}{2}\alpha)$$

$$\sqrt{2}\ 240\text{V} = \frac{4}{\pi} \times 340\text{V} \times \cos(\tfrac{1}{2}\alpha)$$

from which the phase shift is 76.7°, 1.34 radians.

ii. The rms output voltage is given by equation (15.55), that is

$$V_{rms} = V_s\sqrt{1-\alpha/\pi} = 340\text{V}\sqrt{1-1.34/\pi} = 257.5\text{V}$$

iii. The peak values of the first four harmonics are given in the table below.

harmonic n	$v_{an} = \frac{4}{n\pi} V_s \cos(\tfrac{1}{2}n\alpha)$	v_{an}^2
3	-61.4	3765.0
5	-84.7	7175.3
7	-1.4	1.9
9	46.6	2168.5
	$\sqrt{\sum v_{an}^2} =$	114.50

The rms value of the ac of the first four harmonics is 114.5/√2 = 81.0V.

iv. The ac component of the harmonics above the 9th is given by

$$V_{rms\ n>9} = \sqrt{V_{rms}^2 - V_{rms\ n\le 9}^2}$$
$$= \sqrt{257.5\text{V}^2 - \left(240\text{V}^2 + 81.0\text{V}^2\right)} = 46.3\text{V}$$

v. The total harmonic voltage distortion is given by

$$THD_v = \frac{\sqrt{V_{rms}^2 - V_{a1}^2}}{V_{a1}} \times 100 = \sqrt{\left(\frac{V_{rms}}{V_{a1}}\right)^2 - 1} \times 100$$
$$= \sqrt{\left(\frac{257.5\text{V}}{240\text{V}}\right)^2 - 1} \times 100 = 38.9\%$$

♣

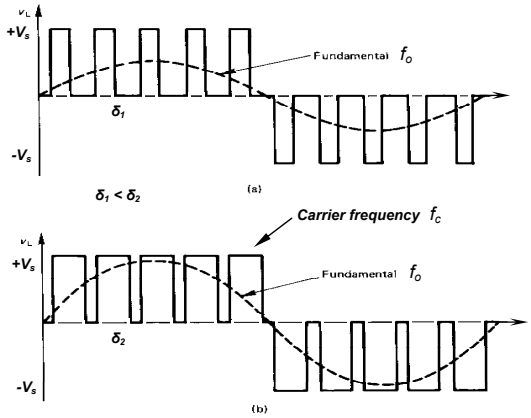


Figure 15.13. *Inverter control giving variable duty cycle of five notches per half cycle: (a) low duty cycle, δ_1, hence low fundamental magnitude and (b) higher duty cycle, δ_2, for a high fundamental voltage output.*

15.1.3iii - Multi-pulse width modulation

An extension of the single-pulse modulation technique is multiple-notching as shown in figure 15.13. The bridge switches are controlled so as to vary the on to off time of each notch, δ, thereby varying the output rms voltage which is given by $V_{rms} = \sqrt{\delta}\, V_s$. Alternatively, the number of notches can be varied.
The harmonic content at lower output voltages is significantly lower than that obtained with single-pulse modulation. The increased switching frequency does increase the magnitude of higher order harmonics and the switching losses. The Fourier coefficients of the output voltage in figure 15.13 are given by

$$V_n = \frac{4}{n\pi} \sum_{j=1,2,3,..}^{f_c/f_o} \left[\cos 2\pi \frac{f_o}{f_c} n(2j-1+\delta) - \cos 2\pi \frac{f_o}{f_c} n(2j-1-\delta) \right] \qquad (15.57)$$

where f_o is the fundamental frequency, f_c the triangular carrier frequency and $0 \le \delta \le 1$ is the duty cycle.

15.1.3iv - Multi-pulse, selected notching modulation – selected harmonic elimination

If a multi-level waveform ($\pm V_s$, 0) is used with quarter wave symmetry, as shown in figure 15.14a, then both the harmonics and total rms output voltage can be controlled. With one pulse per quarter wave, the k^{th} harmonic is eliminated from the output voltage if the centre of the pulse is located such that

$$\sin k\lambda = 0 \qquad (15.58)$$

$$\text{that is } \lambda = \pi/k$$

Independent of the pulse width δ, the k^{th} harmonic is eliminated and the other Fourier components are given by

$$V_n = \frac{8}{n\pi} V_s \sin n\frac{\pi}{k} \sin n\delta \qquad (15.59)$$

The output voltage total rms is solely dependent on the pulse width δ and is given by

$$V_{o\,rms} = V_s \sqrt{\frac{2}{\pi}\delta} \qquad (15.60)$$

On the other hand, the bipolar waveform ($\pm V_s$) in figure 15.14b has an rms value of V_s, independent of the harmonics eliminated.

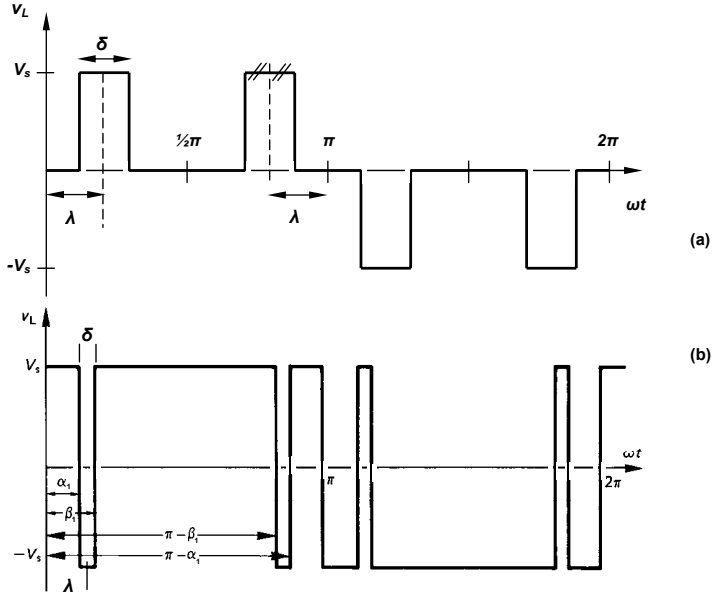


Figure 15.14. *Output voltage harmonic reduction for a single-phase bridge using selected notching: (a) multilevel output voltage and (b) bipolar output voltage.*

Selected elimination of lower-order harmonics can be achieved by producing an output voltage waveform as shown in figure 15.14b. The exact switching points are calculated off-line so as to eliminate the required harmonics. For *n* switchings per half cycle, *n* selected harmonics can be eliminated.
In figure 15.14b two notches per half cycle are introduced; hence any two selected harmonics can be eliminated. The more notches, the lower is the output fundamental. For example, with two notches, the third and fifth harmonics are eliminated. From

$$b_n = \frac{4}{\pi}\int_0^{\frac{1}{2}\pi} f(\theta)\sin n\theta\, d\theta \quad \text{for} \quad n = 1, 2, 3, ... \qquad (15.61)$$

$$b_3 = \frac{4}{3\pi} V_s (1 - 2\cos 3\alpha + 2\cos 3\beta) = 0$$

and

$$b_5 = \frac{4}{5\pi} V_s (1 - 2\cos 5\alpha + 2\cos 5\beta) = 0$$

Solving yields α_1 = 23.6° and β_1 = 33.3°. The total rms output voltage is V_s, independent of the harmonics eliminated. The magnitude (whence rms) of each harmonic component is

$$V_n = \frac{4}{n\pi} V_s (1 - 4 \times \sin n\lambda \times \sin n\delta) \qquad (15.62)$$

The maximum fundamental rms component of the output voltage waveform is 0.84 of a square wave, which is $(2\sqrt{2}/\pi)V_s$ when $\delta = \frac{1}{2}\pi$ which produces a square wave.
Ten switching intervals exist compared with two per cycle for a square-wave, hence switching losses and control circuit complexity are increased.

In the case of a three-phase inverter bridge, the third harmonic does not exist, hence the fifth and seventh (b_5 and b_7) can be eliminated with α_1 = 16.3° and β_1 = 22.1. The 5th, 7th, 11th, and 13th can be eliminated with the angles 10.55°, 16.09°, 30.91°, and 32.87° respectively. Because the waveforms have quarter wave symmetry, only angles for 90° need be stored.

The output rms voltage magnitude can be varied by controlling the dc link voltage or by transformer-adding two phase-displaced bridge outputs as demonstrated in figure 15.12. The output voltage Fourier components in equation (15.62) are modified by equation(15.54) given

$$V_n = \frac{4}{n\pi} V_s (1 - 4 \times \sin n\lambda \times \sin n\delta) \cos \tfrac{1}{2} n\alpha \qquad (15.63)$$

And the total rms output voltage is reduced from V_s , as given by equation (15.55), that is

$$V_{o\,rms} = V_s \sqrt{1 - \alpha/\pi} \quad \text{(V)} \qquad (15.64)$$

Thus the fundamental rms magnitude can be changed by introducing an extra constraint to be satisfied, along with the harmonic eliminating constraints (as a result of the extra constraint, one fewer harmonic can now be eliminated for a given number of switchings per quarter cycle).

The multi-pulse selected notching modulation technique can be extended to the *optimal pulse-width modulation method,* where harmonics may not be eliminated, but minimised according to a specific criterion. In this method, the quarter wave output is considered to have a number of switching angles. These angles are selected so as, for example, to eliminate certain harmonics, minimise the rms of the ripple current, or any other desired performance index. The resultant non-linear equations are solved using numerical methods off-line. The computed angles are then stored in a ROM look-up table for use. A set of angles must be computed and stored for each desired level of the voltage fundamental and output frequency.
The optimal pwm approach is particularly useful for high-power, high-voltage GCT thyristor inverters, which tend to be limited in switching frequency by device switching losses.

Generally *sinωt+ksin3ωt* is generated (since the third harmonic of the square wave is not eliminated)

$$b_n = \frac{4}{\pi}\int_0^{\frac{1}{2}\pi} f(\theta)\sin n\theta\, d\theta = \frac{4}{\pi}\int_0^{\frac{1}{2}\pi} 1 \times \sin n\theta\, d\theta$$

For a fundamental magnitude of √3m of the pu link voltage and to eliminate *N*-1 (*N*-1>2) harmonics:

$$-\sum_{k=1}^{N} (-1)^k \cos\alpha_k = \tfrac{1}{4}\left(2 + (-1)^{N-1} m\pi\right)$$
$$-\sum_{k=1}^{N} (-1)^k \cos n\alpha_k = \tfrac{1}{2} \qquad \text{for } n = 5, 7, ..., 3N - \tfrac{3}{2} - \tfrac{1}{2}(-1)^{N-1} \qquad (15.65)$$

where $\alpha_1 < \alpha_2 < ... < \alpha_N < \frac{1}{2}\pi\left(1 - \frac{1}{6}\left(1 - (-1)^{N-1}\right)\right)$

For *N-1* even, all α_k angles are less than ½π other wise ⅓π (when *N-1* is odd).

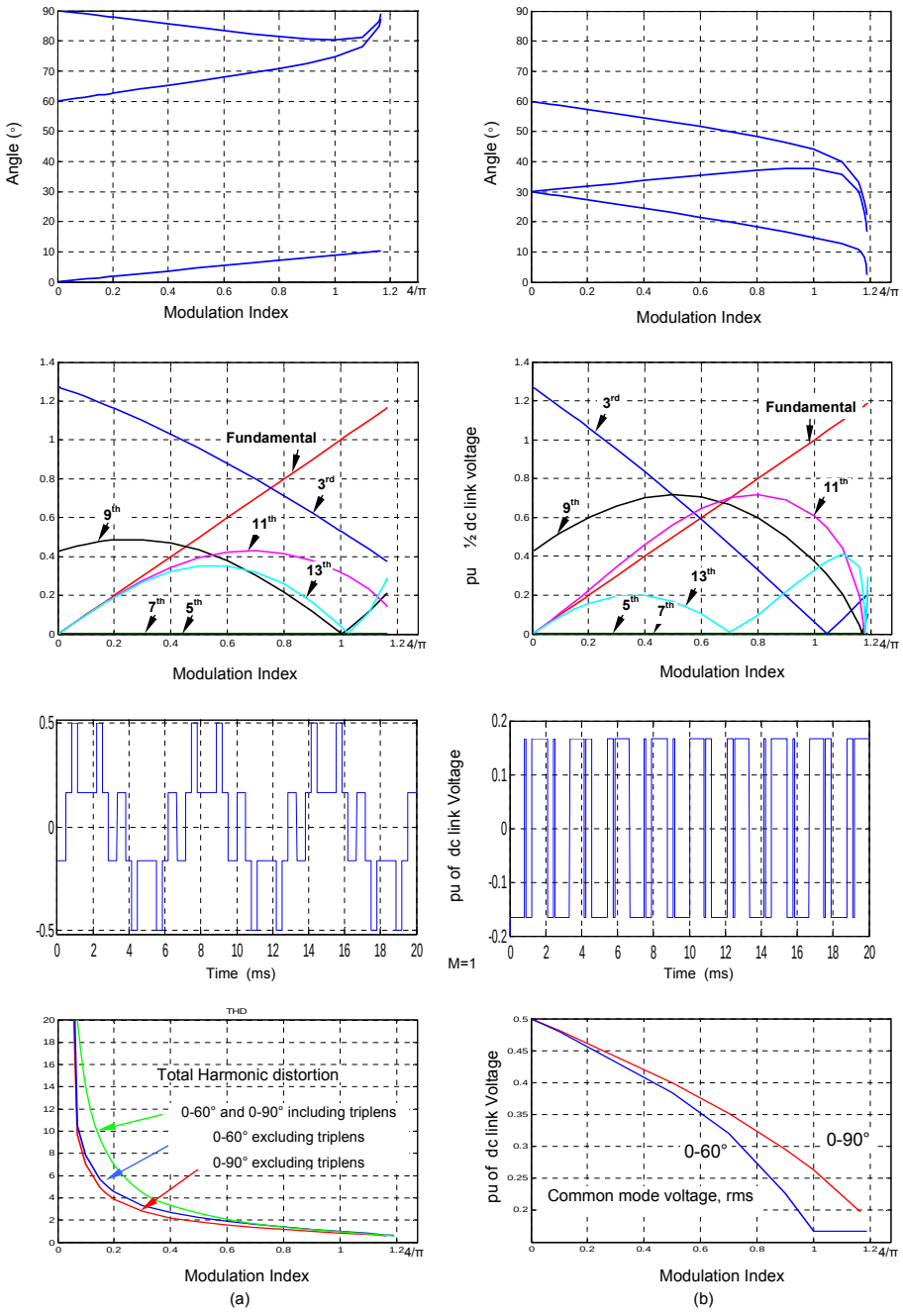


Figure 15.15. *SHE commutation angles, normalised as a function of the fundamental eliminating: the 5th and 7th harmonics with commutation angles (a) 0 to 90° and (b) 0 to 60°.*

To eliminate the 5th and 7th harmonics (an even number of harmonics), with a fundamental magnitude with modulation index *m*, three angles are required, *N*=3 and the three equations to be solved are

$$\cos\alpha_1 - \cos\alpha_2 + \cos\alpha_3 = ¼(2 + m\pi)$$
$$\cos 5\alpha_1 - \cos 5\alpha_2 + \cos 5\alpha_3 = ½$$
$$\cos 7\alpha_1 - \cos 7\alpha_2 + \cos 7\alpha_3 = ½$$

The resultant angles are shown in figure 15.15. The maximum modulation index, with respect to a square wave is 1.166 and 1.188 for angles less that 60 degrees and greater, respectively. Any solution with all angles less than 60° represents dead banding of the three phases, where each phase is alternately clamped to the dc link zero and positive rails (see section 15.1.3vi). The total harmonic distortion is virtually the same in both cases, as is the maximum common mode voltage *dv/dt*, while the rms common mode voltage is greater for the case when the angles can exceed 60 degrees.

To eliminate the 5th, 7th, and 11th harmonics (an odd number of harmonics), with a fundamental magnitude with modulation index *m*, four angles are required, *N*=4 and the four equations to be solved are

$$\cos\alpha_1 - \cos\alpha_2 + \cos\alpha_3 - \cos\alpha_4 = ¼(2 - m\pi)$$
$$\cos 5\alpha_1 - \cos 5\alpha_2 + \cos 5\alpha_3 - \cos 5\alpha_4 = ½$$
$$\cos 7\alpha_1 - \cos 7\alpha_2 + \cos 7\alpha_3 - \cos 7\alpha_4 = ½$$
$$\cos 11\alpha_1 - \cos 11\alpha_2 + \cos 11\alpha_3 - \cos 11\alpha_4 = ½$$

As further harmonics are eliminated, multiple solutions arise, with at least one solution giving a maximum magnitude tending to 1.155 (π/3) in magnitude, compared to 1.27 (4/ π) for a square wave.

15.1.3v - Sinusoidal pulse-width modulation (pwm)

1 - Natural sampling

(a) Synchronous carrier

The output voltage waveform and method of generation for synchronous carrier, natural sampling sinusoidal pwm, suitable for the single-phase bridge of figure 15.1, are illustrated in figure 15.16. The switching points are determined by the intersection of the triangular carrier wave f_c and the reference modulation sine wave, f_o. The output frequency is at the sine-wave frequency f_o and the output voltage is proportional to the magnitude of the sine wave. The amplitude *M* ($0 \le M \le 1$) is called the modulation index. For example, figure 15.16a shows maximum voltage output *(M* = 1), while in figure 15.16b where the sine-wave magnitude is halved *(M* = 0.5*)*, the output voltage is halved.

If the frequency of the modulation sinewave, f_o, is an integer multiple of the triangular wave carrier-frequency, f_c that is, $f_c = nf_o$ where *n* is integer, then the modulation is *synchronous,* as shown in figure 15.16. If *n* is odd then the positive and negative output half cycles are symmetrical and the output voltage contains no even harmonics. In a three-phase system if *n* is a multiple of 3 (and odd), the carrier is a triplen of the modulating frequency and the spectrum does not contain the carrier or its harmonics.

$$f_c = (6q + 3) f_o = nf_o \qquad (15.66)$$

for *q* = 1, 2, 3.

The Fourier harmonic magnitudes of the line to line voltages are given by

$$a_n = V_t \cos\left(\frac{n\pi}{2}\right)\cos\left(\frac{n\pi}{3}\right)$$
$$b_n = V_t \sin\left(\frac{n\pi}{2}\right)\sin\left(\frac{n\pi}{3}\right) \qquad (15.67)$$

where V_t is proportional to the dc supply voltage V_s and the modulation index *M*.

Sinusoidal pwm requires a carrier of much higher frequency than the modulation frequency. The generated rectilinear output voltage pulses are modulated such that their duration is proportional to the instantaneous value of the sinusoidal waveform at the centre of the pulse; that is, the pulse area is proportional to the corresponding value of the modulating sine wave.

If the carrier frequency is very high, an averaging effect occurs, resulting in a sinusoidal fundamental output with high-frequency harmonics, but minimal low-frequency harmonics.

Rather than using two offset triangular carriers, as shown in figure 15.16, a triangular carrier without an offset can be used. Now the output only approximates the ideal. Figure 15.17 shows this pwm generation technique and voltage bipolar output waveform, when applied to the three-phase VSI inverter in figure 15.7. Two offset carriers are not applicable to six-switch, three-phase pwm generation since complementary switch action is required. That is, one switch in the inverter leg must always be on.

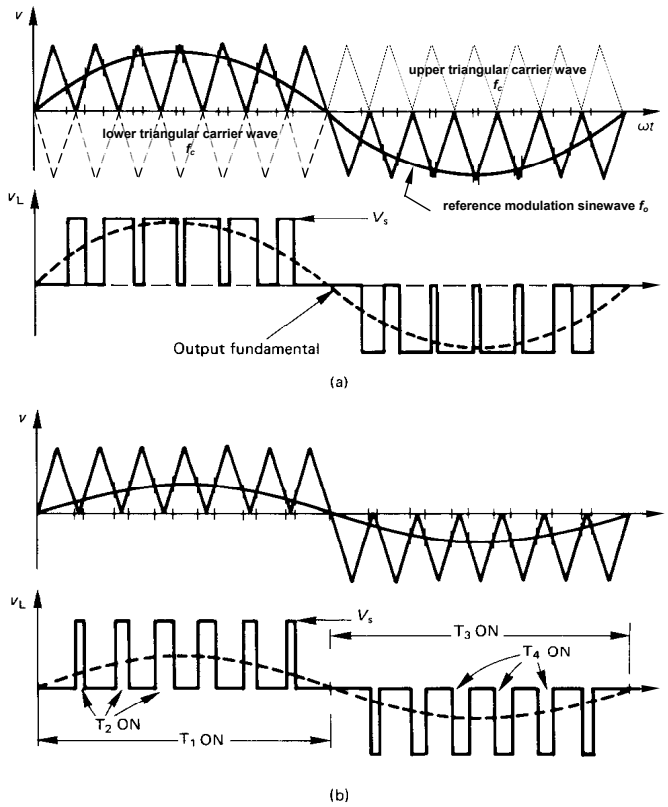


Figure 15.16. *Derivation of trigger signals for multi-level naturally sampled pulse-width modulation waveforms: (a) for a high fundamental output voltage (M = 1) and (b) for a lower output voltage (M = 0.5), with conducting devices shown.*

It will be noticed that, unlike the output in figure 15.16, no zero voltage output periods exist. This has the effect that, in the case of GCT thyristor bridges, a large number of commutation cycles is required. When zero output periods exist, as in figure 15.11, one GCT thyristor is commutated and the complementary device in that leg is not turned on. The previously commutated device can be turned back on without the need to commutate the complementary device, as would be required with the pwm technique illustrated in figure 15.17. Commutation losses are reduced, control circuitry simplified and the likelihood of simultaneous conduction of two series leg devices is reduced.
The alternating zero voltage loop concept can be used, where in figure 15.17b, rather than T_1 being on continuously during the first half of the output cycle, T_2 is turned off leaving T_1 on, then when either T_1 or T_2 must be turned off, T_1 is turned off leaving T_2 on.

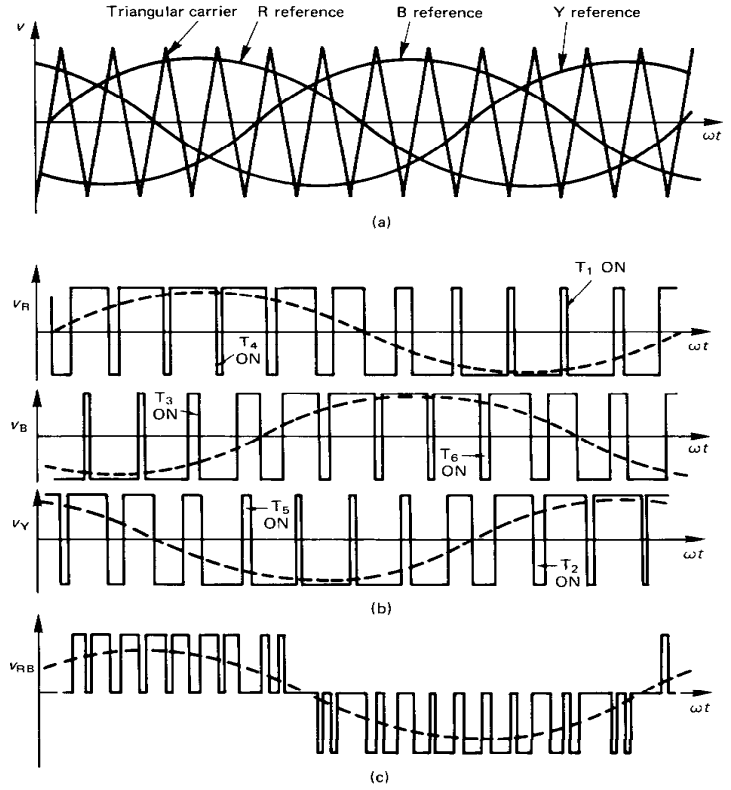


Figure 15.17. *Naturally sampled pulse-width modulation waveforms suitable for a three-phase bridge inverter: (a) reference signals; (b) conducting devices and fundamental sine waves; and (c) one output line-to-line voltage waveform.*

(b) Asynchronous carrier
When the carrier is not an integer multiple of the modulation waveform, asynchronous modulation results. Because the output frequency, f_o, is usually variable over a wide range, it is difficult to ensure $f_c = nf_o$. To achieve synchronism, the carrier frequency must vary with frequency f_o. Simpler generating systems result if a fixed carrier frequency is used, resulting in asynchronism between f_o and f_c at most output frequencies. Left over, incomplete carrier cycles create slowly varying output voltages, called subharmonics, which may be troublesome with low carrier frequencies, as found in high-power drives. Natural sampling, asynchronous sinusoidal pwm is usually restricted to analogue or ASIC implementation. The harmonic consequences of asynchronous-carrier natural-sampling are similar to asynchronous-carrier regular-sampling in 2 to follow.

2 - Regular sampling

(a) Asynchronous carrier
When a fixed carrier frequency is used, usually no attempt is made to synchronise the modulation frequency. The output waveforms do not have quarter-wave symmetry which produces subharmonics. These subharmonics are insignificant if $f_c >> f_o$, usually, $f_c > 20\, f_o$.
The implementation of sinusoidal pwm with microprocessors or digital signal processors is common because of flexibility and the elimination of analogue circuitry associated problems. The digital pwm generation process involves scaling, by multiplication, of the per unit sine-wave samples stored in ROM.

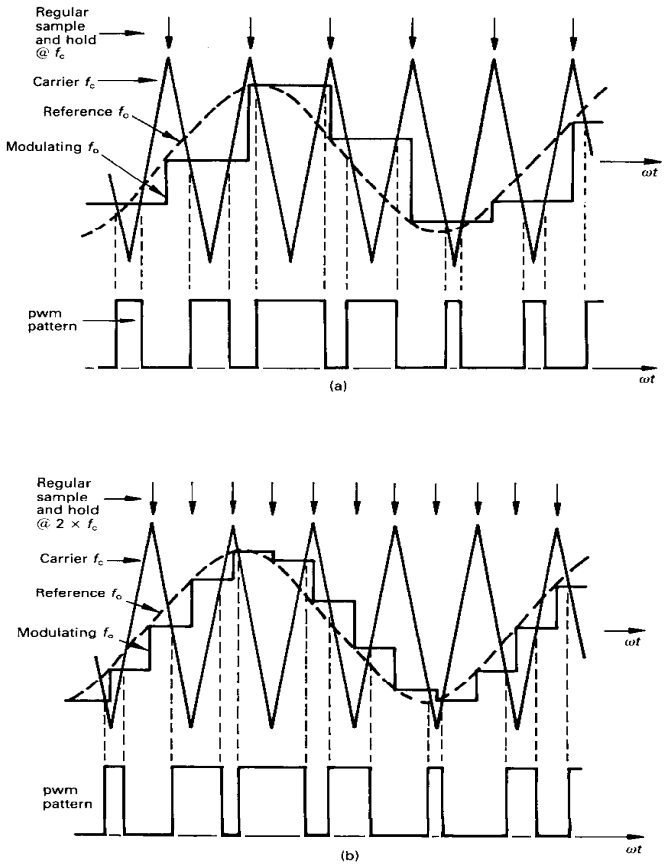


Figure 15.18. *Regular sampling, asynchronous, sinusoidal pulse-width-modulation: (a) symmetrical modulation and (b) asymmetrical modulation.*

The multiplication process is time-consuming, hence natural sampling is not possible. In order to minimise the multiplication rate, the sinusoidal sine-wave reference is replaced by a quantised stepped representation of the sine-wave. Figure 15.18 shows two methods used. Sampling is synchronised to the carrier frequency and the multiplication process is performed at twice the sampling rate for three-phase pwm generation (the third phase can be expressed in terms of two phases, since $v_1 + v_2 + v_3 = 0$).

- *Symmetrical modulation*

Figure 15.18a illustrates the process of symmetrical modulation, where sampling is at the carrier frequency. The quantised sine-wave is stepped and held at each sample point. The triangular carrier is then compared with the step sine-wave sample. The modulation process is termed symmetrical modulation because the intersection of adjacent sides of the triangular carrier with the stepped sine-wave, about the non-sampled carrier peak, are equidistant about the carrier peak. The pulse width, independent of the modulation index *M,* is symmetrical about the triangular carrier peak not associated with sampling, as illustrated by the upper pulse in figure 15.19. The pulse width is given by

$$t_{ps} = \frac{1}{2f_c}\left(1 - M \sin 2\pi f_o t_1\right) \qquad (15.68)$$

where t_1 is the time of sampling.

- *Asymmetrical modulation*

Asymmetrical modulation is produced when the carrier is compared with a stepped sine wave produced by sampling and holding at twice the carrier frequency, as shown in figure 15.18b. Each side of the triangular carrier about a sampling point intersects the stepped waveform at different step levels. The resultant pulse width is asymmetrical about the sampling point, as illustrated by the lower pulse in figure 15.19 for two modulation waveform magnitudes. The pulse width is given by

$$t_{pa} = \frac{1}{2f_c}\left(1 - \tfrac{1}{2}M\left(\sin 2\pi f_o t_1 + \sin 2\pi f_o t_2\right)\right) \qquad (15.69)$$

where t_1 and t_2 are the times at sampling such that $t_2 = t_1 + 1/2f_c$.

Figure 15.19 shows that a change in the modulation index *M* varies the pulse width on each edge, termed *double edge modulation.* A triangular carrier produces double edge modulation, while a sawtooth carrier produces *single edge modulation,* independent of the sampling technique.

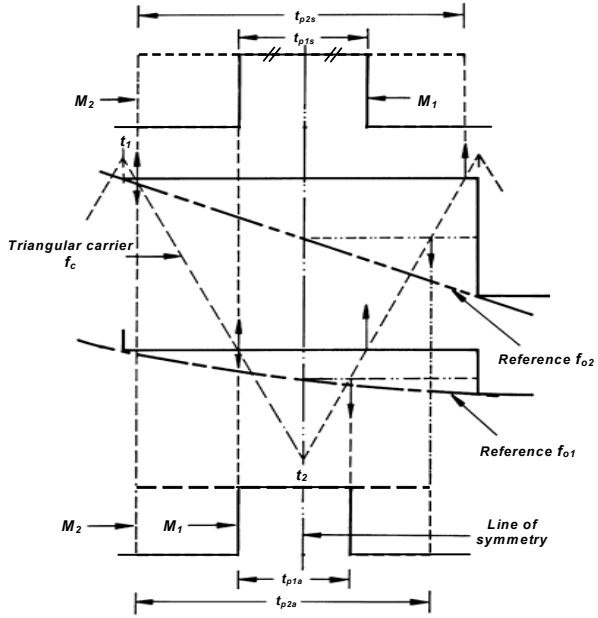


Figure 15.19. *Regular sampling, asynchronous, sinusoidal pulse-width-modulation, showing double edge: (upper) asymmetrical modulation and (lower) symmetrical modulation.*

3 - Frequency spectra of pwm waveforms

The most common form of sinusoidal modulation for three-phase inverters is regular sampling, asynchronous, fixed frequency carrier, pwm. If $f_c > 20f_o$, low frequency subharmonics can be ignored. The output spectra consists of the modulation frequency f_o with magnitude *M.* Also present are the spectra components associated with the triangular carrier, f_c. For any sampling, these are f_c and the odd harmonics of f_c. (The triangular carrier f_c contains only odd harmonics). These decrease in magnitude with increasing frequency. About the frequency nf_c are components of f_o spaced at ± $2f_o$, which generally decrease in magnitude when further away from nf_c. That is, at f_c the harmonics present are f_c, $f_c \pm 2f_o$, $f_c \pm 4f_o$, … while about $2f_c$, the harmonics present are $2f_c \pm f_0$, $2f_c \pm 3f_o$,..., but $2f_c$ is not present. The typical output spectrum is shown in figure 15.20. The relative magnitudes of the sidebands vary with modulation depth and the carrier related frequencies present, f_h, are given by

$$f_h = \tfrac{1}{2}\left(1 + (-1)^{n+1}\right) n f_c \pm \left(2k - \tfrac{1}{2}\left(1 + (-1)^n\right)\right) f_o \qquad (15.70)$$

where $k = 1, 2, 3, \ldots$ (sidebands) and $n = 1, 2, 3, \ldots$ (carrier)

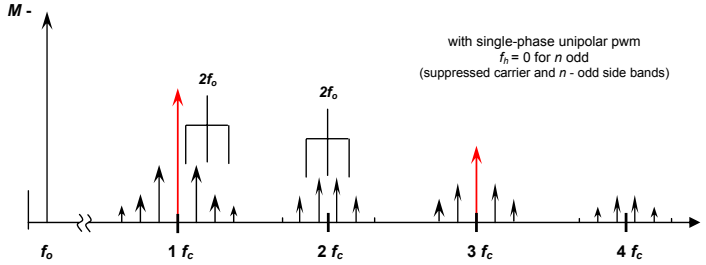


Figure 15.20. *Location of carrier harmonics and modulation frequency sidebands, showing all sideband separated by $2f_m$.*

Although the various pwm techniques produce other less predominate spectra components, the main difference is seen in the magnitude of the carrier harmonics and sidebands. The magnitudes increase as the pwm type changes from naturally sampling to regular sampling, then from asymmetrical to symmetrical modulation, and finally from double edge to single edge. With a three-phase inverter, the carrier f_c and its harmonics do not appear in the line-to-line voltages since the carrier f_c and in particular its harmonics, are co-phase to the three modulation waveforms.

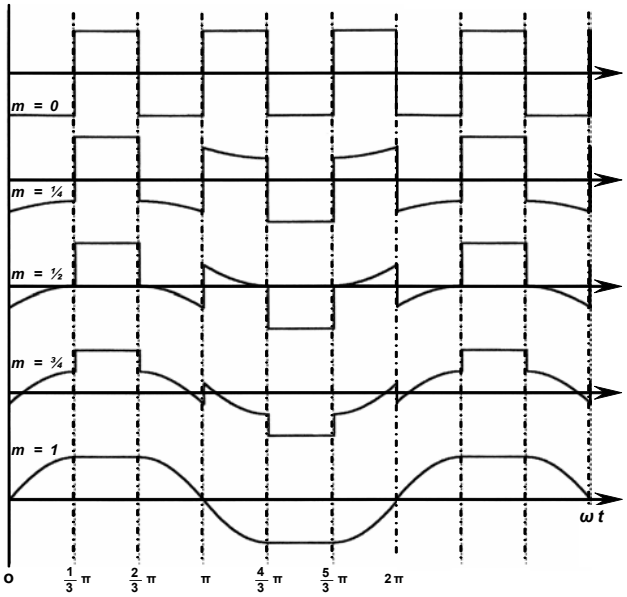


Figure 15.21. *Modulation reference waveform for phase dead banding.*

15.1.3vi - Phase dead-banding

Dead banding is when one phase (leg) is in a fixed on state, and the remaining phases are appropriately modulated so that the phase currents remain sinusoidal. The dead banding occurs for 60° periods of each cycle with the phase with the largest magnitude voltage being permanently turned on. Sequentially each switch is clamped to the appropriate link rail. The leg output is in a high state if it is associated with the largest positive phase voltage magnitude, while the phase output is zero if it is associated with the largest negative phase magnitude. Thus the phase outputs are cycled, being alternately clamped high and low for 60° every 180° as shown in figure 15.21. A consequence of dead banding is reduced switching losses since each leg is not switched at the carrier frequency for 120° (two 60° periods 180° apart). A consequence of dead banding is increased ripple current. Dead banding is achieved with discontinuous modulating reference signals. Dead banding for a continuous 120° per phase leg is also possible but the switching loss savings are not uniformly distributed amongst the six inverter switches.

The magnitude of the fundamental (with respect to the ac mains) when using standard PWM can be increased by $2/\sqrt{3}$ from $3x\sqrt{3}/2\pi$, 0.827pu to $3/\pi$, 0.955pu without introducing output voltage distortion, by the injection of triplen components, which are co-phasal in a three-phase system, and therefore do not appear in the line currents. Two approaches can be used to affect this undistorted output voltage magnitude increase.

- Triplen injection into the modulation waveform or
- Voltage space vector modulation

15.1.3vii - Triplen Injection modulation

1 - Triplens injected into the modulation waveform

An inverter reconstitutes three-phase voltages with a maximum magnitude of 0.827 ($3\sqrt{3}/2\pi$) of the fixed three-phase input ac supply, converted to dc before inversion. A motor designed for the fixed mains supply is therefore under-fluxed at rated frequency and not fully utilised on an inverter. As will be shown, by using third harmonic voltage injection, the flux level can be increased to 0.955 ($3/\pi$) of that produced on the three-phase ac mains supply.
If overmodulation ($M > 1$) is not allowed, then the modulation wave $M \sin \omega t$ is restricted in magnitude to $M = 1$, as shown in figure 15.22a.

If $V_{RN} = M \sin\omega t \le 1$pu
and $V_{YN} = M \sin(\omega t + \frac{2}{3}\pi) \le 1$ pu
then $V_{RY} = \sqrt{3}\, M \sin(\omega t - \frac{1}{6}\pi)$
where $0 \le M \le 1$

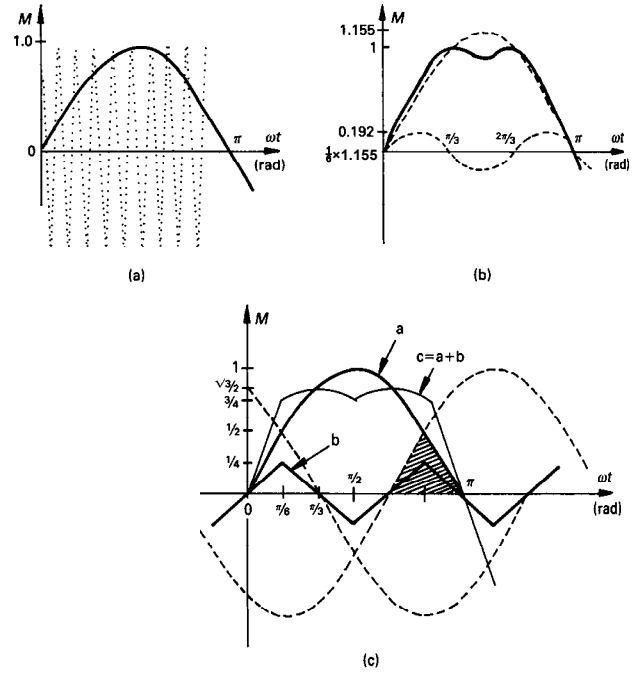


Figure 15.22. *Modulation reference waveforms: (a) sinusoidal reference, sin ωt; (b) third harmonic injection reference, sin $\omega t + \frac{1}{6}\sin 3\omega t$; and (c) triplen injection reference, sin $\omega t + (1/\sqrt{3}\pi)\{9/8 \sin 3\omega t - 80/81 \sin 9\omega t + \ldots\}$ where the near triangular waveform b is half the magnitude of the shaded area.*

In a three-phase pwm generator, the fact that harmonics at $3f_o$ (and odd multiplies of $3f_o$) vectorally cancel can be utilised effectively to increase M beyond 1, yet still ensure modulation occurs for every carrier frequency cycle.

Let $V_{RN} = M' \sin\omega t + \frac{1}{6}\sin 3\omega t) \le 1$ pu
and $V_{YN} = M' (\sin(\omega t + \frac{2}{3}\pi) + \frac{1}{6} \sin 3(\omega t + \frac{2}{3}\pi)) \le 1$ pu
then $V_{RY} = \sqrt{3}\, M' \sin(\omega t - \frac{1}{6}\pi)$

V_{RN} has a maximum instantaneous value of 1 pu at ωt = ±⅓π, as shown in figure 15.22b. Therefore

$$V_{RN}\left(\omega t = \tfrac{1}{3}\pi\right) = \frac{\sqrt{3}}{2} M' = 1$$

that is

$$\widehat{M}' = \frac{2}{\sqrt{3}} \widehat{M} = 1.155 \widehat{M} \qquad (15.71)$$

Thus the fundamental of the phase voltage is M' sin ωt = 1.155 M sin ωt. That is, if the modulation reference sin ωt + $\frac{1}{6}$ sin $3\omega t$ is used, the fundamental output voltage is 15.5 per cent larger than when sin ωt is used as a reference. The increased fundamental is shown in figure 15.22b.

The spatial voltage vector technique injects the triplens according to

$$V_{RN} = M' \left\{ \sin\omega t + \frac{1}{\sqrt{3}\pi} \sum_{r=0}^{\infty} \frac{(-1)^r}{\left[(2r+1)-\frac{1}{3}\right]\left[(2r+1)+\frac{1}{3}\right]} \right\} \sin\left[(2r+1)3\omega t\right] \qquad (15.72)$$

The Fourier triplen series represents half the magnitude of the shaded area in figure 15.22c (the waveform marked 'b'), which is formed by the three-phase sinusoidal waveforms. The spatial voltage vector waveform is defined by

$$\begin{aligned} &\frac{3}{2}\sin(\omega t) && 0 \le \omega t \le \tfrac{1}{6}\pi \\ &\frac{\sqrt{3}}{2}\sin(\omega t + \tfrac{1}{6}\pi) && \tfrac{1}{6}\pi \le \omega t \le \tfrac{1}{2}\pi \end{aligned} \qquad (15.73)$$

The use of this reference increases the duration of the zero volt loops, thereby decreasing inverter output current ripple. The maximum modulation index is 2/√3, 1.155. Third harmonic injection, yielding M = 1.155, is a satisfactory approximation to spatial voltage vector injection.

2 - Voltage space vector pwm

When generating three-phase quasi-square output voltages, the inverter switches step progressively to each of the six switch output possibilities (states). In figure 15.10, when producing the quasi-square output, each of these six states is represented by an output voltage space vector. Each vector has a ⅓π displacement from its two adjacent states, and each has a length V_s which is the pole output voltage relative to the inverter 0V rail. Effectively, the quasi-square three-phase output is generated by a rotating vector of length V_s, jumping successively from one output state to the next in the sequence, and in so doing creating six voltage output sectors. The speed of rotation, in particular the time for one rotation, determines the inverter output frequency. The sequence of voltage vectors {v_1, v_3, v_2, v_6, v_4, v_5} is arranged such that stepping from one state to the next involves only one of the three poles changing state. Thus the number of inverter devices needing to change states (switch) at each transition, is minimised.

[If the inverter switches are relabelled, upper switches T_1, T_2, T_3 - right to left; and lower switches T_4, T_5, T_6 - right to left: then the rotating voltage sequence becomes {v_1, v_2, v_3, v_4, v_5, v_6}]

Rather than stepping ⅓π radians per step, from one voltage space vector position to the next, thereby producing a six-step quasi-square fixed magnitude voltage output, the rotating vector is rotated in smaller steps based on the position being updated at a constant rate (carrier frequency). Furthermore, the vector length can be varied, modulated, to a magnitude less than V_s.

$$\frac{t_a}{T_c} = \frac{|V_a|}{|v_1|} = \frac{\frac{2}{\sqrt{3}} V_{o/p} \sin\left(\frac{1}{3}\pi - \theta\right)}{V_s}$$

$$\frac{t_b}{T_c} = \frac{|V_b|}{|v_3|} = \frac{\frac{2}{\sqrt{3}} V_{o/p} \sin\theta}{V_s} \qquad \text{where } |v_1| = |v_3| \qquad (15.74)$$

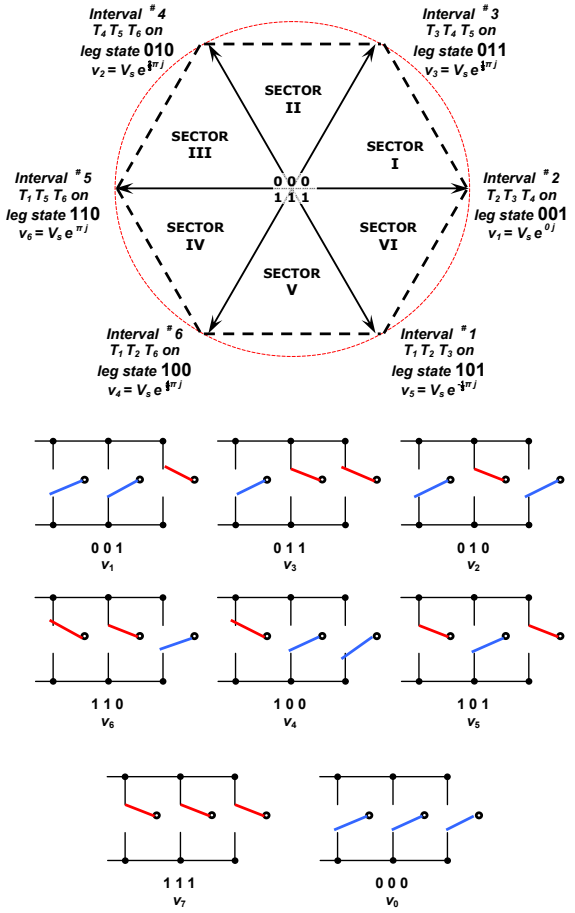


Figure 15.23. *Instantaneous output voltage states for the three legs of an inverter.*

To incorporate a variable rotating **vector length** (modulation depth), it is necessary to vary the average voltage in each carrier period. Hence pulse width modulation is used in the period between each finite step of the rotating vector. Pulse width modulation requires the introduction of zero voltage output states, namely all the top switches on (state 111, v_7) or all the lower switches on (state 000, v_0). These two extra states are shown in figure 15.23, at the centre of the hexagon. Now the pole-to-pole output voltage can be zero, which allows duty cycle variation to achieve variable average output voltage for each phase, within each carrier period, proportional to the magnitude of the position vector.
To facilitate **vector positions** (angles) that do not lie on one of the six quasi-square output vectors, an intermediate vector $V_{o/p}\, e^{j\theta}$ is resolved into the vector sum of the two quasi-square vectors adjacent to the rotating vector. This process is shown in figure 15.24 for a voltage vector $V_{o/p}$ that lies in sector I, between output states v_1 (001) and v_3 (011). The voltage vector has been resolved into the two components V_a and V_b as shown.

The time represented by quasi-square vectors v_1 and v_3 is the carrier period T_c, in each case. Therefore the portion of T_c associated with v_a and v_b is scaled proportionally to v_1 and v_3, giving t_a and t_b.

The two sine terms in equation (15.74) generate two sine waves displaced by 120°, identical to that generated with standard carrier based sinusoidal pwm.

The sum of t_a and t_b cannot be greater than the carrier period T_c, thus

$$t_a + t_b \le T_c$$
$$t_a + t_b + t_o = T_c \qquad (15.75)$$

where the slack variable t_o has been included to form an equality. The equality dictates that vector v_1 is used for a period t_a, v_3 is used for a period t_b, and during period t_o, the null vector, v_0 or v_7, at the centre of the hexagon is used, which do not affect the average voltage during the carrier interval T_c.
A further constraint is imposed in the time domain. The rotating voltage vector is a fixed length for all rotating angles, for a given inverter output voltage. Its length is restricted in both time and space. Obviously the resolved component lengths cannot exceed the pole vector length, V_s. Additionally, the two vector magnitudes are each a portion of the carrier period, where t_a and t_b could be both equal to T_c, that is, they both have a maximum length V_s. The anomaly is that voltages v_a and v_b are added vectorially but their scalar durations (times t_a and t_b) are added linearly. The longest time $t_a + t_b$ possible is when t_o is zero, as shown in figures 15.24a and 15.23a, by the hexagon boundary. The shortest vector to the boundary is where both resolving vectors have a length ½V_s, as shown in figure 15.24b. For such a condition, $t_a = t_b = ½T_c$, that is $t_a + t_b = T_c$. Thus for a constant inverter output voltage, when the rotating voltage vector has a constant length, $\hat{V}_{o/p}$, the locus of allowable rotating reference voltage vectors must be within the circle scribed by the maximum length vector shown in figure 15.24b. As shown, this vector has a length $v_1 \cos 30°$, specifically $0.866V_s$. Thus the full quasi-square vectors v_1, v_2, etc., which have a magnitude of $1 \times V_s$, cannot be used for generating a sinusoidal output voltage. The excess length of each quasi-square voltage (which represents time) is accounted for by using zero state voltage vectors for a period corresponding to that extra length (1- cos30° at maximum output voltage).

Having calculated the necessary periods for the inverter poles (t_a, t_b, and t_o), the carrier period switching pattern can be assigned in two ways.

- Minimised current ripple
- Minimised switching losses, using dead banding

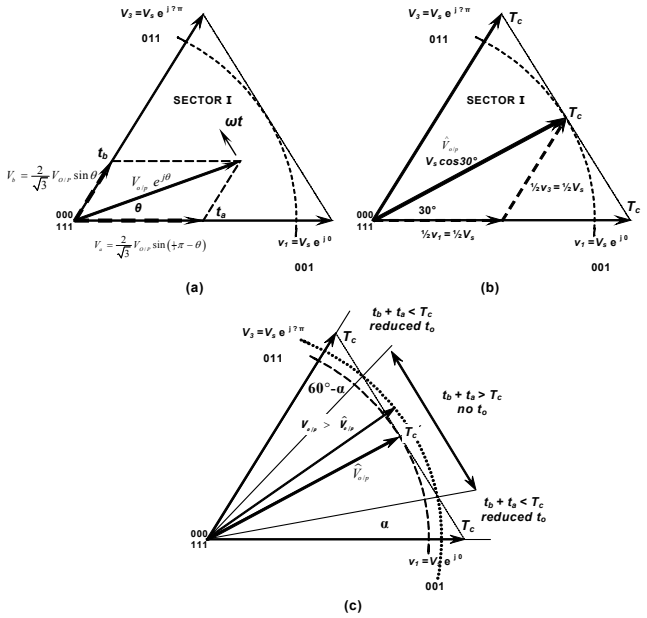


Figure 15.24. *First sector of inverter operational area involving pole outputs 001 and 011: (a) general rotating voltage vector; (b) maximum allowable voltage vector length for undistorted output voltages; and (c) over modulation.*

Each approach is shown in figure 15.25, using single edged modulation. The waveforms are based on the equivalent of symmetrical modulation where the pulses are symmetrical about the carrier trough. By minimising the current ripple, seven switching states are used per carrier cycle, while for loss minimisation (dead banding) only five switching states occur, but at the expense of increased ripple current in the output current. When dead banding, the zero voltage state v_0 is used in even numbered sextants and v_7 is used in odd numbered sextants.
Sideband and harmonic component magnitudes can be decreased if double-edged modulation placement of the states is used, which requires recalculation of t_a, t_b, and t_o at the carrier crest, as well as at the trough.

Over-modulation is when the magnitude of the demanded rotating vector is greater than $\hat{V}_{o/p}$ such that the zero voltage time reduces to zero, t_o = 0, during a portion of the time of one rotation of the output vector. Initially this occurs at 30° $\left(\frac{1}{6}\pi\left(2N_{\text{sector}}-1\right)\right)$ when the output vector length reaches $\hat{V}_{o/p}$, as shown in figure 15.24b. As the demand voltage magnitude increases further, the region around the 30° vector position where t_o ceases to occur, increases as shown in figure 15.24c. When the output rotational vector magnitude increases to V_s, the maximum possible, angle α reduces to zero, and t_o ceases to occur at any rotational angle. The values of t_a, t_b, and t_o (if greater than zero), are calculated as usual, but pulse times are assigned pro rata to fit within the carrier period T_c.

The switching frequency can be decreased by using dc-link clamping like in 15.1.3vi. Each leg is successively clamped, alternately to each dc rail (that is states 000 and 111 are alternated every 60° when a zero state is required). For example, in sector 1, an odd sector number, the sequence would be states [111] [011] [001] [011] [111]. In sector 2, an even sector number, the sequence would be states [000] [010] [001] [010] [000]. In odd sector number zero vector [111] is used, while [000] is used in even numbered sector.

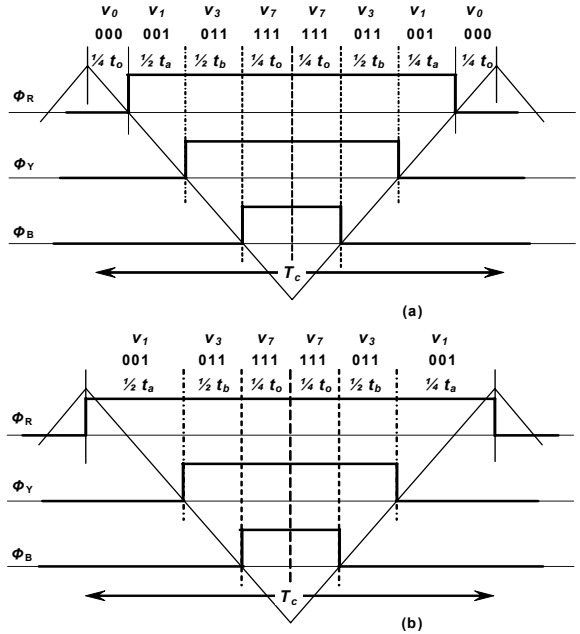


Figure 15.25. *Assignment of pole periods t_a and t_b based on: (a) minimum current ripple and (b) minimum switching transitions per carrier cycle, T_c.*

15.1.4 Common mode voltage

The common mode CM voltage is a remanent voltage with respect to a reference zero 'o' as shown in figure 15.26(a). The line voltages and a common mode voltage V_{cm} can be derived based on leg phase voltages (V_{ao}, V_{bo}, V_{co}) of a two-level three-phase inverter, as follows:

$$V_{ao} = V_{an} + V_{cm}$$
$$V_{bo} = V_{bn} + V_{cm}$$
$$V_{co} = V_{cn} + V_{cm}$$

The sum of the leg voltages is given by :

$$V_{ao} + V_{bo} + V_{co} = (V_{an} + V_{bn} + V_{cn}) + 3V_{cm}$$

In a three-phase system:

$$V_{an} + V_{bn} + V_{cn} = 0$$

thus the common mode voltage source in a three-phase system is:

$$V_{cm}(t) = \tfrac{1}{3}\left(V_{ao}(t) + V_{bo}(t) + V_{co}(t)\right) = \tfrac{1}{3}\left(V_a(t) + V_b(t) + V_c(t)\right)$$

where V_a, V_b, V_c are the phase voltages, n is the load neutral, and o is the split dc link centre voltage node.

The CM voltage is a staircase function with steps of $\frac{1}{3}V_{dc}$ (V_{dc} - dc link voltage) within a frequency of three time the inverter modulation frequency, plus carrier related modulation steps superimposed, as shown in figure 15.26. Because the Kirchhoff voltage loop is usually completed by capacitance, a train of damped sinewave pulses of CM current flows as a result of each CM voltage step. In electrical machines such currents cause bearing deterioration. Generally, in electrical circuits, the CM currents through substrates, isolation capacitance, etc., produce and inject electrical noise and cause EMC interference.

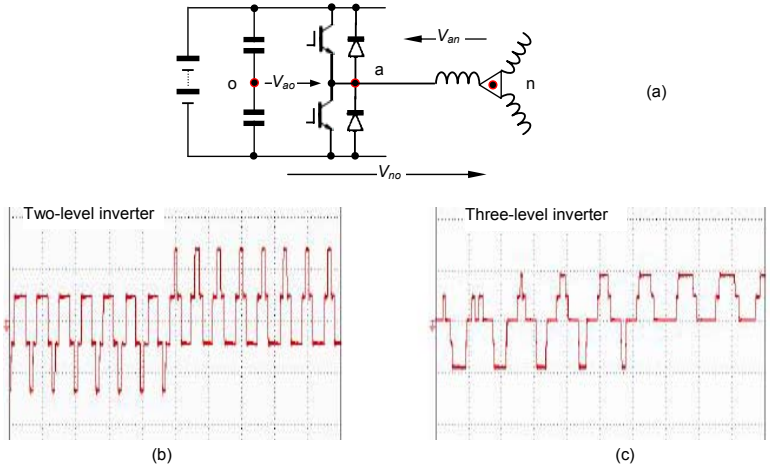


Figure 15.26. *Inverter common mode voltage: (a) source voltages, (b) two level inverter common mode voltage, and (c) three level inverter common mode voltage.*

15.1.5 DC link voltage boosting

Both triplen injection and SVM under flux a 50/60Hz machine designed to operate from a given three-phase ac mains supply. Increased dc rail voltage (over that produced from rectification) can be achieved with the three-phase boost converter shown in figure 15.27. Additionally, being a boost converter, the ac input current can be continuous, and force to track any reference whilst maintaining the transferred power balance between the input and the output. By using the ac mains as references, the input current can be sinusoidal, of high quality and at any desired angle, usual in phase with supply voltage - unity power factor. The bridge acts as an uncontrolled rectifier when the output voltage is below the mains rectified level, as during start-up when the dc link capacitor is uncharged. Near instantaneous power reversal (by current direction reversal) is possible when the converter acts as a dc to ac inverter. The power factor angle and three-phase voltage magnitude can be varied to control the active and reactive power flows back into the ac grid. PWM or SVM control can insure sinusoidal current in both conversion directions.

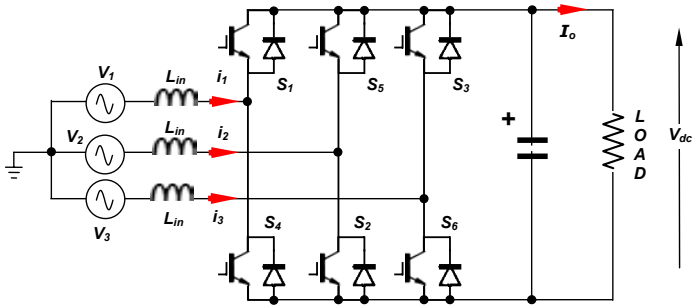


Figure 15.27. *PWM boost three-phase rectifier.*

15.2 dc-to-ac controlled current-source inverters

In the current source inverter, *CSI,* the dc supply is of high reactance, being inductive so as to maintain the required inverter output bidirectional current independent of the inverter load.

15.2.1 Single-phase current source inverter

A single-phase, controlled current-sourced bridge is shown in figure 15.28a and its near square-wave output current is shown in figure 15.28b. No freewheel diodes are required and the thyristors require forced commutation and have to withstand reverse voltages. An inverter current path must be maintained at all times for the source controlled current.

Consider thyristors T_1 and T_2 on and conducting the constant load current. The capacitors are charged with plates X and Y positive as a result of the previous commutation cycle.

- ***Phase I***

Thyristors T_1 and T_2 are commutated by triggering thyristors T_3 and T_4. The capacitors impress negative voltages across the respective thyristors to be commutated off, as shown in figure 15.29a. The load current is displaced from T_1 and T_2 via the path T_3-C_1-D_1, the load and D_2-C_2-T_4. The two capacitors discharge in series with the load, each capacitor reverse biasing the thyristor to be commutated, T_1 and T_2 as well as diodes D_3 to D_4. The capacitors discharge linearly (due to the constant current source).

- ***Phase II***

When both capacitors are discharged, the load current transfers from D_1 to D_2 and from D_3 to D_4, which connects the capacitors in parallel with the load via diodes D_1 to D_2. The plates X and Y now charge negative, ready for the next commutation cycle, as shown in figure 15.29b. Thyristors T_1 and T_2 are now forward biased and must have attained forward blocking ability before the start of phase 2.

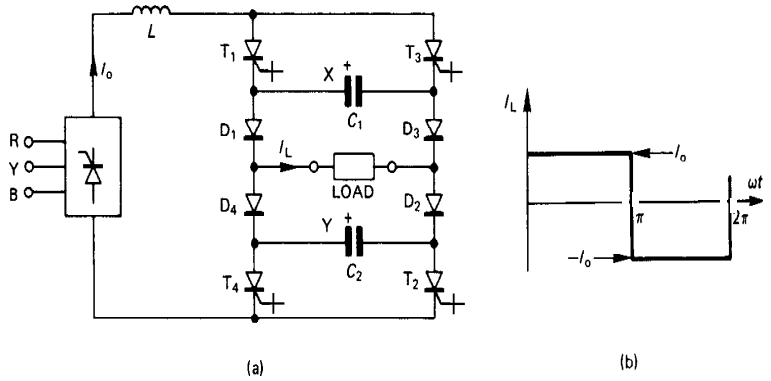


Figure 15.28. *Single-phase controlled-current sourced bridge inverter: (a) bridge circuit with a current source input and (b) load current waveform.*

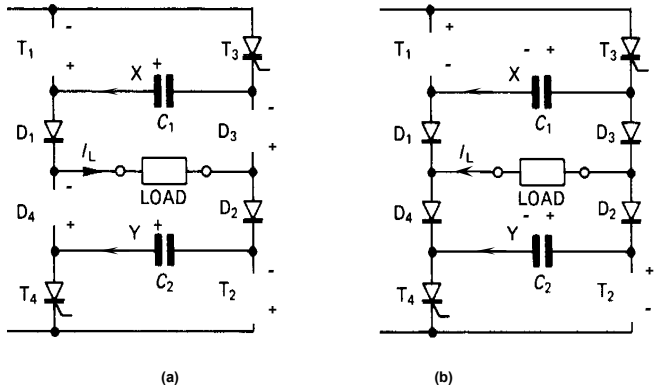


Figure 15.29. *Controlled-current sourced bridge inverter showing commutation of T_1 and T_2 by T_3 and T_4: (a) capacitors C_1 and C_2 discharging and T_1, T_2, D_3, and D_4 reversed biased and (b) C_1, C_2, and the load in parallel with C_1 and C_2 charging.*

The on-going thyristor automatically commutates the outgoing thyristor. This repeated commutation sequencing is a processed termed *auto-sequential thyristor commutation*. The load voltage is load dependent and usually has controlled voltage spikes during commutation.
Since the GTO and GCT both can be commutated from the gate, the two commutation capacitors C_1 and C_2 are not necessary. Commutation overlap is still essential. Also, if the thyristors have reverse blocking capability, the four diodes D_1 to D_4 are not necessary. IGBTs require series blocking diodes, which increases on-state losses. In practice, the current source inverter is only used in very high-power applications (>1MVA), and the ratings of the self-commutating thyristor devices can be greatly extended if the simple external capacitive commutation circuits shown in figure 15.28 are used to reduce thyristor turn-off stresses.

15.2.2 Three-phase current source inverter

A three-phase controlled current-source inverter is shown in figure 15.30a. Only two thyristors can be on at any instant, that is, the 120° thyristor conduction principle shown in figure 15.11 is used. A quasi-square line current results, as illustrated in figure 15.30b. There is a 60° phase displacement between commutation of an upper device followed by commutation of a lower device. An upper device (T_1, T_3, T_5) is turned on to commutate another upper device, and a lower device (T_2, T_4, T_6) commutates another lower device. The three upper capacitors are all involved with each upper device commutation, whilst the same constraint applies to the lower capacitors. Thyristor commutation occurs in two distinct phases.

- ***Phase I***
  In figure 15.31a the capacitors C_{13}, C_{35}, C_{51} are charged with the shown polarities as a result of the earlier commutation of T_5. T_1 is commutated by turning on T_3. During commutation, the capacitor between the two commutating switches is in parallel with the two remaining capacitors which are effectively connected in series. Capacitor C_{13} provides displacement current whilst in parallel, C_{35} and C_{151} in series also provide thyristor T_1 displacement current, thereby reverse biasing T_1.
- ***Phase II***
  When the capacitors have discharged, T_1 becomes forward biased, as shown in figure 15.31b, and must have regained forward blocking capability before the applied positive *dv/dt*. The capacitor voltages reverse as shown in figure 15.31b and when fully charged, diode D_1 ceases to conduct. Independent of this commutation, lower thyristor T_2 is commutated by turning on T_4, 60° later.

As with the single-phase current sourced inverter, assisted capacitor commutation can greatly improve the capabilities of self-commutating thyristors, such as the GTO thyristor and GCT. The output capacitors stiffen the output ac voltage.

A typical application for a three-phase current-sourced inverter would be to feed and control a three-phase induction motor. Varying load requirements are met by changing the source current level over a number of cycles by varying the link inductor input voltage.

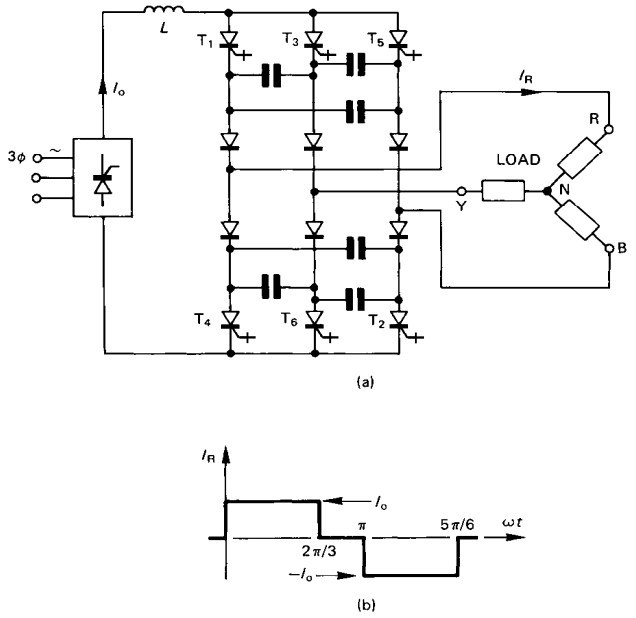


Figure 15.30. *Three-phase controlled-current sourced bridge inverter: (a) bridge circuit with a current source input and (b) load current waveform for one phase showing 120° conduction.*

An important advantage of the controlled current source concept, as opposed to the constant voltage link, is good fault tolerance and protection. An output short circuit or simultaneous conduction in an inverter leg is controlled by the current source. Its time constant is usually longer than that of the input converter, hence converter shut-down can be initiated before the link current can rise to a catastrophic level.

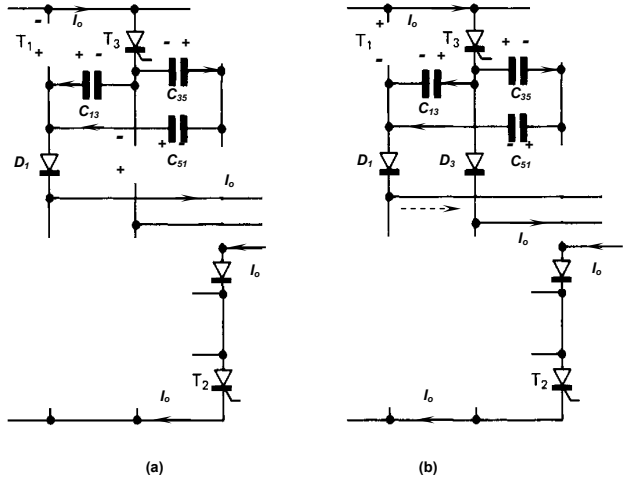


Figure 15.31. *Controlled-current sourced bridge three-phase inverter showing commutation of T_1 and T_3: (a) capacitors C_{13} discharging in parallel with C_{35} and C_{51} discharging in series, with T_1 and D_3 reversed biased (b) C_{13}, C_{35}, and C_{51} charging in series with the load , with T_1 forward biased.*

PWM techniques are applicable to current source inverters in order to reduce current harmonics, thereby reducing load losses and pulsating motor shaft torques. Since current source inverters are most attractive in very high-power applications, inverter switching is minimised by using optimal pwm (selected harmonic elimination). The central 60° portion about the maximums of each phase cannot be modulated, since link current must flow and during such periods both the other phases require the opposite current direction. Attempts to over come such pwm restrictions include using a current sourced inverter with additional parallel current displacement paths as shown in figure 15.32. The auxiliary thyristors, T_{upper} and T_{lower}, and capacitors, C_R, C_Y, and C_B, provide alternative current paths (extra control states) and temporary energy storage. The auxiliary thyristor can be commutated by the extra capacitors. The problem can be alleviated by triplen injection or SVM, which minimises the duration of any such periods in the central 60° blocks, so that the link can be shorted for the short periods.

Characteristics and features of current source inverters

- The inverter is simple and can utilise rectifier grade thyristors. The switching devices must have reverse blocking capability and experience high voltages (both forward and reverse) during commutation.
- Commutation capability is load current dependent and a minimum load is required. This limits the operating frequency and precludes use in UPS systems. The limited operating frequency can result in torque pulsations.
- The inverter can recover from an output short circuit hence the system is rugged and reliable – fault tolerant.
- The converter-inverter configuration has inherent four quadrant capability without extra power components. Power inversion is achieved by reversing the converter average voltage output with a delay angle of $\alpha > \frac{1}{2}\pi$, as in the three-phase fully controlled converter shown in figure 12.11 (or 15.4.3). In the event of a power supply failure, mechanical braking is necessary. Dynamic braking is possible with voltage source systems.
- Current source inverter systems have sluggish performance and stability problems on light loads and at high frequency. On the other hand, voltage source systems have minimal stability problems and can operate open loop.
- Each machine must have its own controlled rectifier and inverter. The dc link of the voltage source scheme can be used by many inverters or many machines can utilise one inverter. A dc link offers limited ride-through.
- Current feed inverters tend to be larger in size and weight, because of the link inductor and filtering requirements.

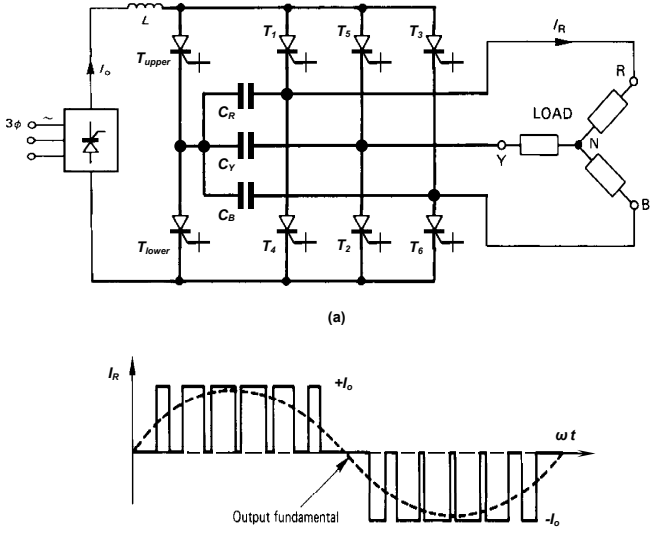


Figure 15.32. *Three-phase controlled-current sourced bridge inverter with alternative commutation current paths: (a) bridge circuit with a current source input and two extra thyristors and (b) load current waveform for one phase showing 180° conduction involving pwm switching.*

15.3 Multi-level voltage-source inverters

The conventional three-phase, six-switch dc to ac voltage-source inverter is shown in figure 15.7. Each of the three inverter legs has an output which can provide one of two voltage levels, V_s, when the upper switch (or diode) is on, and 0 when the lower switch (or diode) conducts. The quality of the output waveform is determined by the resolution and switching frequency of the pwm technique used.

A multilevel inverter (directly or indirectly) divides the dc rail, so that the output of the leg can be more than two discrete levels, as shown in figure 15.33 for a diode clamped multilevel inverter model. In this way, the output quality is improved because both pulse width modulation and amplitude modulation can be used. The output pole is made from more than two series connected, clamped switches, so the total dc voltage rail can be the sum of the voltage rating of the individual switches. Very high output voltages can be achieved, where each device does not experience a voltage in excess of its individual rating.

A multilevel inverter allows higher output voltages with low distortion (due to the use of both pulse width and amplitude modulation) and reduced output *dv/dt*.

There are three main types of multilevel converters

- Diode clamped
- Flying capacitor, and
- Cascaded H-bridge

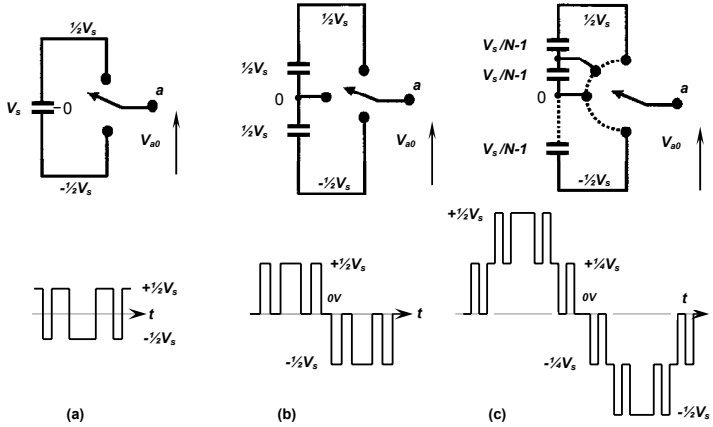


Figure 15.33. *One phase leg of a voltage-source bridge inverter with: (a) two levels; (b) three levels; and (c) N-levels, with N-1 capacitors and waveform for five levels.*

15.3.1 Diode clamped multilevel inverter

Figure 15.33 shows the basic principle of the diode clamped (or neutral point clamped, NPC) multilevel inverter, where only one dc supply, V_s, is used and N is the number levels present in the output voltage between the leg output and the inverter negative terminal, $V_{a\text{-}neg}$. The capacitors split the dc rail voltage into a number of lower voltage levels, each of which can be tapped and connected to the leg output through switches (and diodes). Only one string of series connected capacitors is necessary for any number of output phase legs.

The number of levels in the line-to-line voltage waveform will be

$$k = 2N - 1 \tag{15.76}$$

while the number of levels in the line to load neutral of a star (wye) load will be

$$p = 2k - 1 \tag{15.77}$$

The number of capacitors required, independent of the number of phase, is

$$N_{cap} = N - 1 \tag{15.78}$$

while the number of clamping diodes per phase is

$$D_{clamp} = 2(N-2) \text{ but } (N-1)(N-2) \text{ if rated at switch voltage} \tag{15.79}$$

The number of possible switch states is

$$n_{states} = N^{phases} \tag{15.80}$$

and the number of switches (and inverse parallel diodes) in each leg is

$$S_n = 2(N-1) \tag{15.81}$$

The basic three-level inverter (±½V_s, 0) is shown in figure 15.34, along with the basic three-level voltage from the leg output to centre tap of the capacitor string, *R* (neutral point). When switch T_1 is on, its complement T_1' is off, and visa versa. Similarly for the pair of switches T_2 and T_2'. Specifically T_1 and T_2 on give the output +½V_s, T_1' and T_2' on give the output -½V_s, and T_2 and T_1' on give the output 0. Essential to attaining these output levels, are the clamping diodes D_u and $D_ł$. These two diodes clamp the outer switches to the capacitor string mid-point, which is half the dc rail voltage. In this way, no switch experiences a voltage in excess of half the dc rail voltage. Inner switches must be turned on (or off) before outer switches are turned on (or off).

The five-level inverter uses four capacitors, and eight switches in each inverter leg. A set of clamping diodes (three in total for each leg) clamp the complementary switches in each leg. The output is characterised by having five levels, ±½V_s, ±¼ V_s, and zero. Some of the clamping diodes experience voltages in excess of that experienced by the main switches. Series connection of some of the clamping diodes avoids this limitation, but at the expense of increasing the number of clamping diodes from 2×(*N*-2) to (*N*-1)×(*N*-2) per phase. Thus, depending on the diode position in the structure, two diodes have blocking requirements of

$$V_{RB} = \frac{N-1-k}{N-1} V_s \qquad (15.82)$$

where $1 \le k \le N-2$. These diodes require series connection of diodes, if all devices in the structure are to support $V_s/(N-1)$. For $N > 2$, capacitor imbalance occurs at high modulation indices.

The general output voltage, to the centre of the capacitor string is given by

$$V_{an} = \frac{V_s}{N-1}(T_1 + T_2 + .. \quad .. + T_{N-1} - ½(N-1)) \qquad (15.83)$$

Common to all diode clamped inverter, each phase leg is identical in structure, and all legs share a common dc link capacitor string.

Table 15.5 in combination with the six parts of figure 15.35, show the conducting devices for the six different output voltage and current combinations of the NPC inverter leg. The commercial inverter, HVDC Light, uses the NPC structure in figure 15.34, but uses extensive series connection of devices to achieve a high dc link voltage. The main problem in increasing the number of output voltage levels, other than increased circuit complexity, is voltage balancing the dc link series connected capacitors at higher modulation levels, $M > ½(N-1)$. This capacitor voltage balancing problem can be avoided when two multilevel inverters are used in an ac-dc-ac back to back converter arrangement, where the link capacitors are common to both converters.

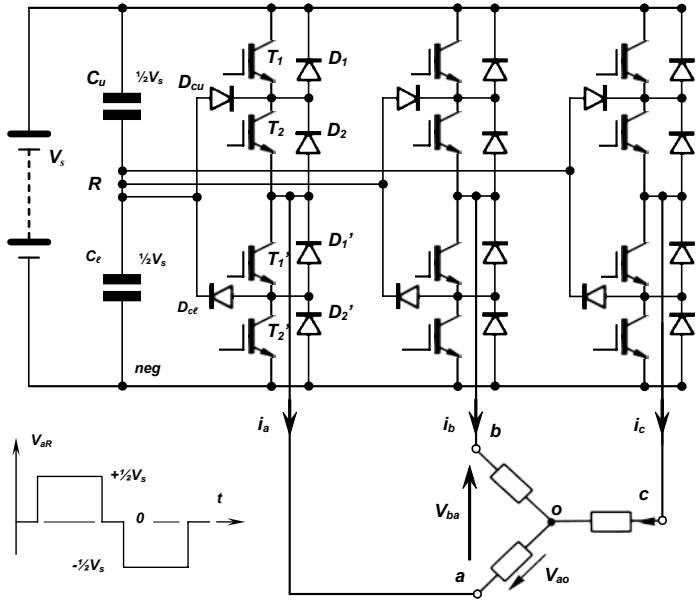


Figure 15.34. *Three-phase, voltage-source, three-level, diode-clamped (NPC) bridge inverter.*

Table 15.5: Conduction paths in the diode clamped three-level inverter

V_{out}	On switches	Output current and path $+i_L$		$-i_L$		Active clamping diodes
½ V_s	T_1 T_2	T_1 T_2	Fig 15.35a	D_1 D_2	Fig 15.35d	none
0	T_1' T_2	D_{cu} T_2	Fig 15.35b	T_1' $D_{cł}$	Fig 15.35e	D_{cu} $D_{cł}$
-½ V_s	T_1' T_2'	D_1' D_2'	Fig 15.35c	T_1' T_2'	Fig 15.35f	none

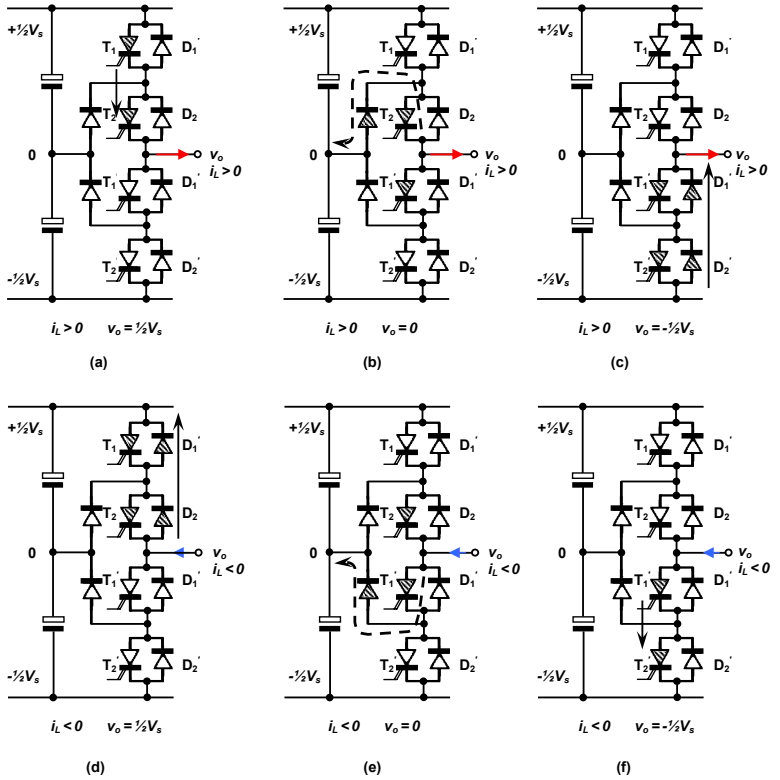


Figure 15.35. *The six output voltage and current combinations for the NPC bridge inverter: (a), (b), (c) output current $i_L > 0$; and (d), (e), (f) output current $i_L < 0$.*

15.3.2 Flying capacitor multilevel inverter

One leg of a fly-capacitor clamped five-level voltage source inverter is shown in figure 15.36b, where capacitors are used to clamp the switch voltages to ¼V_s. The available output voltages are ±½V_s, ±¼V_s, and 0, where the output is connected to the dc link (V_s and 0) indirectly via capacitors. Figure 15.36 shows that in general, switches T_n and T_{n+1} connect to capacitor C_n. The configuration offers more usable switch states than the clamped diode inverter, and this redundancy allows better, flexible control of capacitor voltages. For example, Table 15.5 shows that there are six states for obtaining 0V output, and four states for each of ±¼V_s. The output states ±½V_s do not involve the capacitors, hence they offer no redundant states. The basic switch restriction is that only one complementary switch (for example, T_4 or T_4') is on at any time, so as to prevent shorting of a flying capacitor (e.g., T_4 and T_4' would short C_3).

The number of levels in the line-to-line voltage waveform will be

$$k = 2N - 1 \tag{15.84}$$

while the number of levels in the line to load neutral of a star (wye) load will be

$$p = 2k - 1 \tag{15.85}$$

The number of capacitors required, which is dependent of the number of phase, is for each phase

$$N_{cap} = \tfrac{1}{2}(N-1)(N-2) \tag{15.86}$$

The number of possible switch states is

$$n_{states} = N^{phases} \tag{15.87}$$

and the number of switches in each leg is

$$S_n = 2(N-1) \tag{15.88}$$

The current output paths in Table 15.6 are made up by the series (and parallel) connection of the flying capacitors through the turn-on of the appropriate switches. Capacitors shown as negative are discharging in the formed path, while those shown as positive are charging. Use of the shown redundant states allows control to maintain the necessary voltage level on all the flying capacitors, while providing the desired output voltages.

A feature of the flying capacitor multilevel inverter is its ride through capability due to the large capacitance used. On the other hand, the capacitors have a high voltage rating and suffer from high current ripple, since they conduct the full load current when connected into an active output voltage state. Capacitor initial charging is also problematic, especially given the capacitors for each leg, and between the different legs, are independent.

If all the flying capacitors are voltage rated at the switch voltage level, then C_2 comprises two series connect capacitors and C_3 comprises three series capacitors, and all the same voltage rating as C_1.

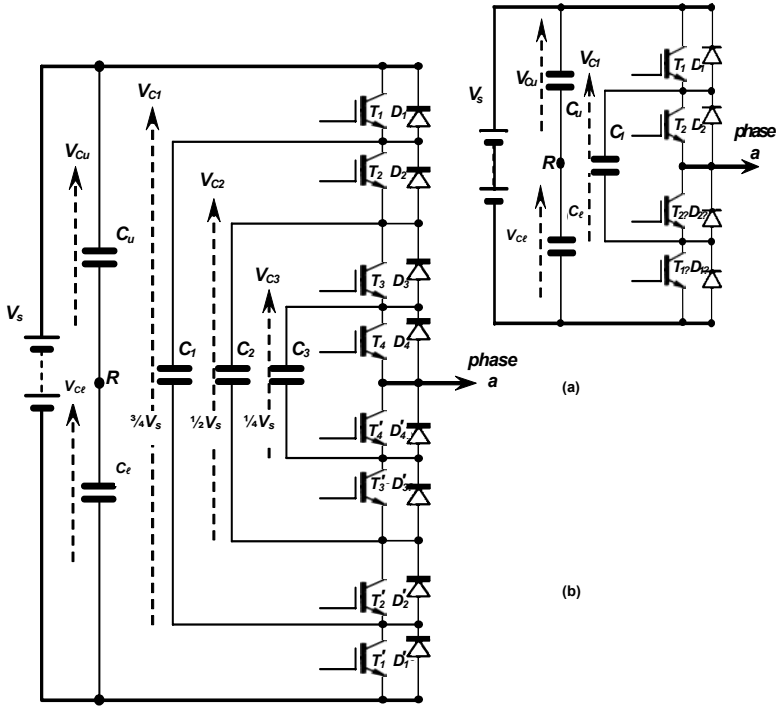


Figure 15.36. *One leg of a voltage-source:*
(a) three-level and (b) five-level, flying capacitor clamped bridge inverter.

Table 15.6: Five-level flying-capacitor inverter output states (phase A to R)

mode	V_{AR}	switching states				capacitors			paths
		T_1	T_2	T_3	T_4	C_1	C_2	C_3	
1	$\frac{1}{2}V_s$	1	1	1	1	=	=	=	$\frac{1}{2}V_s$
2 N-1 states	$\frac{1}{4}V_s$	1	1	1	0	=	=	+	$\frac{1}{2}V_s$ $-V_{C3}$
		1	1	0	1	=	+	-	$\frac{1}{2}V_s$ $-V_{C2}+V_{C3}$
		1	0	1	1	+	-	=	$\frac{1}{2}V_s-V_{C1}+V_{C2}$
		0	1	1	1	-	=	=	$-\frac{1}{2}V_s+V_{C1}$
3 N^2-4N+1 states	0	1	1	0	0	=	+	=	$\frac{1}{2}V_s$ $-V_{C2}$
		1	0	1	0	+	-	+	$\frac{1}{2}V_s-V_{C1}+V_{C2}$ $-V_{C3}$
		0	1	1	0	-	=	+	$-\frac{1}{2}V_s+V_{C1}-V_{C3}$
		1	0	0	1	+	=	-	$\frac{1}{2}V_s-V_{C1+}-V_{C3}$
		0	1	0	1	-	+	-	$-\frac{1}{2}V_s+V_{C1}-V_{C2}+V_{C3}$
		0	0	1	1	=	-	=	$-\frac{1}{2}V_s$ $+V_{C2}$
4 N-1 states	$-\frac{1}{4}V_s$	1	0	0	0	+	=	=	$\frac{1}{2}V_s-V_{C1}$
		0	1	0	0	-	+	=	$-\frac{1}{2}V_s+V_{C1}-V_{C2}$
		0	0	1	0	=	-	+	$-\frac{1}{2}V_s$ $-V_{C2}$ $-V_{C3}$
		0	0	0	1	=	=	-	$-\frac{1}{2}V_s$ $+V_{C3}$
5	$-\frac{1}{2}V_s$	0	0	0	0	=	=	=	$-\frac{1}{2}V_s$

15.3.3 Cascaded H-bridge multilevel inverter

The *N*-level cascaded H-bridge, multilevel inverter comprises ½(*N*-1) series connected single-phase H-bridges per phase, for which each H-bridge has its own isolated dc voltage source. For each bridge, as shown in table 15.7, three output voltages are possible, $\pm V_s$, and zero, giving a total number of states of $3^{\frac{1}{2}(N-1)}$, where *N* is odd. Figure 15.37 shows one phase of a seven-level cascaded H-bridge inverter.

The cascaded H-bridge multilevel inverter is based on multiple two level inverter outputs (each H-bridge), with the output of each phase shifted. Despite four diodes and switches, it achieves the greatest number of output voltage levels for the fewest switches.

Its main limitation lies in the need for isolated power sources for each H-bridge and for each phase, although for VA compensation, capacitors replace the dc voltage supplies, and the necessary capacitor energy is only to replace losses due to inverter losses. Its modular structure of identical H-bridges is a positive design feature.

The number of levels in the line-to-line voltage waveform will be

$$k = 2N - 1 \tag{15.89}$$

while the number of levels in the line to load neutral of a star (wye) load will be

$$p = 2k - 1 \tag{15.90}$$

The number of capacitors or isolated supplies required per phase is

$$N_{cap} = \tfrac{1}{2}(N-1) \tag{15.91}$$

The number of possible switch states is

$$n_{states} = N^{phases} \tag{15.92}$$

and the number of switches in each leg is

$$S_n = 2(N-1) \tag{15.93}$$

Table 15.7: Three output states of H-bridges and their current paths.

V_ℓ	On switches	Bidirectional current paths	
		$+ i_L$	$- i_L$
V_s	T_2 T_3	T_2 T_3	D_2 D_3
0	none	D_4 D_1	D_2 D_3
$-V_s$	T_1 T_4	T_1 T_4	D_2 D_3

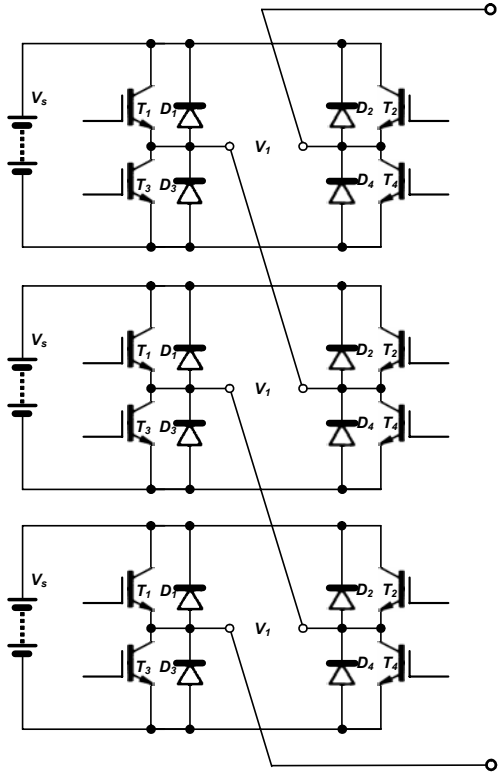
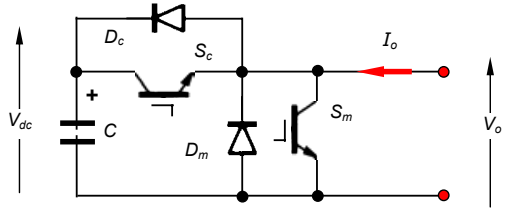


Figure 15.37. *One leg of a voltage-source, seven-level, cascaded H-bridge inverter.*

15.3.4 Capacitor clamped multilevel inverter

The capacitor-clamped multilevel inverter has several advantages compared to conventional multilevel inverters, such as: modular construction; can be extended to any number levels; capacitor voltage balance is attainable for any number of voltage levels; for a large number of voltage levels extremely low total harmonic distortion can be achieved without the need for filters. In addition, to reduced voltage stress on switching device *dv/dt*; and it has failure management capability in the case of device failures.
Figure 15.38 shows one cell of a capacitor-clamped multilevel converter, when the switching device S_m is turned on and S_c is turned off, the voltage $V_o = 0$; when the switching device S_m is turned off and S_c is turned on, the voltage $V_o = V_{dc}$. Table 15.8 and figure 15.38b summarizes the switching states of a cell shown in Fig. 15.38 and their influence in capacitor voltage. Switches S_m and S_c are complementary, when any one of them is on.
Figure 15.39 shows one-phase of the three-level capacitor-clamped inverter in which each voltage level can be synthesized by turning four switching devices simultaneously; in each instant two switches must belong to (S_{y1}, S_{y2}, S_{y3} and S_{y4}) and the remaining two from auxiliary switches (S_{x1}, S_{x2}, S_{x3} and S_{x4}). There are four complementary switch pairs in each phase, turning on one of the pair switches will excluded the other from being turned on. The four complementary switches are (S_{y1}, S_{x1}), (S_{y2}, S_{x2}), (S_{y3}, S_{x3}) and (S_{y4}, S_{x4}). For a dc bus of V_{dc}, the voltage across each cell capacitor is $\frac{1}{2}V_{dc}$ and each switching device voltage stress is limited to one capacitor voltage. For an *n*-level converter, the voltage across each capacitor and switching device is limited to $V_{dc}/(n-1)$ while the number of switching devices (IGBT plus free wheeling diode) required per phase is doubled that for multipoint clamped converters. The number of capacitors required for three-phases is *6n-6* (*2n-2* per phase), while no clamping diodes are required. To explain how the multilevel waveform voltage is synthesized, the supply mid point is assumed the output voltage reference. Using the three-level converter circuit shown in Figure 15.39 as an example, there are six switching combinations to synthesize three-level voltages between *x* and *o*.

1. For voltage level $V_{xo} = \frac{1}{2}V_{dc}$, turn on all the upper main switches (S_{y1} and S_{y2}) and all the lower auxiliary switches (S_{x3} and S_{x4}).
2. For voltage level $V_{xo} = 0$, there are four different switch combinations:
   a) Turn on S_{y1}, S_{y3}, S_{x2} and S_{x4}.
   b) Turn on S_{y2}, S_{y3}, S_{x1} and S_{x4}.
   c) Turn on S_{y2}, S_{y4}, S_{x1} and S_{x3}.
   d) Turn on S_{y1}, S_{y4}, S_{x2} and S_{x3}.
3. For voltage level $V_{xo} = -\frac{1}{2}V_{dc}$, turn on all the upper auxiliary switches (S_{x1} and S_{x2}) and all the lower main switches (S_{y3} and S_{x4}).



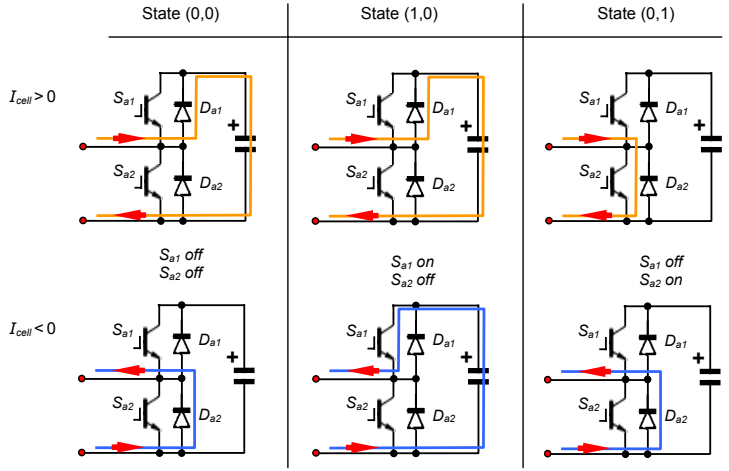


Figure 15.38. *Structure of one cell in capacitor-clamped multilevel converter and its current direction dependant conduction states.*

Table 15.8: Switching states of one cell

S_m	S_c	V_o	Current direction	Power path	Capacitor state
ON	OFF	0	$i_o > 0$	S_m	unchanged
ON	OFF	0	$i_o < 0$	D_m	unchanged
OFF	ON	V_{dc}	$i_o > 0$	D_c	charging
OFF	ON	V_{dc}	$i_o < 0$	S_c	discharging

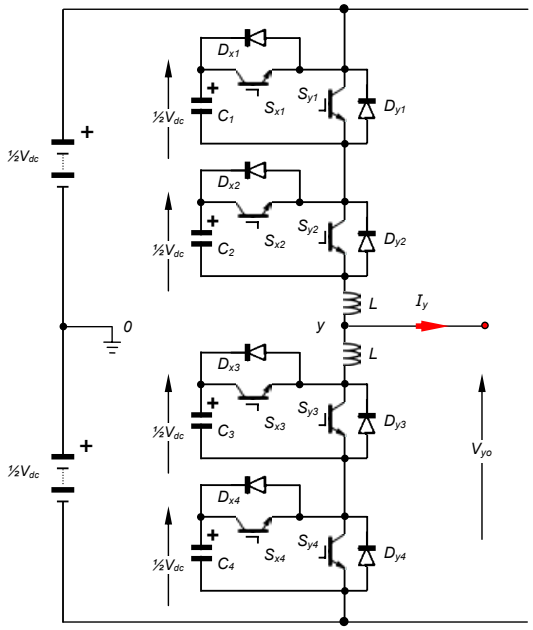


Figure 15.39. *Schematic for one-leg of a three-level capacitor-clamped multilevel inverter.*

Table 15.9 lists the voltage levels and their corresponding switch states. State condition 1 means the switch is on, and 0 means the switch is off. In order to maintain equal voltage stressing on the switching devices, the voltage across each cell capacitor must be maintained at $½V_{dc}$.

Table 15.9: Switching combinations for three-level capacitor-clamped converter

output voltage V_{xo}	switch states								
	S_{y1}	S_{y2}	S_{y3}	S_{y4}	S_{x1}	S_{x2}	S_{x3}	S_{x4}	
$½V_{dc}$	1	1	0	0	0	0	1	1	
0	1	0	1	0	0	1	0	1	(a)
	0	1	1	0	1	0	0	1	(b)
	0	1	0	1	1	0	1	0	(c)
	1	0	0	1	0	1	1	0	(d)
$½V_{dc}$	0	0	1	1	1	1	0	0	

Multilevel topology comparison

A comparison between the three basic multilevel inverters is possible from the numerical summary of component numbers for each inverter, as in Table 15.10.

The diode clamped inverter requires many clamping diodes; the flying capacitor inverter requires many independent capacitors; while the cascaded inverter requires many isolated dc voltage power supplies.

Table 15.10: Multilevel inverter component count, per phase

Inverter type * either /or	levels			switches & // diodes	diodes clamping	flying capacitors	level capacitors	isolated supplies	Outer hex redundant states
	$V_{A\text{-}0V}$	$V_{A\text{-}B}$	$V_{A\text{-}N}$						
diode clamped	*N*	*2N-1*	*4N-3*	*2(N-1)*	*(N-1)(N-2)*	0	*(N-1)*	0	0
flying capacitor	*N*	*2N-1*	*4N-3*	*2(N-1)*	0	*½(N-1)(N-2)*	*(N-1)*	0	yes
cascade	*N (odd)*	*2N-1*	*4N-3*	*2(N-1)*	0	0	*½ (N-1)**	*½(N-1)**	n/a
capacitor clamped	*N*	*2N-1*	*4N-3*	*2(N-1)*	0	0	*2(N-1)*	0	yes

15.3.5 PWM for multilevel inverters

Two basic approaches can be used to generate the necessary pwm signals for multilevel inverters. Each approach is based on the extension of a two level equivalent.

- Modulating waveform comparison with offset triangular carriers
- Space vector modulation based on a rotating vector in multilevel space

15.3.4i - Multiple offset triangular carriers

Various sinusoidal pwm techniques were considered in sections 15.1.3v and 15.1.3vi of this chapter. Figure 15.40 shows how a triangular carrier is associate with each complementary switch pair, four carriers (*N*-1) for the five-level inverter as illustrated. The parts of figure 15.40 show how the four individual carriers can be displaced with respect to one another. The figure also shows how triplen injection is incorporated. The appropriate five-level switch states, as in tables 15.4 to 15.6, can be used to decode the necessary switching sequences. To minimise losses, switching is restricted to only occur between adjacent levels.

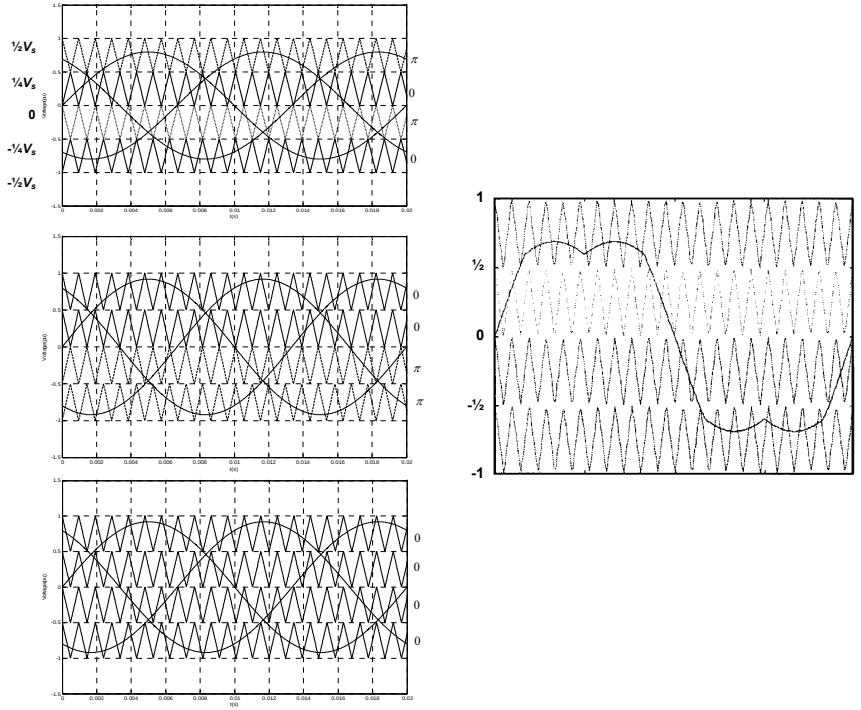


Figure 15.40. *Multi-carrier based pwm generation for a voltage-source, 5-level, inverter.*

15.3.4ii - Multilevel rotating voltage space vector

Space vector modulation for the two-level inverter was considered in section 15.1.3vi of this chapter. The basic hexagon shape for two levels is extended to higher levels as shown in figure 15.41, for three levels. The number of triangles, vectors, and states increases rapidly as the level number increases.

Table 15.11: Properties of *N*-level vector spaces

levels	states	triangles	vectors	vectors in each hexagon
N	N^3	$6(N-1)^2$	$3N(N-1)+1$	
2	8	6	7	(1+6)
3	27	24	19	(1+6)+12
5	125	96	61	(1+6)+12+18+24

From Table 15.11, the states for the two and three level inverters can be specified as follows.

The 2-level inverter

The zero state matrix is

$$[000 \quad 111]$$

The first and only hexagon is shown in figure 15.23a.

$$[100 \quad 110 \quad 010 \quad 011 \quad 001 \quad 101]$$

The three level inverter

The zero state matrix is

$$[000 \quad 111 \quad 222]$$

The first hexagon matrix is

$$\begin{bmatrix} 100 & 110 & 010 & 011 & 001 & 101 \\ 211 & 221 & 121 & 122 & 112 & 212 \end{bmatrix}$$

The second hexagon matrix is

$$[200 \quad 210 \quad 220 \quad 120 \quad 020 \quad 021 \quad 022 \quad 012 \quad 002 \quad 102 \quad 202 \quad 201]$$

These pole states are shown figure 15.41.

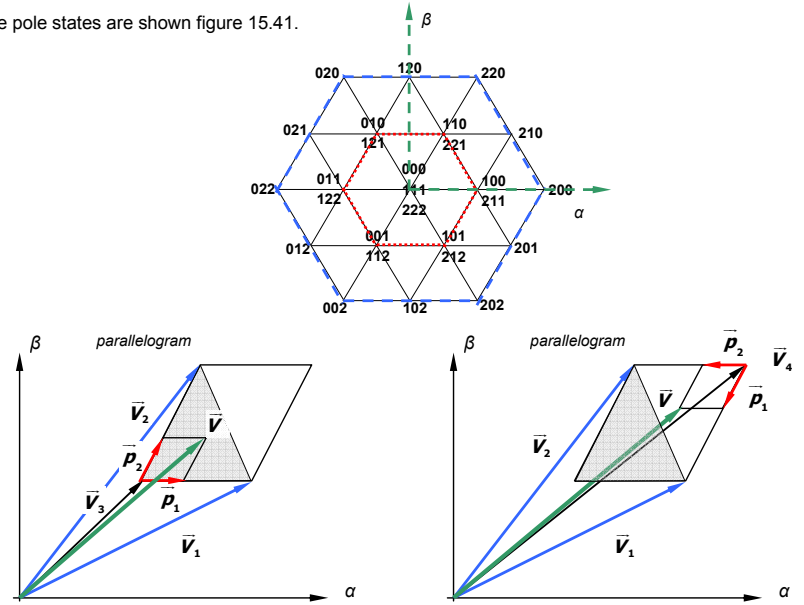


Figure 15.41. *Rotating voltage space vector approached applied to three phases of a voltage-source three-level, inverter and decomposition of the vector in a given parallelogram.*

A `0´ represents the minimum voltage obtainable from the multilevel converter and *N-1* represents the maximum value. For example, in a two-level converter, `0´ is equivalent to 0V and `1´ is equivalent to $V_s$, where $V_s$ is the converter DC link voltage. In a three-level converter `0´ is equivalent to $-½V_s$, `1´ is equivalent to 0 V, and '2' is equivalent to $½V_s$ where V_s is the dc link voltage of the multilevel converter.
When the rotating vector is drawn in the vector space, it is decomposed into vectors bordering the triangle it lies in. When operating in the outer hexagon, the vectors states used in the inner most hexagon mean that that level of the converter is operating with a six-step quasi-square output voltage waveform, to which is added a modulated square waveform for the next higher level.

Generally, the rotating vector can lay in one of two triangles of any parallelogram. Once the parallelogram has been uniquely decoded, for the triangle nearer the origin:

$$\begin{aligned} \vec{V} &= p_1\left(\vec{V}_1 - \vec{V}_3\right) + p_2\left(\vec{V}_2 - \vec{V}_3\right) + \vec{V}_3 = p_1\vec{V}_1 + p_2\vec{V}_2 + (1 - p_1 - p_2)\vec{V}_3 \\ &= \vec{p}_1 + \vec{p}_2 + \vec{V}_3 \end{aligned} \tag{15.94}$$

while for the triangle further from the origin:

$$\begin{aligned} \vec{V} &= p_1\left(\vec{V}_1 - \vec{V}_4\right) + p_2\left(\vec{V}_2 - \vec{V}_4\right) + \vec{V}_4 = p_1\vec{V}_1 + p_2\vec{V}_2 + (1 - p_1 - p_2)\vec{V}_4 \\ &= \vec{p}_1 + \vec{p}_2 + \vec{V}_4 \end{aligned} \tag{15.95}$$

where p_1 and p_2 are the relative duration lengths of the active vectors V_1 and V_2.
The zero vectors, V_3 and V_4, constitute the remaining time in the carrier period not assigned to the V_1 and V_2 associated duration components. The periods associated with p_1V_1 and p_2V_2 are distributed within the carrier period as for to level SVM in section 15.1.3vii-2.
Since SVM decomposes a single rotating vector into three identical components displaced by 120°, such a modulation strategy may not be applicable to any inverter used for FACTS type applications since phases may be unbalanced, distorted or phase shifted.

15.4 Reversible dc link converters

Power inversion by phase angle control is attained with a fully controlled single-phase converter as discussed in section 12.2.3. Power regeneration is also possible with the fully controlled three-phase converter shown in figure 12.11. If a fully controlled converter supplies a dc machine, two-quadrant control is possible (QI and QIV), motoring in one direction of rotation and generating in the other direction. Power regeneration into the supply is achieved by reversing the dc output voltage by controlling the converter phase delay angle. The converter current is uni-directional, that is, the converter output current can not reverse.
The dual or double converter circuit in figure 15.42a and b will accommodate four-quadrant dc machine operation, where the circuit performs as two fully controlled converters in anti-parallel. Each converter is able to rectify and invert, but because of their inverse parallel connection, one converter (the positive converter P) operates in quadrants QI and QIV, while the other (the negative converter N) operates in quadrants QII and QIII, as shown in figure 15.43.
The two converters can be operated synchronously, called *simultaneous control* or independently where one is always blocking, called *independent control*.

15.4.1 Independent control

Simultaneous converter control can be used if continuous load current can be guaranteed. Otherwise only one converter, depending on the quadrant, need operate at anyone time (the other is in a blocking state), as shown in figure 15.42a. No circulating currents arise due to possible mismatched N and P converter output voltages. The continuous current condition may be difficult to ensure at light load levels. Additional series armature inductance, *L* in figure 15.42a and b, helps with current smoothing and ensuring continuous machine current.

A machine rotational direction change is affected by the following converter operating procedure.

- Initially the motor is operating in quadrant I, with $0° \le \alpha_1 \le 90°$ for the positive converter P. The negative converter, N, is in the fully blocking state, with all thyristors turned off.
- The positive converter is put into the inverting mode with $90° \le \alpha_1 \le 180°$, changing the average output voltage from positive to negative. The machine current rapidly falls to zero. The machine rotational speed slows, the rate depending on the load inertia.
- After a dead time, the positive converter blocks and the negative converter N starts in a motor braking mode in quadrant II. The motor speed falls rapidly to zero.
- The second converter operates in quadrant III and rapidly accelerates the motor in the opposite direction, with $0° \le \alpha_2 \le 90°$.

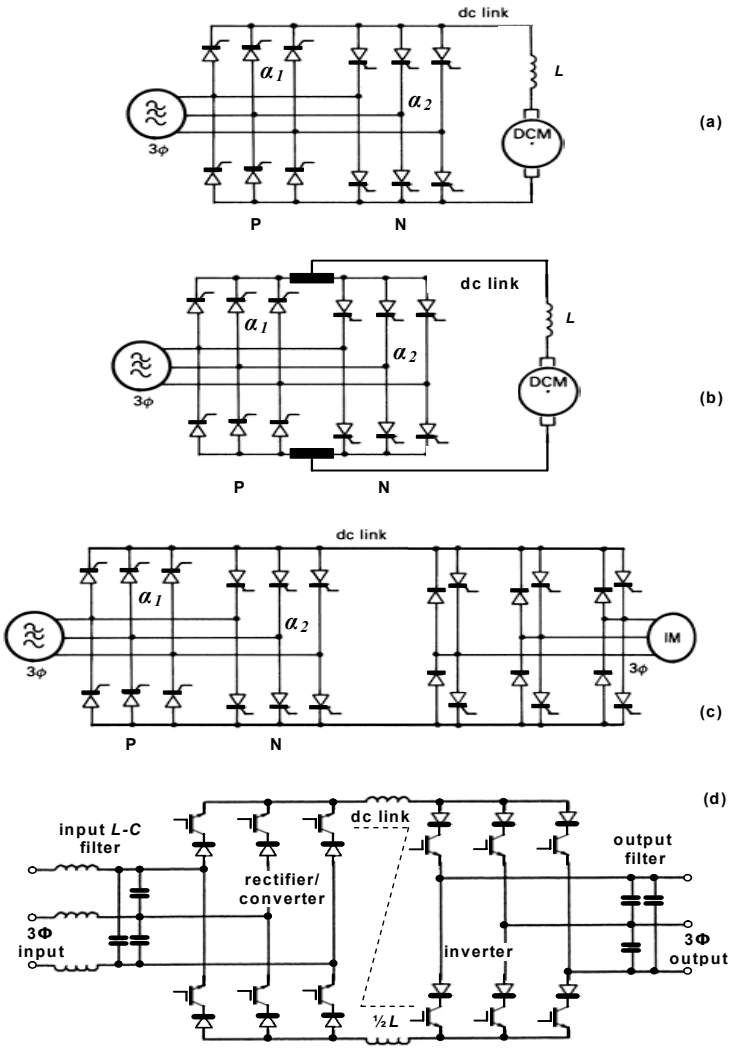


Figure 15.42. *Reversible converter allowing four-quadrant control of: (a) a dc machine with independent converters; (b) a dc machine with simultaneously controlled converters; and (c) voltage and (d) current fed induction machine.*

The dead time before turning on the negative converter N is to ensure the positive converter P is fully off, otherwise the three-phase input voltage lines may short through the two converters. Such a current condition cannot be controlled with line-commutated thyristors. Operation is characterised by transitions from QI to QII to QIII for reversal, and transitions from QIII to QIV to QI for returning to the original direction of rotation.

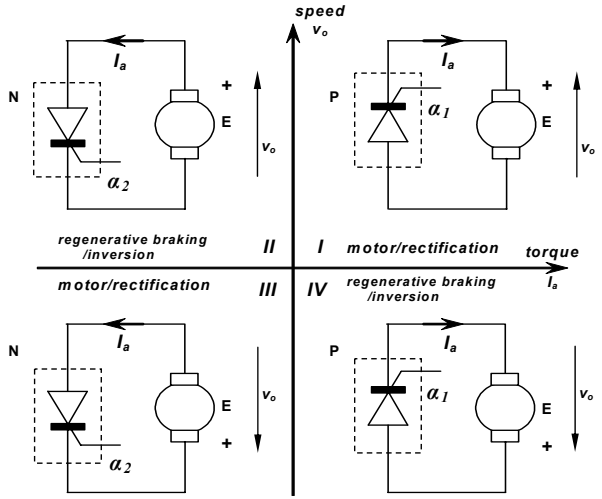


Figure 15.43. *Four quadrants of reversible converter operation.*

15.4.2 Simultaneous control

Simultaneous converter control, also called circulating current control, functions with both converters always in operation which gives a faster dynamic response than when the converters are used mutually exclusively. To avoid supply short circuits requires that the output voltage of both converters (rectifier V_r and inverter V_i) be the same in order to minimise circulating currents.

$$\begin{aligned} \overline{V}_r + \overline{V}_i &= 0 \\ V\cos\alpha_1 + V\cos\alpha_2 &= 0 \\ \cos\alpha_1 + \cos\alpha_2 &= 0 \\ \text{that is } \alpha_1 + \alpha_2 &= 180° \end{aligned} \qquad (15.96)$$

Equation (15.96) implies that both converters operate with firing angles that sum to 180°. Each converter produces the opposite polarity output voltage, which is cancelled by reversing the relative output connections. Under such conditions the load current can be maintained continuous. To minimize any circulating current due to ripple voltage produced by instantaneous voltage differences between the two converters, inductance is usually inserted between each converter and the dc machine load, as shown in figure 15.42b. Adversely the cost and weight are increased, and the supply power factor and drive efficiency are decreased, compared to that obtained with independently controlled converters.

A machine rotational direction change is affected by the following converter operating procedure.

- Initially the motor is operating in quadrant I for the rectifying, positive converter, with $0° \le \alpha_1 \le 90°$. The other converter is operating in the inverting mode with $90° \le \alpha_2 \le 180°$, such that $\alpha_1 + \alpha_2 = 180°$. The output voltage for both converters is the same, and the negative converter N carries only the circulating current.
- For rotational direction reversal, $\alpha_1 \ge 90°$ and $\alpha_2 \le 90°$, such that $\alpha_1 + \alpha_2 = 180°$. The armature back emf voltage now exceeds the converter output voltages, and current diverts to the negative converter N and the machine regeneratively brakes, operating in quadrant II. The current rapidly falls to zero and the positive converter P carries only the ac circulating current.
- The speed rapidly falls to zero, with $\alpha_1 = \alpha_2 = 90°$ giving zero output voltage, so as to control the armature current since the back emf is zero. Then with $\alpha_2 < 90°$ the machine rapidly accelerates in quadrant III, in the reverse direction to the original rotation.

For reversing the direction of rotation from Q III the operation sequence is QIII to QIV to QI. Since no converter dead time is introduced, a fast dynamic response can be attained. A small dc circulating current is deliberately maintained, that is greater in magnitude than the peak of the ac ripple current. The ac current can then flow continuously in both converters, both of which can operate in the continuous conduction mode without the need for continuous converter current reversal operation.

15.4.3 Inverter regeneration

The bridge freewheel diodes of a three-phase inverter restrict the dc rail or dc link voltage from reversing. The dual or double converter circuit in figure 15.42c will allow inversion with a three-phase voltage source inverter. One converter rectifies, the other converter inverts, functioning as a self-commutated inverter, transferring power from the dc link to the ac supply. Complete four-quadrant control of the three-phase ac machine on the inverter is achieved in conjunction with control of the dc to ac inverter. That is, motor reversal is achieved by effectively interchanging the pwm control signals associated with two phases. The real power flow back into the ac supply is controlled by the converter phase delay angle, while the reactive power flow is controlled by the voltage magnitude. The angle and voltage are not independent. In the case of a pwm controlled inverter fed ac machine, the ac to dc converter can be uncontrolled, using all diodes, since dc output voltage reversal is not utilised.

Figure 15.42d shows a fully reversible current controlled converter/inverter configuration, using self-commutating devices. The use of self-commutated switches (rather than mains commutated converter thyristors) offers the possibility to minimise the input current distortion and to reduce the inductor size hence improve the dynamic current response. The switch series diodes are essential since the shown IGBTs have no useable reverse blocking capability. The use of reverse blocking GCTs avoids the need for the series blocking diodes, which reduces the on-state voltage losses but increases gate drive complexity and power rating. Series connection of devices is necessary above a few kV, and above 1 MVA the GCT dominates.

15.5 Standby inverters and uninterruptible power supplies

Standby inverters and uninterruptible power supplies (UPS's) provide a 50/60 Hz supply in the event of an ac mains failure. A UPS must provide ac output such that mains failure is undetected by the load. To achieve this, a UPS continually feeds the load from an inverter. A load that can tolerate a short interruption of the ac supply is fed from a *standby inverter* which becomes operational within 1-5 ms after the ac supply failure. In communications, computing, and automated production lines, UPS's are essential for even brownouts (V and f outwith bounds for reliable equipment operation), while in lighting and heating applications, standby inverters are used since a few missing ac cycles (due to a blackout – total interruption of the mains power) may be tolerated. In each power supply case, the alternative energy source is a standby dc battery. The UPS keeps the battery charged when the ac input is supplying the output power.

15.5.1 Single-phase UPS

A basic single-phase UPS is shown in figure 15.44. A key safety objective is to retain the supply neutral at both the supply input and the ac output, without resorting to any from of isolating transformer. Consequently, the input ac mains is half-wave rectified by diodes D_R^+ and D_R^-. Boost converters on the positive and negative groups ensure supply sinusoidal input current and unity power factor. The output H-bridge (T_1-T_4) uses pwm and feedback control to produce a fixed frequency and magnitude output (and ac mains phase synchronisation if required), which is filtered by an *L-C* filter. In the event of a loss of the ac supply, the backup batteries, V^+ and V^-, provide energy to the boost converters, hence to the output inverter. The battery backup voltage magnitude is much less than the ac supply magnitude and diodes, D_B^+ and D_B^-, isolate the batteries from the rectified ac supply voltage. The shown UPS has two basic limitations that manufactures strive to overt.

- If the battery is to be connected to neutral, then two batteries are necessary. Proprietary attempts using only one battery involve circuit complications and limitations. At best, with one battery, it is one forward biased diode voltage drop from neutral.
- Because the batteries supplies are not isolated during normal operation, during part of the mains cycle near zero voltage, the batteries alternately provide energy. This decreases their lifetime and necessitates more complicated trickle charge circuits. The input current is also distorted at the 0V crossover. Replacement of the blocking diodes D_B by switches involves complexity and battery backup operation requires detection and is not fail safe.

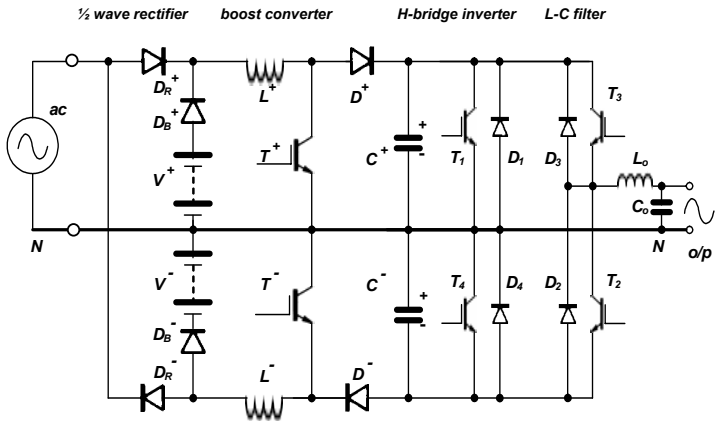


Figure 15.44. *Single-phase uninterruptible power supply.*

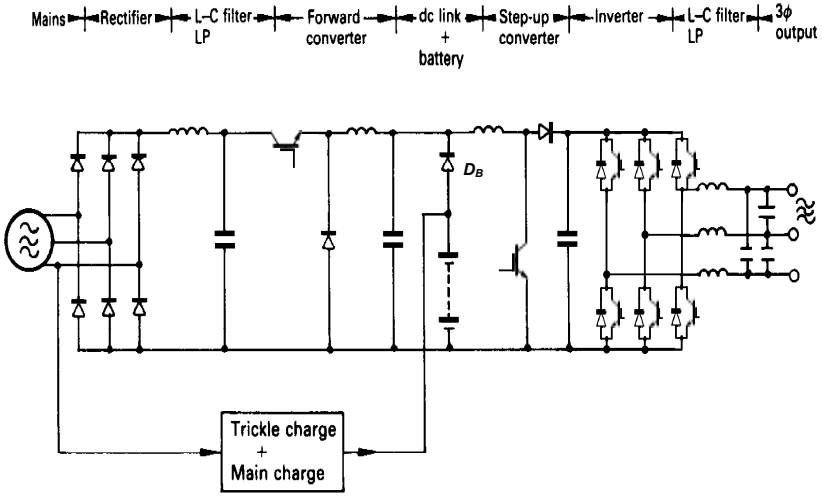


Figure 15.45. *Three-phase uninterruptible power supply.*

15.5.2 Three-phase UPS

Figure 15.45 shows a basic three-phase UPS, used up to a few tens of kilowatts. The ac supply is rectified and filtered. A forward converter controls the dc link voltage to just above the battery voltage level. This dc voltage is boosted to a dc level such that after inversion it provides the required output voltage magnitude. If the input ac fails or droops, the dc link power is provided by the battery via diode D_B. The output inverter is usually operational in a pwm mode, which allows precise frequency control, voltage control, ac mains phase synchronisation, and minimisation of low frequency output harmonics. With pwm control minimal filtering is required, which minimises the filter weight, cost, size, and losses. A three-phase UPS can utilise third harmonic injection (15.1.4(iv)).
A three-phase boost input converter can be used to maintain sinusoidal ac supply input currents at unity power factor.

15.6 Power filters

Power *L*-*C* filters are used to reduce harmonics or ripple from

- the rectifier output (dc filter)
- the inverter output (ac filter).

L-*C* low-pass, second-order filters are shown in figures 15.42, 15.44, and 15.45. In figure 15.45, the *L*-*C* smoothing filter at the rectifier output, filters the ac mains frequency components leaving dc. The same type of filter is used in the inverter output to filter pwm harmonics, leaving the relative low frequency modulation frequency.
The *L*-*C* filter fundamental cut-off frequency is dependent on *L, C,* and the load impedance Z_L

$$\frac{v_o}{v_i} = \frac{1}{1 + j\omega L\left(\frac{1}{Z_L} + j\omega C\right)} = \frac{1}{1 - \omega^2 LC + j\frac{\omega L}{Z_L}} \qquad (15.97)$$

The simplest design approach is to assume a non-load condition, $Z_L \rightarrow \infty$, whence the filter cut-off frequency is $f_o = 1/\left(2\pi\sqrt{LC}\right)$.
Frequency components below f_o, including dc, are passed. Those components above f_o are attenuated by a second order fall-off in gain. Any frequency components inadvertently around the resonant frequency, f_o, will be amplified. For this reason, the filter may be damped with parallel connected *R*-*C* snubbers.

Reading list

See chapter 11 reading list.

Hart, D.W., *Introduction to Power Electronics,*
Prentice-Hall, Inc, 1994.

Mohan, N., *Power Electronics,* 3rd Edition,
Wiley International, 2003.

Problems

15.1. The inverter in figure 15.7 is supplied from a 340 V dc source. The load has a resistance of 10Ω and an inductance of 10mH. The basic operating frequency is 50 Hz, with three notches per half cycle giving half the maximum output, similar to that shown in figure 15.13.
Determine the load current waveform over the first two cycles and determine the power delivered to the load based on the current waveform of the final half cycle.

15.2. The inverter and load in problem 15.1 are controlled so as to eliminate the third and fifth harmonics in the output voltage.
Determine the load current waveform over the first two cycles and the power delivered to the load based on the current waveform of the last half cycle.

15.3. Output voltage harmonic reduction can be achieved by employing multiphase, selected notching modulation control on a three-phase bridge as discussed in 15.1.4. An output as in figure 15.14b with α_1 = 16.3° and β_1 = 22.1° eliminates the 5th and 7th harmonics.
Determine the fundamental voltage output component and compare it with that of a square wave. Determine the output rms voltage.

15.4. With the aid of figure 15.11 determine the line-to-neutral and line-to-line output voltage of a dc to three-phase inverter employing 120° device conduction.
Calculate the interphase:
i. mean half-cycle voltage
ii. rms voltage
iii. rms voltage of the fundamental.

15.5. The three-phase inverter bridge in figure 15.4 has a 600V dc rail and a 10Ω per phase load. For 180° and 120° conduction calculate:
i. the rms phase current
ii. the power delivered to the load
iii. the switch rms current.
[24.5 A, 18 kW, 17.3 A; 28.3 A, 24 kW, 14.15 A]

15.6 A single-phase square-wave inverter is supplied from a 340V dc source and the load is a 17Ω resistor. Determine switch average and rms current ratings. What power is delivered to the load?

15.7 A single-phase square-wave inverter is supplied from a 340V dc source and the series *R*-*L* load is a 20Ω resistor and *L*=20mH. Determine:
i. an expression for the load current, hence the maximum switch current
ii. rms load current
iii. average and rms switch current
iv. maximum switch voltage
v. average source current, hence power delivered to the load
vi. load current total harmonic distortion.

16

DC to AC Inverters – Resonant Mode

Inversion (in this chapter) is the conversion of dc power to ac power at a desired output voltage or current and frequency. A static semiconductor inverter circuit performs this electrical energy inverting transformation. The terms voltage-source and current-source are used in connection with the output from inverter circuits.

A *voltage-source inverter (VSI)* is one in which the dc input voltage is essentially constant and independent of the load current drawn. The inverter specifies the load voltage while the drawn current shape, near sinusoidal, is dictated by the series resonant load, in this case.

A *current-source inverter (CSI)* is one in which the source, hence the load current is predetermined and the load impedance, a parallel resonant circuit in this case, determines the, near sinusoidal output voltage. The supply current cannot change quickly. This current is controlled by series dc supply inductance which prevents sudden changes in current. Being a current source, the inverter can survive an output short circuit thereby offering fault ride-through properties.

Inverter switching losses (either turn-on or turn-off) can be significantly reduced if zero current or voltage switching can be utilised. This switching loss reduction allows higher operating frequencies hence smaller *L* and *C* components (in size, weight, and value). Also radiated switching noise is significantly reduced.

Two main techniques can be used to achieve near zero switching losses

- a resonant load that provides natural voltage or current zero instances for switching
- a resonant circuit across the switch which feeds energy to the load as well as introducing zero current or voltage instances for switching.

The inverter and its output are single-phase and the output is controlled around the load resonant frequency. Zero current, ZCS, and zero voltage, ZVS, switching occurs when the inverter switches are operated either side of resonance.

16.1 Resonant dc-ac inverters

The voltage source inverters considered in 15.1 involve inductive loads and the use of switches that are hard switched. That is, the switches experience simultaneous maximum voltage and current during turn-on and turn-off with an inductive load. The current source inverters considered in 15.2 required capacitive circuits to commutate the bridge switches. When self-commutatable devices are used in current source inverters, hard switching occurs. In resonant inverters, the load enables commutation of the bridge switches with near zero voltage or current switch conditions, resulting in low switching losses.

A characteristic of *L*-*C*-*R* resonant circuits is that at regular, definable instants

- for a step load voltage, the series *L*-*C*-*R* load current sinusoidally reverses or
- for a step load current, the parallel *L*-*C*-*R* load voltage sinusoidally reverses.

If the load can be resonated, as considered in chapter 6.2.3, then switching stresses can be significantly reduced for a given power through put, provided switching is synchronised to the *V* or *I* zero crossing.

Three types of resonant converters utilise zero voltage or zero current switching.

- load-resonant converters
- resonant-switch dc-to-dc converters
- resonant dc link and forced commutated converters

The single-phase load-resonant converter, which is extensively used in induction heating applications, is presented and analysed in this chapter. Such resonant load converters use an *L*-*C* load which oscillates, thereby providing load zero current or voltage intervals at which the converter switches can be commutated with minimal electrical stress. Resonant switch dc-to-dc converters are presented in chapter 18.

Two basic resonant-load single-phase inverters are used, depending on the *L*-*C* load arrangement:

- current source inverter with a parallel *L*-*C* resonant (tank) load circuit:
  switch turn-off at zero load voltage instants and turn-on with zero voltage
  switch overlap is essential (a continuous source current path is required)
- voltage source inverter with a series connected *L*-*C* resonant load:
  switch turn-off at zero load current instants and turn-on with zero current
  switch under lap is essential (to avoid dc voltage source short circuiting)

Each load circuit type can be fed from a single leg (or arm) circuit or H-bridge circuit depending on the load *Q* factor. This classification is divided according to

- symmetrical full bridge for low *Q* load circuits (class D)
- single bridge leg circuit for a high *Q* load circuit (class E)

High *Q* circuits can also use a full bridge inverter configuration, if desired, for higher through-put power.

In induction heating applications, the resistive part of the resonant load, called the work-piece, is the active load to be heated - melted, where the heating load is usually transformer coupled. Energy transfer control complication is usually associated with the fact that the resistance of the load work-piece changes as it heats up and melts, since resistivity is temperature dependant. However, control is essentially independent of the voltage and current levels and is related to the resonant frequency which is *L* and *C* dependant. Inverter bridge operation is near the load resonant frequency so that the output waveform is essentially sinusoidal. By ensuring operation is below the resonant frequency, such that the load is capacitive, the resultant leading current can be used to self commutate thyristor converters which may be used in high power series resonant circuits. This same capacitive load commutation effect is obtained for parallel resonant circuits with thyristor current source inverters operating just above resonance. The output power is controlled by controlling the converter output frequency, with maximum power being transferred at the resonant frequency.

16.2 L-C resonant circuits

L-*C*-*R* resonant circuits, whether parallel or series connected are characterised by the load impedance being capacitive at low frequency and inductive at high frequency for the series circuit, and vice versa for the parallel case. The transition frequency between being capacitive and inductive is the resonant frequency, ω_o, at which frequency the *L*-*C*-*R* load circuit appears purely resistive and maximum power is transferred to the load, *R*. *L*-*C*-*R* circuits are classified according to circuit quality factor *Q*, resonant frequency, ω_o, and bandwidth, *BW*, for both parallel and series circuits. The characteristics for the parallel and series resonant circuits are related since every practical series *L*-*C*-*R* circuit has a parallel equivalent, and vice versa. The parallel circuit can be series *R*-*L* in parallel with the capacitor *C*.

As shown in figure 16.1 each resonant half cycle is characterised by

- the series resonant circuit current is zero at maximum capacitor stored energy
- the parallel resonant circuit voltage is zero at maximum inductor stored energy

The capacitor in a series resonant circuit must have an external path through which to release its stored energy. The parallel resonant circuit can release its stored inductive energy within its parallel circuit, without an external circuit. The stored energy can internally resonate, transferring energy back and forth between the *L* and *C*, gradually dissipating energy in the circuit *R*, as heat.

16.2.1 - Series resonant *L*-*C*-*R* circuit

The series *L*-*C*-*R* circuit current for a step input voltage V_s, with initial capacitor voltage v_o and series inductor current i_o is given by

$$i(\omega t) = \frac{V_s - v_o}{\omega L} \times e^{-\alpha t} \times \sin \omega t + i_o \times e^{-\alpha t} \times \frac{\omega_o}{\omega} \times \cos(\omega t + \phi) \qquad (16.1)$$

where

$$\omega^2 = \omega_o^2 (1 - \xi^2) = \omega_o^2 - \alpha^2 \qquad \omega_o = \frac{1}{\sqrt{LC}} \qquad \alpha = \frac{R}{2L} \qquad \frac{1}{2Q_s} = \xi = \frac{R}{2\omega_o L} \quad \text{and} \quad \tan\phi = \frac{\alpha}{\omega}$$

ξ is the damping factor. The capacitor voltage is important because it specifies the energy retained in the *L*-*C*-*R* circuit at the end of each half cycle.

$$v_c(\omega t) = V_s - (V_s - v_o)\frac{\omega_o}{\omega} e^{-\alpha t} \cos(\omega t - \phi) + \frac{i_o}{\omega C} e^{-\alpha t} \sin \omega t \qquad (16.2)$$

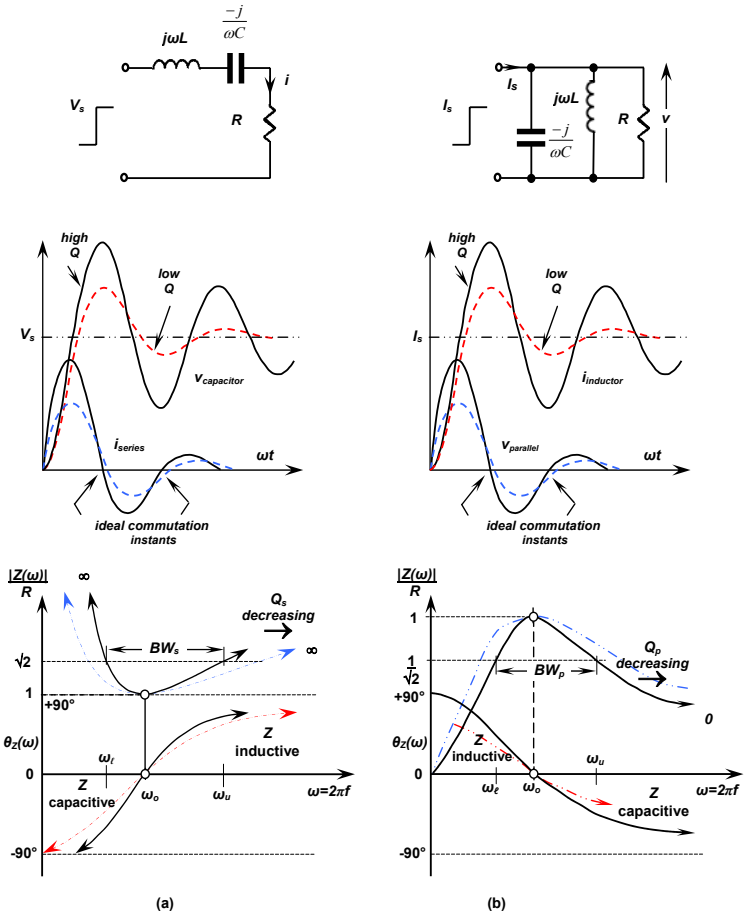


Figure 16.1. *Resonant circuits, step response, and frequency characteristics: (a) series L-C-R circuit and (b) parallel L-C-R circuit.*

At the series circuit resonance frequency ω_o, the lowest possible circuit impedance results, $Z = R$ as shown in figure 16.1a, hence it can be termed, *low-impedance resonance*. The series circuit quality factor or figure of merit, Q_s, is defined by

$$Q_s = \frac{\text{reactive power}}{\text{average power}} = \frac{2\pi \times \text{maximum stored energy}}{\text{energy dissipated per cycle}} = \frac{2\pi \cdot \tfrac{1}{2} L i^2}{\tfrac{1}{2} R i^2 / f_o} = \frac{\omega_o L}{R} = \frac{1}{2\xi} = \frac{Z_o}{R} \qquad (16.3)$$

where the characteristic impedance is

$$Z_o = \sqrt{\frac{L}{C}} \qquad (\Omega)$$

The series circuit half-power bandwidth BW_s is given by

$$BW_s = \frac{\omega_o}{Q_s} = \frac{2\pi f_o}{Q_s} \qquad (16.4)$$

and upper and lower half-power frequencies are related by $\omega_o = \sqrt{\omega_\ell \, \omega_u}$.

$$\omega_\ell^u = \omega_o \pm \alpha$$
$$f_\ell^u = f_o \pm \frac{R}{4\pi L} \qquad (16.5)$$

Figure 16.1a shows the time-domain step-response of the series *L-C-R* circuit for a high *Q* load and a lower *Q* case. In the lower *Q* case, to maintain and transfer sufficient energy to the load *R*, the circuit requires re-enforcement every half sine cycle, while with a high circuit *Q*, re-enforcement is only necessary once per sinusoidal cycle. Thus for a high circuit *Q*, full bridge excitation is not essential, yielding a simpler power circuit as shown in figure 16.2a and b.

Table 16.1: Characteristics and parameters of parallel and series resonant circuits

characteristic		series	parallel
Resonant period/time constant	s	$\tau = \sqrt{LC}$	
Resonant angular frequency	rad/s	$\omega_o = 2\pi f_o = \frac{1}{\tau} = \frac{1}{\sqrt{LC}}$	
Damping factor	pu	$\xi_s = \frac{1}{2}\frac{R}{\omega_o L} = \frac{1}{2}\omega_o C R$	$\xi_p = \frac{1}{2}\frac{\omega_o L}{R} = \frac{1}{2}\frac{1}{\omega_o C R}$
Damping constant	/s	$\alpha_s = \frac{R}{2L}$	$\alpha_p = \frac{1}{2CR}$
Characteristic impedance	Ω	$Z_o = \sqrt{\frac{L}{C}} = \omega_o L = \frac{1}{\omega_o C}$	
Damped resonant angular frequency $\omega = \omega_o\sqrt{1-\xi^2} = \sqrt{\omega_o^2 - \alpha^2}$	rad/s	$\omega = \omega_o\sqrt{1-\xi_s^2}$	$\omega = \omega_o\sqrt{1-\xi_p^2}$
Quality factor $Q_s = \frac{1}{Q_p}$	pu	$Q_s = \frac{1}{2\xi_s} = \frac{Z_o}{R} = \frac{\sqrt{L/C}}{R} = \frac{\omega_o L}{R} = \frac{1}{\omega_o CR} = \frac{2\pi\left(\frac{1}{2}LI_p^2\right)}{\left(\frac{1}{2}RI_p^2\right)\tau}$	$Q_p = \frac{1}{2\xi_p} = \frac{R}{Z_o} = \frac{R}{\sqrt{L/C}} = \omega_o CR = \frac{R}{\omega_o L} = \frac{2\pi\left(\frac{1}{2}CV_p^2\right)}{\left(\frac{1}{2}\frac{V_p^2}{R}\right)\tau}$
Bandwidth	rad/s	$BW_s = \frac{\omega_o}{Q_s}$	$BW_p = \frac{\omega_o}{Q_p}$

The energy transferred to the load resistance *R*, per half cycle $1/2f_r$, is

$$W_{\frac{1}{2}} = \int_0^\pi i(\omega t)^2 R \, d\omega t \qquad (16.6)$$

The active power transferred to the load depends on the repetition rate of the excitation, f_r.

$$P = W_{\frac{1}{2}} \times f_r \qquad (\text{W}) \qquad (16.7)$$

16.2.2 - Parallel resonant *L-C-R* circuit

The load for the parallel case is a parallel *L-C* circuit, where the active load is represented by series resistance in the inductive path. For analysis, the series *L-R* circuit is converted into its parallel *R-L* equivalent circuit, thus forming the equivalent parallel *L-C-R* circuit shown in figure 16.1b. A parallel resonant circuit is used in conjunction with a current source inverter, thus the parallel circuit is excited with a step input current. The voltage across a parallel *L-C-R* circuit for a step input current I_s, with initial capacitor voltage v_o and initial inductor current i_o is given by

$$v(\omega t) = v_c(\omega t) = \frac{I_s - i_o}{\omega C} \times e^{-\alpha t} \times \sin \omega t + v_{co} \times e^{-\alpha t} \times \frac{\omega_o}{\omega} \times \cos(\omega t + \phi) \qquad (16.8)$$

The inductor current is important since it specifies the tank circuit stored energy at the end of each half cycle.

$$i_L(\omega t) = I_s - (I_s - i_o) \times \frac{\omega_o}{\omega} \times e^{-\alpha t} \times \cos(\omega t - \phi) + \frac{v_o}{\omega L} \times e^{-\alpha t} \times \sin \omega t \qquad (16.9)$$

where

$$\alpha = \frac{1}{2CR}$$

The parallel circuit *Q* for a parallel resonant circuit is

$$Q_p = \frac{2\pi \tfrac{1}{2} C v^2}{\tfrac{1}{2} v^2 / R f_o} = \omega_o RC = \frac{R}{\omega_o L} = \frac{R}{Z_o} = \frac{1}{Q_s} \qquad (16.10)$$

where Z_o and ω_o are defined as in equations (16.1) and (16.3), except *L, C,* and *R* refer to the parallel circuit values.
The half-power bandwidth BW_p is given by

$$BW_p = \frac{\omega_o}{Q_p} = \frac{2\pi f_o}{Q_p} \qquad (16.11)$$

and upper and lower half power frequencies are related by $\omega_o = \sqrt{\omega_l \omega_u}$.
At the parallel circuit resonance frequency ω_o, the highest possible circuit impedance results, $Z = R$ as shown in figure 16.1b, hence it can be termed, *high-impedance resonance*.
The energy transferred to the load resistance *R*, per half cycle $1/2f_r$, is

$$W_{\frac{1}{2}} = \int_0^{\pi} v(\omega t)^2 / R \, d\omega t \qquad (16.12)$$

The active power to the load depends on the repetition rate of the excitation, f_r.

$$P = W_{\frac{1}{2}} \times f_r \qquad \text{(W)} \qquad (16.13)$$

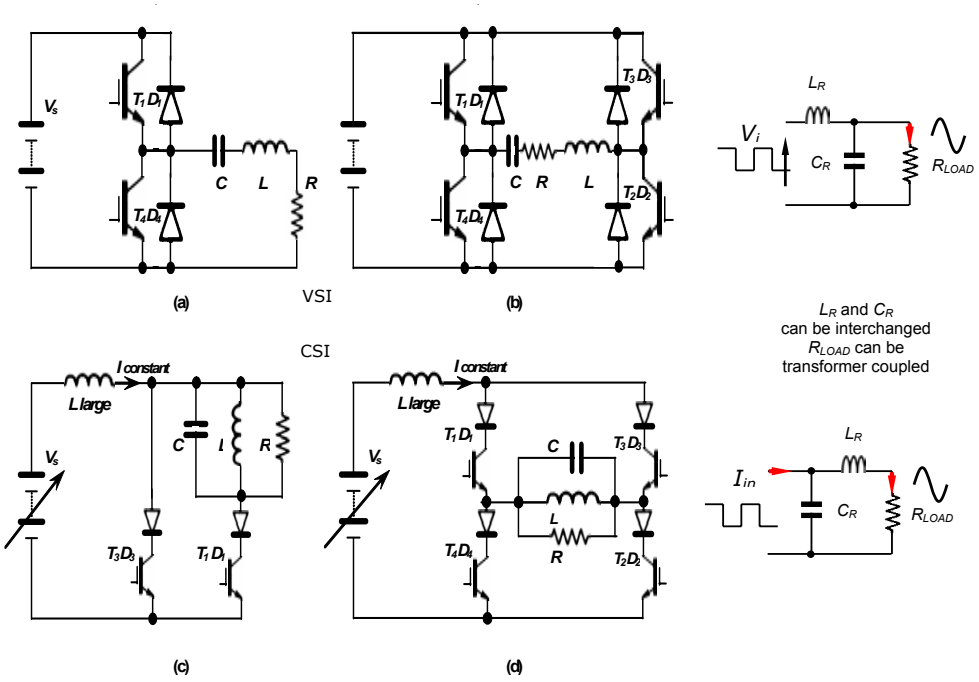


Figure 16.2. *Resonant converter circuits: (a) series L-C-R with a high Q; (b) low Q series L-C-R; (c) parallel L-C-R and high Q; and (d) low Q parallel L-C-R circuit.*

16.3 Series-resonant voltage-source inverters

Series resonant circuits use a voltage source inverter (class D series) as considered in 16.1.1 and shown in figure 16.3a and b. If the load *Q* is high, then the resonance of energy from the energy source, V_s, need only be re-enforced every second half-cycle, thereby simplifying converter and control requirements. A high *Q* circuit is characterised by successive half-cycle capacitor voltage peak magnitudes being of similar magnitude, that is the decay rate is

$$\frac{v_{c_n}}{v_{c_{n+1}}} = e^{\frac{\pi}{2Q}} \approx 1 \quad \text{for } Q \gg 1 \qquad (16.14)$$

Thus there is sufficient energy stored in *C* to be transferred to the load *R*, without need to involve the supply V_s. The circuit in figure 16.3a is simpler and control is easier.

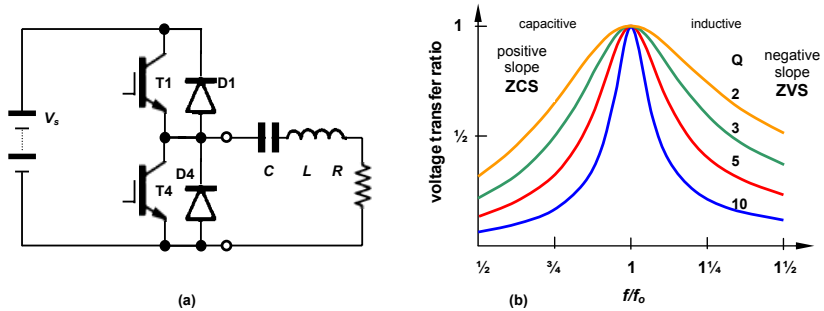


Figure 16.3. *Series resonant voltage source converter: (a) circuit and (b) voltage transfer function.*

Also, for any *Q*, each converter can be used with or without the shown freewheel diodes. Without freewheel diodes, the switches have to block high reverse voltages due to the energy stored by the capacitor. MOSFET and IGBTs require series diodes to achieve the reverse voltage blocking requirements. In high power resonant applications, the reverse blocking abilities of the GTO and GCT make them ideal converter switches. Better load resonant control is obtained if freewheel diodes are not used.

16.3.1 – Series-resonant voltage-source inverter – single inverter leg

Operation of the series load single leg circuit in figure 16.3a depends on the timing of the switches.

1 - Lagging operation (advancing the switch turn-off angle, f > f_o)

If the converter is operated at a frequency above resonance (effected by commutating the switches before the end of an oscillation cycle), the inductor reactance dominates and the load appears inductive. The load current lags the voltage as shown in figure 16.4. This figure shows the conducting devices and that a switch is turned on when its parallel connected diode is conducting. Turn-on therefore occurs at a low voltage (hence low switch turn-on loss and no need for fast recovery diodes), while turn-off (premature) is as with a hard switched inductive load (associated with switch high turn-off loss and turn-off Miller capacitance effects). The turn-off switching loss can be eliminated by adding a shunt capacitor across one of the leg switches and using a dead time between the gate drive voltages.

Operation and switch timing are as follows:

Switch T1 is turned on while its anti-parallel diode D1 is conducting and the current in the diode reaches zero and the current transfers to, and begins to oscillate through the switch T1. The capacitor charges to a maximum voltage and before the current reverses, the switch T1 is hard turned off. The current is diverted through diode D4. T4 is turned on which allows the oscillation to reverse. Before the current in T4 reaches zero, it is turned off and current is diverted to diode D1, which returns energy to the supply. The resonant cycle is repeated when T1 is turned on before the current in diode D1 reaches zero and the process continues.

2 - Leading operation (delaying the switch turn-on angle, f < f_o)

By operating the converter at a frequency below resonance (effectively by delaying switch turn-on until after the end of an oscillation cycle), the capacitor reactance dominates and the load appears capacitive. The load current leads the voltage as shown in figure 16.5. This figure shows the conducting devices and that a switch is turned off when its parallel diode is conducting. Turn-off therefore occurs at a low current, while turn-on (diode reverse recovery) is as with a hard switched inductive load. Fast recovery diodes are therefore essential. Switch output capacitance charging and discharge ($\frac{1}{2}CV^2$) and the Miller effect at turn-on (requiring increased gate power) are factors to be accounted for.

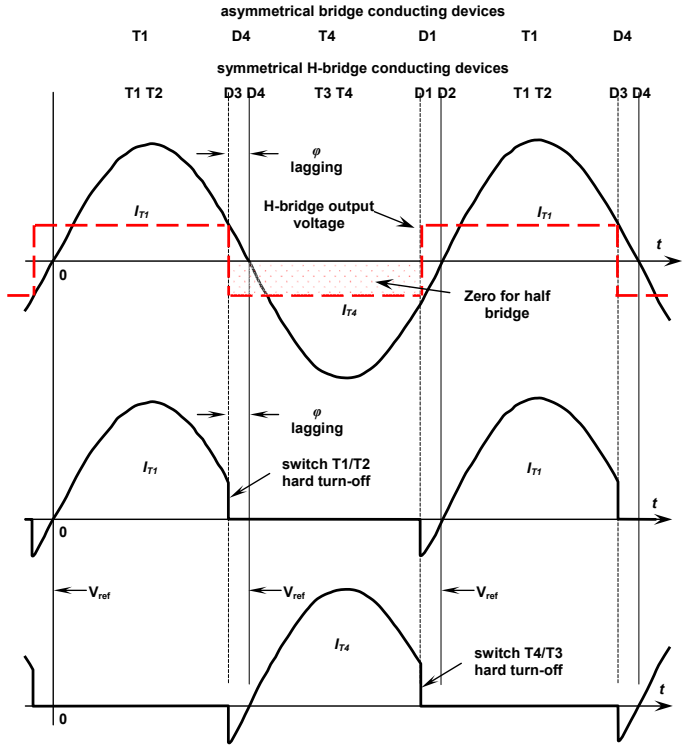


Figure 16.4. *Series L-C-R high Q resonance using the converter circuit in figure 16.2a and b, with f > f_o, a lagging power factor φ.*

Operation and switch timing are as follows:

Diode D4 is conducting when switch T1 is turned on, which provides a step input voltage V_s to the series *L-C-R* load circuit, and the current continues to oscillate. The capacitor charges to a maximum voltage and the current reverses through D1, feeding energy back into the supply. T1 is then turned off with zero current.

The switch T4 is turned on, commutating D1, and the current oscillates through the zero volt loop created through T4 and the load. The oscillation current reverses through diode D4, when T4 is turned off with zero current.

T1 is turned on and the process continues.

Without the freewheel diodes the half oscillation cycles are controlled completely by the switches. On the other hand, with freewheel diodes, the timing of switch turn-on and turn-off is determined by the load current zeros, if maximum energy transfer to the load is to be gained.

Analysis – *single inverter leg – figure 16.3a*

For a square wave input voltage, 0 to V_s, of frequency $\omega \approx \omega_o$, the input voltage fundament of magnitude $2V_s/\pi$ produces the dominant load current component, since higher frequency components are attenuated by second order *L-C* filtering action. That is, the resonant circuit excitation voltage is $|V_i| = 2V_s/\pi$. Key characteristic equations are $\omega_o = 1/\sqrt{LC}$, $Z_o = \sqrt{L/C}$, and $Q = Z_o/R$.

The series circuit steady-state current at resonance for the single-leg half-bridge can be approximated by assuming $\omega_o \approx \omega$, such that in equation (16.1) $i_o = 0$:

$$i(\omega t) = \frac{1}{1-e^{\frac{-\alpha\pi}{\omega}}} \times \frac{V_s}{\omega L} \times e^{-\alpha t} \times \sin \omega t \qquad 0 \le \omega t \le \pi \tag{16.15}$$

which is valid for the + V_s loop (through T1) and zero voltage loop (through T4) modes of cycle operation at resonance, provided the time reference is moved to the beginning of each half-cycle.

In steady-state the successive capacitor voltage absolute maxima are

$$\hat{V}_c = V_s \frac{1}{1-e^{-\alpha\pi/\omega}} \quad \text{and} \quad \check{V}_c = -V_s \frac{e^{-\alpha\pi/\omega}}{1-e^{-\alpha\pi/\omega}} \tag{16.16}$$

The peak-to-peak capacitor voltage is therefore

$$V_{c_{p-p}} = \frac{1+e^{-\alpha\pi/\omega}}{1-e^{-\alpha\pi/\omega}} \times V_s = V_s \times \coth(\alpha\pi/2\omega) \approx \frac{2\omega}{\alpha\pi} \times V_s \tag{16.17}$$

The energy transferred to the load *R*, per half sine cycle (per current pulse) is

$$W = \int_0^{\pi/\omega} i^2 R dt = \int_0^{\pi/\omega} \left(\frac{1}{1-e^{\frac{-\alpha\pi}{\omega}}} \times \frac{V_s}{\omega L} \times e^{-\alpha t} \times \sin \omega t \right)^2 R\, dt$$
$$= \tfrac{1}{2} C V_s^2 \coth\left(\alpha\pi/2\omega\right) \tag{16.18}$$

The input impedance of the series circuit is

$$Z_s = Ze^{j\varphi} = R + j\left(\omega L - \frac{1}{\omega C}\right) = R\left(1 + jQ_s\left(\frac{\omega}{\omega_o} - \frac{\omega_o}{\omega}\right)\right) \tag{16.19}$$

where $\varphi = \tan^{-1}\left[Q_s\left(\frac{\omega}{\omega_o} - \frac{\omega_o}{\omega}\right)\right]$

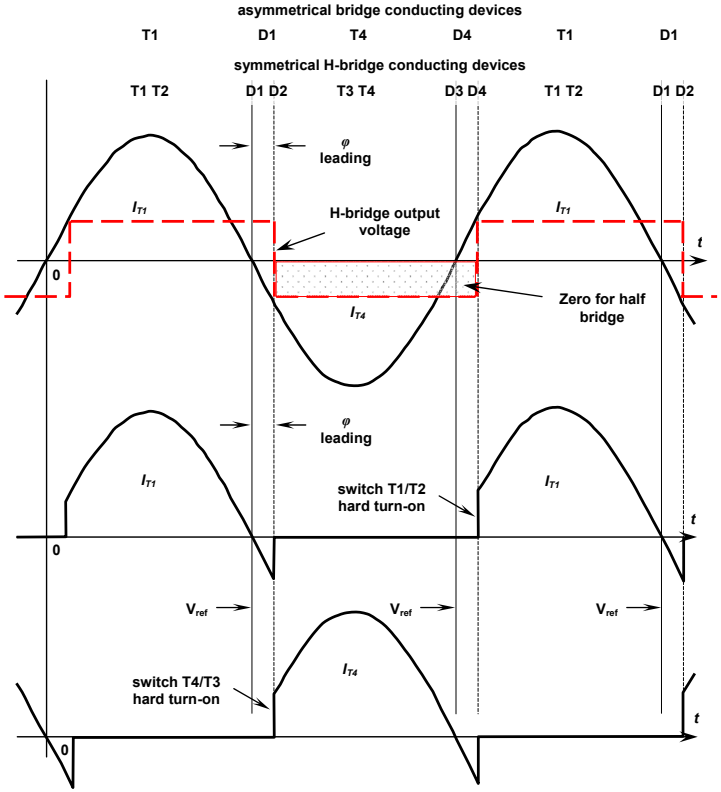


Figure 16.5. *Series L-C-R high Q resonance using the converter circuit in figure 16.2a and b, with f < f_o, a leading power factor φ.*

The frequency ratio terms in the equation for the input phase angle φ show that the resonant circuit is inductive (φ > 0, lagging current) when ω > ω_o and capacitive (φ < 0, leading current) when ω<ω_o.
From the series ac circuit, the voltage across the resistor, v_R, at a given frequency, ω, is given by

$$v_R(\omega) = V_i \frac{R}{R + j\left(\omega L - \frac{1}{\omega C}\right)} \tag{16.20}$$

The magnitude of the resistor voltage is therefore

$$v_R(\omega) = V_i \frac{R}{\sqrt{R^2 + \left(\omega L - \frac{1}{\omega C}\right)^2}} = V_i \frac{1}{\sqrt{1 + \left(\frac{\omega L}{R} - \frac{1}{\omega RC}\right)^2}} = V_i \frac{1}{\sqrt{1 + Q^2\left(\frac{\omega}{\omega_o} - \frac{\omega_o}{\omega}\right)^2}} \tag{16.21}$$

16.3.2 – Series-resonant voltage-source inverter – H-bridge voltage-source inverter – figure 16.2b

When the load *Q* is not high, the capacitor voltage between successive absolute peaks decays significantly, leaving insufficient energy to maintain high efficiency energy transfer to the load *R*. In such cases the resonant circuit is re-enforced with energy from the dc source V_s every half-resonant cycle, by using a full H-bridge as shown in figure 16.2b.
Operation is characterised by turning on switches T1 and T2 to provide energy from the source during one half of the cycle, then having turned T1 and T2 off, T3 and T4 are turned on for the second resonant half cycle. Energy is again drawn from the supply V_s, and when the current reaches zero, T3 and T4 are turned off.
Without bridge freewheel diodes, the switches support high reverse bias voltages, but the switches control the start of each oscillation half cycle. With freewheel diodes the oscillations can continue independent of the switch states. The diodes return energy to the supply, hence reducing the energy transferred to the load. Correct timing of the switches minimises currents in the freewheel diodes, hence minimises the energy needlessly being returned to the supply. Net energy to the load is maximised. As with the single-leg half-bridge, the switches can be used to control the effective load power factor. By advancing turn-off to before the switch current reaches zero, the load can be made to appear inductive, while delaying switch turn-on produces a capacitive load effect. The timing sequencing of the conducting devices, for load power factor control, are shown in figures 16.4 and 16.5.
The series circuit steady-state current at resonance for the symmetrical H-bridge can be approximated by assuming $\omega_o \approx \omega$, such that in equation (16.1) $i_o = 0$:

$$i(\omega t) = \frac{2}{1 - e^{\frac{-\alpha\pi}{\omega}}} \times \frac{V_s}{\omega L} \times e^{-\alpha t} \times \sin \omega t \qquad 0 \le \omega t \le \pi \tag{16.22}$$

which is valid for the ± V_s voltage loops of cycle operation at resonance, provided the time reference is moved to the beginning of each half-cycle.
In steady-state the capacitor voltage absolute maxima are

$$\hat{V}_c = V_s \frac{1 + e^{-\alpha\pi/\omega}}{1 - e^{-\alpha\pi/\omega}} = V_s \times \coth(\alpha\pi/2\omega) = -\check{V}_c \tag{16.23}$$

The peak-to-peak capacitor voltage is therefore

$$V_{c_{p-p}} = 2 \times \frac{1 + e^{-\alpha\pi/\omega}}{1 - e^{-\alpha\pi/\omega}} V_s = 2V_s \coth(\alpha\pi/2\omega) \approx \frac{4\omega}{\alpha\pi} \times V_s \tag{16.24}$$

The energy transferred to the load *R*, per half sine cycle (per current pulse) is

$$W = \int_0^{\pi/\omega} i^2 R\, dt = \int_0^{\pi/\omega} \left(\frac{2}{1 - e^{\frac{-\alpha\pi}{\omega}}} \times \frac{V_s}{\omega L} \times e^{-\alpha t} \times \sin \omega t \right)^2 R\, dt = 2CV_s^2 \coth\left(\alpha\pi/2\omega\right) \tag{16.25}$$

Notice the voltage swing is twice that with the single-leg half-bridge, hence importantly, the power delivered to the load is increased by a factor of four.
From the series ac circuit, the voltage across the resistor, v_R, at a given frequency, ω, is given by

$$v_R(\omega) = V_i \frac{R}{R + j\left(\omega L - \frac{1}{\omega C}\right)} \tag{16.26}$$

The magnitude of the resistor voltage is therefore

$$v_R(\omega) = V_i \frac{R}{\sqrt{R^2 + \left(\omega L - \frac{1}{\omega C}\right)^2}} = V_i \frac{1}{\sqrt{1 + \left(\frac{\omega L}{R} - \frac{1}{\omega RC}\right)^2}} = V_i \frac{1}{\sqrt{1 + Q^2\left(\frac{\omega}{\omega_o} - \frac{\omega_o}{\omega}\right)^2}} \tag{16.27}$$

The frequency ratio terms in the equation for the input phase angle φ show that the resonant circuit is inductive (φ > 0, lagging current) when ω > ω_o and capacitive (φ < 0, leading current) when ω<ω_o.
These resonant circuit resistor expressions are the same as for the half bridge case except the input voltage V_i for the full bridge is twice that of the half bridge case, for the same supply voltage V_s.

If the input voltage V_i is expressed as a Fourier series then the resistor current can be derived in terms of the summation of all the harmonic component according to

$$\sum_{n=1}^{\infty} i_R(n\omega) = \sum_{n=1}^{\infty} v_R(n\omega)/R \tag{16.28}$$

For a square wave input voltage, ±V_s, of frequency $\omega \approx \omega_o$, the input voltage fundament of magnitude $4V_s/\pi$ produces the dominant load current component, since higher frequency components are attenuated by second order *L-C* filtering action. That is, $|V_i| = 4V_s/\pi$.

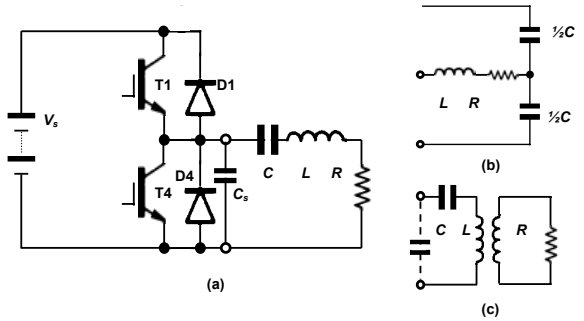


Figure 16.6. *Different resonant load arrangements: (a) switch turn-off snubber capacitor C_s; (b) split capacitor; and (c) series coupled circuit for induction heating.*

16.3.3 – Series circuit variations

Figure 16.6a shows a single-leg half-bridge with a turn-off snubber C_s, where C_s << *C*, hence resonant circuit properties are not affected. The capacitive turn-off snubber is only effective if switch turn-off is advanced such that switch hard turn-off would normally result, that is, the resonant circuit appears capacitive. The snubber acts on both switches since small signal wise (short dc sources), switches T1 and T4 are in parallel.
Figure 16.6b shows a series resonant load used with split resonant capacitance which is in parallel, ac circuit wise. Resonance re-enforcement occurs every half cycle as with the full H-bridge topology, but only two switches are used.
Figure 16.6c shows a transformer-coupled series circuit which equally could be a parallel circuit with *C* in parallel with the coupled circuit, as shown. Under light loads, the transformer magnetising current influences operation.

16.4 Parallel-resonant voltage-source inverter – single inverter leg

The load resistance *R* in this inverter is connected in parallel with the resonant capacitor (or inductor), as shown in figure 16.7a. As a result, if the load resistance is much higher than the reactance of the resonant capacitor, the current through the resonant inductor and the switches is virtually independent of the load. As the load resistance increases, the voltage across the resonant capacitor and the load increases, causing the output power to increase. Parallel in the voltage source case means that the load *R* is in parallel with one of the L-C resonant circuit components: not that *L-C-R* are all in parallel.

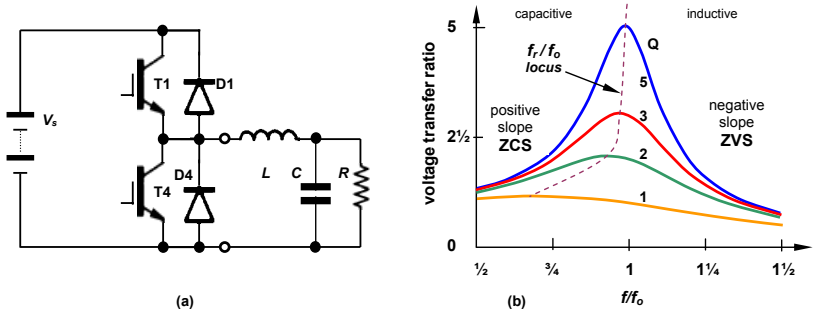


Figure 16.7. *Parallel resonant voltage source converter: (a) circuit and (b) voltage transfer function.*

The resonant circuit input impedance is

$$Z = j\omega L + \frac{R\frac{1}{j\omega C}}{R+\frac{1}{j\omega C}} = \frac{R\left[1-\left(\frac{\omega}{\omega_o}\right)^2 + j\frac{1}{Q}\frac{\omega}{\omega_o}\right]}{1+jQ\frac{\omega}{\omega_o}} \qquad (16.29)$$

where $\omega_o = 1/\sqrt{LC} = f_o/2\pi$, $Z_o = \sqrt{L/C}$, and $Q = R/Z_o > 1$.

At resonance, with a large Q

$$Z(\omega_o) = \frac{Z_o}{\sqrt{Q^2+1}} \approx \frac{Z_o}{R} \qquad (16.30)$$

The circuit resonant frequency, f_r, where the current phase shift is zero, is load R dependant (as opposed to the natural resonant frequency, f_o, which is defined by $\omega_o = 1/\sqrt{LC} = 2\pi f_o$), where

$$\frac{f_r}{f_o} = \sqrt{1-\frac{1}{Q^2}} \text{ for } Q \geq 1 \qquad (16.31)$$

When the bridge voltage excitation frequency is less than the circuit resonant frequency, $f < f_r$, the resonant circuit appears capacitive, and inductive if $f > f_r$.
The voltage gain magnitude transfer function, shown in figure 16.7b, is

$$v_R(\omega) = V_i \frac{1}{\sqrt{\left[1-\left(\frac{\omega}{\omega_o}\right)^2\right]^2 + \frac{1}{Q^2}\left(\frac{\omega}{\omega_o}\right)^2}} \qquad (16.32)$$

which for a large Q tends to

$$v_R(\omega) = V_i \frac{1}{1-\left(\frac{\omega}{\omega_o}\right)^2} \quad \text{as} \quad Q \to \infty \qquad (16.33)$$

The output voltage can be in excess of the input voltage magnitude.

16.5 Series-parallel-resonant voltage-source inverter – single inverter leg

The topology of this inverter is similar to that of the parallel resonant inverter except for an additional capacitor in series with the resonant inductor, or the same as that of the series resonant inverter except for an additional capacitor in parallel with the load, as shown in figure 16.8a. As a result, the inverter exhibits third-order characteristics that are intermediate between those of the series and parallel resonant inverters. In particular, it has a high light-load efficiency.
The resonant circuit input impedance magnitude is

$$Z = Z_o Q \sqrt{\frac{\left[\left(1+A^2\right)\left[1-\left(\frac{\omega}{\omega_o}\right)^2\right]^2 + \frac{1}{Q}\left(\frac{\omega}{\omega_o}-\frac{\omega_o}{\omega}\frac{A}{1+A}\right)^2\right]}{1+\left[Q\frac{\omega}{\omega_o}(1+A)\right]^2}} \qquad (16.34)$$

where $A=C_2/C_1$ and $C=C_1C_2/(C_1+C_2)$ *such that* $\omega_o = 1/\sqrt{LC} = f_o/2\pi$, $Z_o = \sqrt{L/C}$, and $Q = R/Z_o > 1$.

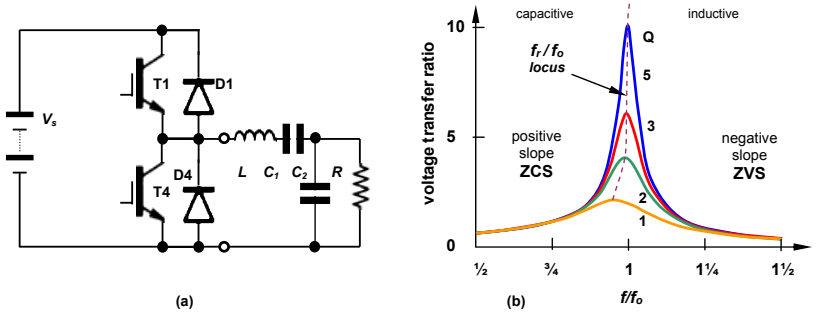


Figure 16.8. *Series-parallel resonant voltage source converter: (a) circuit and (b) voltage transfer function, when C_1=C_2, that is, A=1*

At resonance, with a large Q

$$Z(\omega_o) = \frac{Z_o}{(1+A)\sqrt{1+Q^2(1+A)^2}} \approx \frac{Z_o^2}{R(1+A)^2} \text{ for } Q(1+A)^2 \gg 1 \qquad (16.35)$$

As with the parallel resonant inverter case, series-parallel resonance is load resistance R dependant:

$$\frac{f_r}{f_o} = \sqrt{\frac{Q^2(1+A)^2 - 1 + \sqrt{\left[Q^2(1+A)^2-1\right]^2 + 4Q^2A(1+A)}}{2Q^2(1+A)^2}} \qquad (16.36)$$

When the bridge voltage excitation frequency is less than the circuit resonant frequency, $f < f_r$, the resonant circuit appears capacitive, and inductive if $f > f_r$.

The voltage gain magnitude transfer function, shown in figure 16.8b, is

$$v_R(\omega) = V_i \frac{1}{\sqrt{(1+A)^2\left[1-\left(\frac{\omega}{\omega_o}\right)^2\right]^2 + \frac{1}{Q^2}\left(\frac{\omega}{\omega_o}-\frac{\omega_o}{\omega}\frac{A}{1+A}\right)^2}} \qquad (16.37)$$

The output voltage can be in excess of the input voltage magnitude.

The inverter is not safe under short-circuit and the open-circuit conditions. At R = 0, the capacitor C_2 is shorted-circuited and the resonant circuit consists of L and C_1. If the switching frequency f is equal to the resonant frequency of this circuit $f_{r1} = 1/2\pi\sqrt{LC_1}$, the magnitude of the current through the switches and the L-C_1 resonant circuit is limited only by low switch resistance and the reactive components. This current may become excessive and destroy the circuit. If f is remote from f_{r1}, the current amplitude is limited by the reactance of the resonant circuit. Because $f_{r1} < f_o$, the inverter is safe for switching frequencies above f_o. At $R = \infty$, the resonant circuit consists of L and the series combination of C_1 and C_2. Consequently, its resonant frequency is equal to f_o and the inverter is not safe at or close to this frequency, as with the series resonant inverter under light loads.

Summary of voltage source resonant inverters
The maximum voltage across the switches in voltage-source inverters (both half-bridge and full-bridge) is equal to the dc input voltage V_s.

- Inverter operation above the resonant frequency f_r is preferred. Such operation results in an inductive load seen by the bridge switches. The switches turn on at zero voltage, thereby reducing the turn-on switching loss, Miller's effect is absent, the switch input capacitance is low, the switch drive-power requirement is low, and turn-on speed is fast. However, switch turn-off is hard switched. The anti-parallel diodes turn off with a low *di*/*dt* and without diode recovery voltage snap.
- During operation below resonance, the anti-parallel fast recovery diodes turn off with a high *di*/*dt* and generate reverse recovery current spikes. These spikes are present in the switch current at both turn-on and turn-off and stress the switches. For operation below resonance, the switches are turned on at the supply voltage V_s and the switch output capacitance is discharged into a low switch on-resistance, producing a high turn-on switching loss.

The resonant frequency $f_r = f_o$ is constant in the series resonant inverter but f_r depends on the load R in the parallel and series-parallel resonant inverters.

The series resonant inverter can operate safely with an open circuit output, although the output voltage can not be regulated. It is, however, exposed to excessive currents, $>V_s/Z_o$,, which builds up with successive operational cycles, if the output is short-circuited at the operating frequency f close to the resonant frequency $f = f_o$. Any output short circuit protection can exploit the time it takes for the current build up. The parallel resonant inverter output is protected by the impedance of the inductor from a short circuit output at any switching frequency. Large output currents occur when the output is open-circuited at a switching frequency close to the corner frequency f_o. Series-parallel resonant inverter operation is not safe with an open-circuited output at frequencies close to the corner frequency f_o and with a short-circuited output at frequencies close to the resonant frequency f_r.
The output voltage of resonant inverters is regulated by changing the switching frequency. However, the required frequency changes in the series resonant inverter are large for no-load or light-load conditions. The series resonant inverter output voltage can be controlled by varying the duty cycle of the square-wave excitation, whilst operating at the resonance frequency. The parallel resonant inverter exhibits good light-load regulation, by operating above resonance. The output voltage at resonance is a function of load, thus can rise to high voltages at no load, if the operating frequency is no increased. It has, however, a low light-load efficiency due to a relatively constant current through the resonant circuit. The series-parallel resonant inverter combines the advantages, and eliminates the weaknesses, of the series and parallel resonant inverter topologies at the expense of an additional resonant capacitor. Alleviated are the poor light load regulation of the series resonant converter and the circulating current independent of load of the parallel resonant inverter.
In the series resonant converter, the series capacitor tends to act is a dc blocking capacitor, facilitating H-bridge operation, prevent core saturation when the load is magnetically coupled. Also the current in the semiconductors decreases as the load decreases, which helps maintain the efficiency at light loads.
The input voltage of the resonant circuit in the switched mode full-bridge inverters is a square wave with the with voltage levels of $\pm V_s$. The peak-to-peak voltage across the resonant circuit in the full-bridge inverter is twice that in the half-bridge inverter. Therefore, the output voltage of the full-bridge inverter is also twice as high and the output power is four times higher than that from the half-bridge inverter at the same operating conditions (load, input voltage, and switching frequency).

16.6 Parallel-resonant current-source inverters

Parallel resonant circuits use a current source inverter (class D, parallel) as considered in 16.2.1 and shown in figure 16.2 parts c and d. If the load Q is high, then resonance need only be re-enforced every second half-cycle, thereby simplifying converter and control requirements. A common feature of parallel resonant circuits fed from a current source, is that commutation of the switches involves overlap where the output of the current source can be briefly shorted.

16.6.1 – Parallel-resonant current-source inverter – single inverter leg – figure 16.2c

Figure 16.2c shows a single-leg half-bridge converter for high Q parallel load circuits. Energy is provided from the constant current source every second half cycle by turning on switch T1. When T1 is turned on (and T3 is subsequently turned off) the voltage across the L-C-R circuit resonates from zero to a maximum and back to zero volts. The energy in the inductor reaches a maximum at each zero voltage instant. T3 is turned on (at zero volts) to divert current from T1, which is then turned off with zero terminal voltage. The energy in the load inductor resonates within the load circuit, with the load in an open circuit state, since T1 is off. The sequence continues when the load voltage resonates back to zero as shown in figure 16.1b.
The parallel circuit steady-state voltage at resonance for the single-leg half-bridge can be approximated by assuming $\omega_o \approx \omega$, such that in equation (16.8) $v_o = 0$:

$$v(\omega t) = \frac{1}{1-e^{\frac{-\alpha\pi}{\omega}}} \times \frac{I_s}{\omega C} \times e^{-\alpha t} \times \sin \omega t \qquad 0 \le \omega t \le \pi \tag{16.38}$$

which is valid for both the $+I_s$ loop and open circuit load modes of cycle operation, provided the time reference is moved to the beginning of each half-cycle.
In steady-state the successive inductor current absolute maxima are

$$\hat{I}_L = I_s \frac{1}{1-e^{-\alpha\pi/\omega}} \quad \text{and} \quad \check{I}_L = I_s \frac{-e^{-\alpha\pi/\omega}}{1-e^{-\alpha\pi/\omega}} \tag{16.39}$$

The energy transferred to the load R, per half sine cycle (per voltage pulse) is

$$W = \int_0^{\pi/\omega} v^2/R \; dt = \int_0^{\pi/\omega} \left(\frac{1}{1-e^{\frac{-\alpha\pi}{\omega}}} \times \frac{I_s}{\omega C} \times e^{-\alpha t} \times \sin \omega t \right)^2 / R \, dt$$
$$= \tfrac{1}{2} L I_s^2 \coth\left(\frac{\alpha\pi}{2\omega}\right) \tag{16.40}$$

To drive a parallel circuit from a voltage source inverter leg the resonant circuit inductance is series connected to the parallel R-C circuit. The input impedance of the series plus parallel circuit is

$$Z_p = Ze^{j\varphi} = R \left(\frac{1-\left(\frac{\omega}{\omega_o}\right)^2 + j\frac{1}{Q_p}\frac{\omega}{\omega_o}}{1+jQ_p\frac{\omega}{\omega_o}} \right) \tag{16.41}$$

where $\varphi = \tan^{-1}\left[Q_p \frac{\omega}{\omega_o}\left(\left(\frac{\omega}{\omega_o}\right)^2 - \frac{1}{Q_p^2} - 1 \right) \right]$

For a voltage source inverter leg, from the series plus parallel ac circuit, the voltage across the resistor, v_R, at a given frequency, ω, is given by

$$v_R(\omega) = Ve^{j\varphi} = V_i \frac{\frac{R}{j\omega C}}{R+\frac{1}{j\omega C}} \Big/ \frac{j\omega L + \frac{R}{j\omega C}}{R+\frac{1}{j\omega C}} = V_i \frac{1}{1-\left(\frac{\omega}{\omega_o}\right)^2 + j\frac{1}{Q_p}\frac{\omega}{\omega_o}} \tag{16.42}$$

The magnitude of the resistor voltage is therefore

$$v_R(\omega) = V_i \frac{1}{\sqrt{\left[1-\left(\frac{\omega}{\omega_o}\right)^2\right]^2 + \frac{1}{Q_p^2}\left(\frac{\omega}{\omega_o}\right)^2}} \tag{16.43}$$

where $\varphi = -\tan^{-1} \dfrac{\frac{1}{Q_p}\frac{\omega}{\omega_o}}{1-\left(\frac{\omega}{\omega_o}\right)^2}$

The maximum resistor voltage is $Q_p / \sqrt{1-1/4Q_p^2}$ at $f = f_o\sqrt{1-1/2Q_p^2}$. The effective input voltage V_i is $2V_s/\pi$.

16.6.2 – Parallel-resonant current-source inverter – H-bridge current-source inverter–figure 16.2d

If the load Q is low, or maximum energy transfer to the load is required, the full bridge converter shown in figure 16.1d is used.
Operation involves T1 and T2 directing the constant source current to the load and when the load voltage falls to zero, T3 and T4 are turned on (and T1 and T2 are then turned off). Overlapping the switching sequence ensures a path always exists for the source current. At the next half sinusoidal cycle voltage zero, T1 and T2 are turned on and then T3 and T4 are turned off.
The parallel circuit steady-state voltage for the symmetrical H-bridge can be approximated by assuming $\omega_o \approx \omega$, such that in equation (16.8) $v_o = 0$:

$$v(\omega t) = \frac{2}{1-e^{\frac{-\alpha\pi}{\omega}}} \times \frac{I_s}{\omega C} \times e^{-\alpha t} \times \sin \omega t \qquad 0 \le \omega t \le \pi \tag{16.44}$$

which is valid for both the $+I_s$ loops of cycle operation, provided the time reference is moved to the beginning of each half-cycle.

In steady-state the successive inductor current absolute maxima are

$$\hat{I}_L = I_s \frac{1+e^{-\alpha\pi/\omega}}{1-e^{-\alpha\pi/\omega}} = I_s \times \coth(\alpha\pi/2\omega) = -\check{I}_L \tag{16.45}$$

The energy transferred to the load R, per half sine cycle (per voltage pulse) is

$$W = \int_0^{\pi/\omega} v^2/R \; dt = \int_0^{\pi/\omega} \left(\frac{2}{1-e^{\frac{-\alpha\pi}{\omega}}} \times \frac{I_s}{\omega C} \times e^{-\alpha t} \times \sin \omega t \right)^2 / R \, dt$$
$$= 2LI_s^2 \coth\left(\frac{\alpha\pi}{2\omega}\right)$$

As with a series resonant circuit, the full bridge delivers four times more power to the load than the single-leg half-bridge circuit. Similarly, the load power and power factor can be controlled by operating above or below the resonant frequency, by delaying or advancing the appropriate switching instances. In the case of a voltage source, the expressions for the voltage across the load resistor are the same as equations (16.41) to (16.43), except the input voltage V_i is doubled, from $2V_s/\pi$ to $4V_s/\pi$.

Example 16.1: ***Single-leg half-bridge with a series L-C-R load***

A single-leg half-bridge inverter as shown in the figure 16.2a, with the dc rail *L-C* decoupling shown in figure 16.9, supplies a 1 ohm resistance load with series inductance 100 μH from a 340 V dc source. If the bridge is to operating at a resonant frequency of 10kHz, determine:

i. the necessary series *C* for resonance at 10kHz and the resultant *Q*
ii. the peak load current, its steady-state time domain solution, and peak capacitor voltages
iii. the bridge rms voltage and fundamental voltage across the series *L-C-R* load
iv. the power delivered to the load and the frequency when half power is delivered to the load. What is the switching advance/delay time?
v. the peak blocking voltage of each semiconductor type (and for the case when the freewheel diodes are not employed)
vi. the average, rms, and peak current in the switches and diodes
vii. the resonant capacitor specification
viii. the dc supply current and the dc link capacitor rms current
ix. summarise conditions if the load is supplied from an H-bridge and also calculate the load power supplied at the third harmonic frequency, $3\omega_o$.

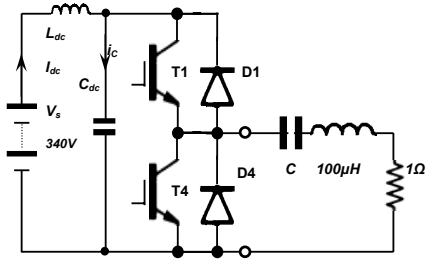


Figure 16.9. *Single-leg half-bridge series-resonance circuit.*

Solution

i. From $\omega_o = 2\pi f_o = 1/\sqrt{LC}$ the necessary capacitance for resonance at 10kHz with 100μH is

$$C = \frac{1}{(2\times\pi\times 10\text{kHz})^2 \times 100\mu\text{H}} = 2.5\mu\text{F}$$

The circuit quality factor *Q* is given by

$$Q = \frac{Z_o}{R} = \sqrt{\frac{L}{C}}/R = \sqrt{\frac{100\mu\text{H}}{2.5\mu\text{F}}}/1\Omega = 6.3$$

Therefore

$\alpha = 5\times10^3$ Ω/H $\quad \omega = 62.6$ krad/s (9.968 kHz) $\quad Z_o = 6.3$ Ω
$\xi = 0.079$ $\quad BW_s = 9.97$ krad/s (1.587kHz)

ii. The steady-state current is given by equation (16.15)

$$i(\omega t) = \frac{1}{1-e^{\frac{-\alpha\pi}{\omega}}} \times \frac{V_s}{\omega L} \times e^{-\alpha t} \times \sin\omega t$$
$$= 245.5 \times e^{-5000t} \times \sin(2\pi 10\text{kHz} \times t)$$

Since the Q is high (6.3), a reasonably accurate estimate of the peak current results if the current expression is evaluated at sin(½π), that is t = 25μs, which yields $\hat{i}$ = 216.7A. The rms load current is 216.7A/√2 = 153.2A rms.

From equation (16.16) the maximum capacitor voltage extremes are

$$\hat{V}_c = V_s \frac{1}{1-e^{-\alpha\pi/\omega}} \quad \text{and} \quad \check{V}_c = V_s \frac{-e^{-\alpha\pi/\omega}}{1-e^{-\alpha\pi/\omega}}$$
$$= \frac{340\text{V}}{1-e^{-0.25}} \qquad = -\frac{340\text{V}e^{-0.25}}{1-e^{-0.25}}$$
$$= 1537\text{V} \qquad = -1197\text{V}$$

iii. The bridge output voltage is a square wave of magnitude 340V and 0V, with a 50% duty cycle. The rms output voltage is therefore 340/√2 = 240.4V.

Since the load is at resonance, the current is in phase with the fundamental of the bridge output voltage.

The fundament voltage magnitude is given by

$$b_1 = \frac{1}{\pi}\int_0^{\pi} V_s \sin 1\omega t = \frac{2V_s}{\pi} = 216.5\text{V peak}$$
$$\equiv \frac{\sqrt{2}V_s}{\pi} = 153\text{V rms}$$

The rms load current results because of the fundamental voltage, that is, the peak sine current is 216.5V/1Ω = 216.5A peak or 153V/1Ω = 153A rms. This agrees with the current values calculated in part b.

iv. The power delivered to the load is given by

$$P = i_{rms}^2 R = i_{b1}^2 R$$
$$= 153\text{A}^2 \times 1\Omega = 23.41\text{kW}$$

Substitution into equation (16.18) gives 23.15kW at a pulse rate of 2×10kHz. Alternately

$$P = V_s \times \bar{I} = V_s \times 0.45 \times I_{rms}$$
$$= 340\text{V} \times 0.45 \times 153\text{A} = 23.42\text{kW}$$

The half-power frequencies are when the reactive voltage magnitude equals the resistive voltage magnitude.

$$f_\ell^u = f_o \pm \frac{R}{4\pi L}$$
$$= 10\text{kHz} \pm 796\text{Hz}$$

Thus at 9204 Hz and 10796 Hz the voltage across the resistive part of the load is reduced to 1/√2 of the inverter output voltage, since the voltage vectors are perpendicular. The power (proportional to voltage squared) is therefore halved (11.71kW) at the half-power frequencies.

Operating above resonance, $f > f_o$ produces an inductive load and this is achieved by turning T1 and T4 off prematurely. Zero current turn-on occurs, but hard switching results at turn-off. To operate at the 10796Hz (92.6μs) upper half-power frequency the period has to be reduced from 100μs (10kHz) to 92.6μs. The period of each half cycle has to be reduced by ½×(100μs - 92.6μs) = 3.7μs

Operating below resonance, $f < f_o$ produces a capacitive load and this is achieved by turning T1 and T4 on late. Zero current turn-off occurs, but hard switching results at turn-on. By delaying turn-on of each switch by ½×(109μs - 100μs), 4.5μs, the effective oscillation frequency will be decreased to the lower half-power frequency, 9204Hz.

v. The bridge diodes, which do not conduct at resonance, clamp switch and diode maximum supporting voltages to the rail voltage, 340V dc.

Note that if clamping diodes were not employed the device maximum off-state voltages would occur during switch change over, when one switch has just been turned off, and just before the on-going switch is turned on. The load current is zero, so the load terminal voltage is the capacitor voltage.

Switch T1 would need to support
a forward voltage of $V_s - \check{v}$ = 340V + 1197V = 1537V = $\hat{v}$ and
a reverse voltage of $\hat{v} - V_s$ = 1537V - 340V = 1197V = $-\check{v}$, while
Switch T4 supports
a forward voltage of $\hat{v}$ = 1537V and
a reverse voltage of $-\check{v}$ = 1197V.

Thyristor family devices must be used, or devices with a series connected diode, which will increase the converter on-state losses.

vi. At resonance the two freewheel diodes do not conduct.
The rms load current is 153.2 A at 10 kHz, where switch T1 conducts half the cycle and T4 conducts the other half which is the opposite polarity of the cycle. Each switch therefore has an rms current rating of 153.2/√2 = 108.3A rms.
Since both switches conduct the same current shape, each has an average current rating of a half-wave rectified sine of magnitude 216.5A, that is

$$\bar{I}_{T1} = \frac{1}{2\pi}\int_0^{\pi} 216.5\sin\omega t\,dt = \frac{1}{\pi}\times 216.5\text{A}$$
$$= 0.45\times 216.5/\sqrt{2} = 68.9\text{A}$$

By Kirchhoff's current law, this current value for T1 is also equal to the average dc input current from the supply V_s.

vii. The 2.5μF capacitor has a bipolar voltage and current requirement of ±1537V and ±216.7 A. The rms ratings are therefore ≈1087V rms and 153A rms. A metallised polypropylene capacitor capable of 10kHz ac operation, with a maximum *dv/dt* rating of approximately ½×(1537+1197)×ω, that is 85.6V/μs, is required.

viii. The dc supply current is the average value of the half-wave rectified sinusoidal load current, which is the average current in T1. That is

$$I_{dc} = 0.45\times 153.1\text{A rms}$$
$$= 68.9\text{A dc}$$

The rms current in the dc link capacitor C_{dc} is related to the dc input current and switch T1 rms current (as found in part vi.), by

$$I_c = \sqrt{I_{rms}^2 - I_{dc}^2}$$
$$= \sqrt{108.3^2 - 68.9^2} = 83.6\text{A rms}$$

ix. For the full H-bridge, the load dependant parameters *C*, ω_o, ω, α, *Q*, *BW*, ξ, and half power points remain unchanged, being independent of bridge type and switching frequency.

From equation (16.22) the steady-state current is double that for the asymmetrical bridge,

$$i(\omega t) = \frac{2}{1-e^{\frac{-\alpha\pi}{\omega}}}\times\frac{V_s}{\omega L}\times e^{-\alpha t}\times\sin\omega t$$
$$= 491\times e^{-5000t}\times\sin(2\pi 10\text{kHz}\times t)$$

The peak current is $\hat{i}$ = 433.4A.
The rms load current is 433.4A/√2 = 306.4A rms

From equation (16.23) both the maximum capacitor voltages are

$$\hat{V}_c = V_s\frac{1+e^{-\alpha\pi/\omega}}{1-e^{-\alpha\pi/\omega}} = -\check{V}_c$$
$$= 340\text{V}\frac{1+e^{-0.25}}{1-e^{-0.25}} = 2734\text{V}$$

The power delivered to the load is four times the single-leg half-bridge case and is

$$P = i_{rms}^2 R = 306.4\text{A}^2\times 1\Omega = 93.88\text{kW}$$

The average switch current is 194.8A, but the average supply current is four times the single-leg half-bridge case and is 275.5.6A.
For a square wave, the third harmonic voltage is a third the magnitude of the fundamental. From equation (16.27), for operation at the lower half power frequency 9204Hz, (which would result in the largest harmonic component magnitude after *L*-*C* filtering attenuation) f_3 = 27.6kHz.

$$v_R(\omega_{\frac{1}{3}}) = \tfrac{1}{3}\times\frac{4V_s}{\pi}\times\frac{1}{\sqrt{1+Q^2\left(\frac{3\omega_{\frac{1}{3}}}{\omega_o}-\frac{\omega_o}{3\omega_{\frac{1}{3}}}\right)^2}}$$
$$= \tfrac{1}{3}\times\frac{4\times 340\text{V}}{\pi}\times\frac{1}{\sqrt{1+6.3^2\left(\frac{3\times 2\pi 9.204\text{kHz}}{2\pi 10\text{kHz}}-\frac{2\pi 10\text{kHz}}{3\times 2\pi 9.204\text{kHz}}\right)^2}}$$

$$= \tfrac{1}{3}\times\frac{4\times 340\text{V}}{\pi}\times\frac{1}{\sqrt{1+6.3^2\left(\frac{3\times 9.204}{10}-\frac{10}{3\times 9.204}\right)^2}}$$
$$= 144.3\text{V}\times 0.066 = 9.53\text{V}$$

The magnitude of the third harmonic current is therefore 9.5V/1Ω = 9.5A or 6.7A rms. The load power at this frequency is $6.7\text{V}^2/1\Omega$ = 45.1W. This is clearly insignificant compared to the fundament power of 93.88kW being delivered to the 1 Ω load.

♣

16.7 Single-switch, current source, series resonant inverter

The single switch inverter in figure 16.10 is applicable to high *Q* load circuits such that the output is essentially sinusoidal, with zero average current. Based on the operating mechanisms, a sinusoidal current implies the switch has a 50% duty cycle. The switch turns on and off at zero volts so switch losses are low, thus the operating frequency can be high. The input inductor L_{large} in conjunction with the input voltage source, during steady state operation, act as a current source input, I_s, for the resonant circuit, such that $V_s I_s$ is equal to the power delivered to the load *R*.

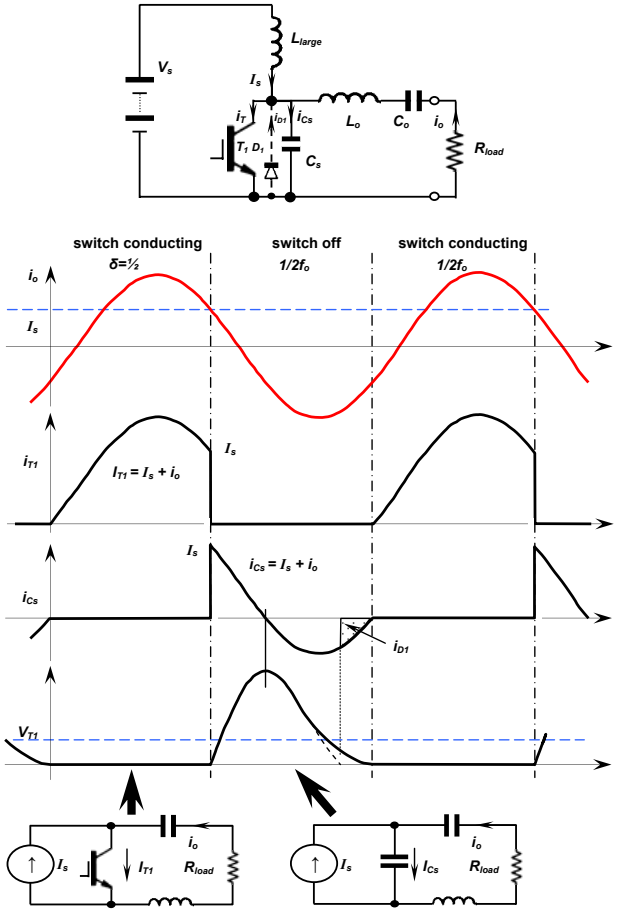


Figure 16.10. *Single-switch, current-source series resonant converter circuit and waveforms.*

When the switch T1 is turned on, with zero terminal voltage, it conducts both the constant current I_s and the current i_o resonating in the output circuit, as shown in the circuit waveforms in figure 16.10. The resonating load current builds up. The switch T1, which is in parallel with C_s, is turned off. Current from the switch is diverted to C_s, which charges from an initial voltage of zero. C_s thus forms a turn-off snubber in parallel with T1. The charge on C_s eventually resonates back to zero at which instant the switch is turned on, again, with zero turn-on loss.

The resonant frequency is $\omega_o = 1/\sqrt{L_o C_o}$ and because of the high Q, a small change in the switching frequency significantly decreases the output current, hence output voltage.

As with any current source inverter, the peak switch voltage is in excess of V_s. Since the current is sinusoidal, the average load voltage and inductor voltage are zero. Therefore the average voltage across C_o and C_s is the supply voltage V_s. The peak switch voltage can be estimated to be in excess of V_s /0.45 which is based on a half-wave rectified average sinusoidal voltage.

If the load conditions change and the switch duty cycle is varied from δ = ½, circuit voltages increase and capacitor C_s voltage discharges before the circuit current reaches zero. The capacitor and switch are bypassed with current flowing through the diode D1. This diode prevents the switch from experiencing a negative voltage and the capacitor C_s from charging negatively.

Although such resonant converters offer features such as low switching losses and low radiated EMI, optimal control and performance are difficult to maintain and extremely high circuit voltages occur at low duty cycles.

Reading list

Hart, D.W., *Introduction to Power Electronics,*
Prentice-Hall, Inc, 1994.

Mohan, N., *Power Electronics,* 3rd Edition,
Wiley International, 2003.

Problems

Blank

17

DC to DC Converters
- Switched-Mode

A switched-mode power supply (smps) or switching regulator, efficiently converts a dc voltage level to another dc voltage level, via an intermediate magnetic (inductor) storage/transfer stage, such that a continuous, possibly constant, load current flows, usually at power levels below a few kilowatts.

Shunt and series linear regulator power supplies dissipate much of their energy across the regulating transistor, which operates in the linear mode. An smps achieves regulation by varying the on to off time duty cycle of the switching element. This switching minimises losses, irrespective of load conditions.

Figure 17.1 illustrates the basic principle of the ac-fed smps in which the ac mains input is rectified, capacitively smoothed, and supplied to a high-frequency transistor chopper. The chopped dc voltage is transformed, rectified, and smoothed to give the required dc output voltage. A high-frequency transformer is used if an isolated output is required. The output voltage is sensed by a control circuit that adjusts the duty cycle of the switching transistor in order to maintain a constant output voltage with respect to load and input voltage variation. Alternatively, the chopper can be configured and controlled such that the input current tracks a scaled version of the input ac supply voltage, therein producing unity (or controllable) power factor *I*-*V* input conditions.

The switching frequency can be made much higher than the 50/60Hz line frequency; then the filtering and transformer elements used can be made small, lightweight, low in cost, and efficient.

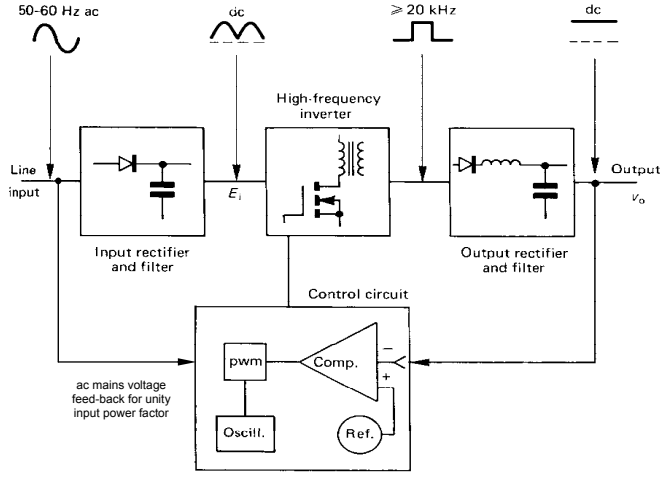


Figure 17.1. *Functional block diagram of a switched-mode power supply.*

Depending on the requirements of the application, the dc-to-dc converter can be one of four basic converter types, namely

- forward
- flyback
- balanced
- resonant.

17.1 The forward converter

The basic *forward converter,* sometimes called a step-down or *buck converter,* is shown in figure 17.2a. The input voltage E_i is chopped by transistor T. When T is on, because the input voltage E_i is greater than the load voltage v_o, energy is transferred from the dc supply E_i to *L*, *C*, and the load *R*. When T is turned off, stored energy in *L* is transferred via diode D to *C* and the load *R*.

If all the stored energy in *L* is transferred to *C* and the load before T is turned back on, operation is termed *discontinuous* inductor current, since the inductor current has reached zero. If T is turned on before the current in *L* reaches zero, that is, if continuous current flows in *L*, inductor operation is termed *continuous.*

Parts b and c respectively of figure 17.2 illustrate forward converter circuit current and voltage waveforms for continuous (figure 17.1b) and discontinuous (figure 17.1c) current conduction of inductor *L*.

For analysis it is assumed that components are lossless and the output voltage v_o is maintained constant because of the large magnitude of the capacitor *C* across the output. The input voltage E_i is also assumed constant, such that $E_i \geq v_o$.

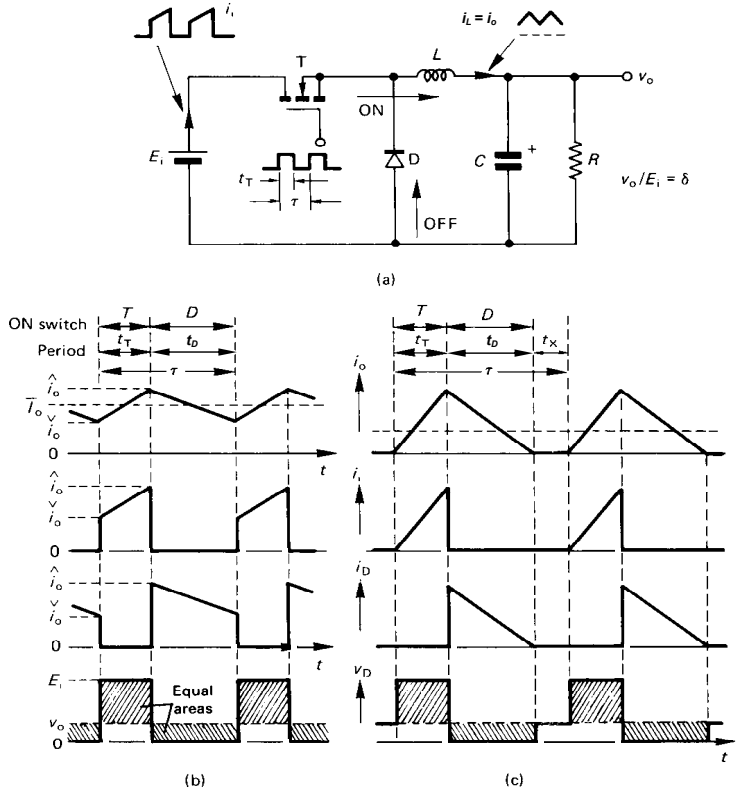


Figure 17.2. *Non-isolated forward converter (buck converter) where $v_0 \leq E_i$: (a) circuit diagram; (b) waveforms for continuous output (inductor) current; and (c) waveforms for discontinuous output (inductor) current.*

17.1.1 Continuous inductor current

The inductor current is analysed first when the switch is on, then when the switch is off. When transistor T is turned on for period t_T, the difference between the supply voltage E_i and the output voltage v_o is impressed across L. From $V = L\,di/dt = L\,\Delta i/\Delta t$, the linear current change through the inductor will be

$$\Delta i_L = \hat{i}_L - \check{i}_L = \frac{E_i - v_o}{L} \times t_T \tag{17.1}$$

When T is switched off for the remainder of the switching period, $t_D = \tau - t_T$, the freewheel diode D conducts and $-v_o$ is impressed across L. Thus, using $V = L\,\Delta i/\Delta t$, rearranged, assuming continuous conduction

$$\Delta i_L = \frac{v_o}{L} \times (\tau - t_T) \tag{17.2}$$

Equating equations (17.1) and (17.2) gives

$$(E_i - v_o)\, t_T = v_o\, (\tau - t_T) \tag{17.3}$$

This expression shows that the inductor average voltage is zero, and after rearranging ($P_{out} = P_{in}$):

$$\frac{v_o}{E_i} = \frac{\bar{I}_i}{\bar{I}_o} = \frac{t_T}{\tau} = \delta = t_T f \qquad 0 \le \delta \le 1 \tag{17.4}$$

This equation also shows that for a given input voltage, the output voltage is determined by the transistor conduction duty cycle δ and the output is always less than the input voltage. This confirms and validates the original analysis assumption that $E_i \ge v_o$. The voltage transfer function is independent of the load R, circuit inductance L and capacitance C.

The inductor rms ripple current (and capacitor ripple current in this case) from equations (17.1) and (17.2) , for continuous inductor current, is given by

$$i_{Lr} = \frac{\Delta i_L}{2\sqrt{3}} = \frac{1}{2\sqrt{3}} \frac{v_o}{L}(1-\delta)\tau = \frac{1}{2\sqrt{3}} \frac{E_i}{L}(1-\delta)\delta\tau \tag{17.5}$$

while the inductor total rms current is

$$i_{L\,rms} = \sqrt{\bar{I}_L^2 + i_{Lr}^2} = \sqrt{\bar{I}_L^2 + \left(\tfrac{1}{2}\Delta i_L / \sqrt{3}\right)^2} = \sqrt{\frac{1}{3}\left(\hat{i}_L^2 + \hat{i}_L \times \check{i}_L + \check{i}_L^2\right)} \tag{17.6}$$

The switch and diode average and rms currents are given by

$$\begin{aligned} \bar{I}_T &= \bar{I}_i = \delta \bar{I}_o & I_{Trms} &= \sqrt{\delta}\; i_{L\,rms} \\ \bar{I}_D &= \bar{I}_o - \bar{I}_i = (1-\delta)\bar{I}_o & I_{Drms} &= \sqrt{1-\delta}\; i_{L\,rms} \end{aligned} \tag{17.7}$$

If the average inductor current, hence output current, is $\bar{I}_L$, then the maximum and minimum inductor current levels are given by

$$\begin{aligned} \hat{i}_L &= \bar{I}_L + \tfrac{1}{2}\Delta i_L = \bar{I}_o + \tfrac{1}{2}\frac{v_o}{L}(1-\delta)\tau \\ &= v_o\left[\frac{1}{R} + \frac{1-\delta}{2fL}\right] \end{aligned} \tag{17.8}$$

and

$$\begin{aligned} \check{i}_L &= \bar{I}_L - \tfrac{1}{2}\Delta i_L = \bar{I}_o - \tfrac{1}{2}\frac{v_o}{L}(1-\delta)\tau \\ &= v_o\left[\frac{1}{R} - \frac{1-\delta}{2fL}\right] \end{aligned} \tag{17.9}$$

respectively, where Δi_L is given by equation (17.1) or (17.2). The average output current is $\bar{I}_L = \tfrac{1}{2}(\hat{i}_L + \check{i}_L) = \bar{I}_o = v_o / R$. The output power is therefore v_o^2 / R , which equals the input power, namely $E_i \bar{I}_i = E_i \bar{I}_T$. Circuit waveforms for continuous inductor current conduction are shown in figure 17.2b.

Switch utilisation ratio

The switch utilisation ratio, SUR, is a measure of how fully a switching device's power handling capabilities are utilised in any switching application. The ratio is defined as

$$SUR = P_{out} \Big/ p \hat{V}_T \hat{I}_T \tag{17.10}$$

where p is the number of power switches in the circuit; $p=1$ for the forward converter. The switch maximum instantaneous voltage and current are $\hat{V}_T$ and $\hat{I}_T$ respectively. As shown in figure 17.2b, the maximum switch voltage supported in the off-state is E_i, while the maximum current is the maximum inductor current $\hat{i}_L$ which is given by equation (17.8). If the inductance L is large such that the ripple current is small, the peak inductor current is approximated by the average inductor current $\hat{I}_T \approx \bar{I}_L = \bar{I}_o$, that is

$$SUR = \frac{v_o \bar{I}_o}{1 \times E_i \times \bar{I}_o} = \frac{v_o}{E_i} = \delta \tag{17.11}$$

which assumes continuous inductor current. This result shows that the higher the duty cycle, that is the closer the output voltage v_o is to the input voltage E_i, the better the switch I-V ratings are utilised.

17.1.2 Discontinuous inductor current

The onset of discontinuous inductor current operation occurs when the minimum inductor current $\check{i}_L$, reaches zero. That is, with $\check{i}_L = 0$ in equation (17.9), the last equality

$$\frac{1}{R} - \frac{(1-\delta)}{2fL} = 0 \tag{17.12}$$

relates circuit component values (R and L) and operating conditions (f and δ) at the verge of discontinuous inductor current. Also, with $\check{i}_L = 0$ in equation (17.9)

$$\bar{I}_L = \bar{I}_o = \tfrac{1}{2}\Delta i_L \tag{17.13}$$

which, after substituting equation (17.1) or equation (17.2), yields

$$\bar{I}_L = \bar{I}_o = \frac{(E_i - v_o)}{2L}\tau\delta \quad \text{or} \quad \frac{E_i}{2L}\tau\delta(1-\delta) \tag{17.14}$$

If the transistor on-time t_T is reduced (or the load current is reduced), the discontinuous condition dead time t_x is introduced as indicated in figure 17.2c. From equations (17.1) and (17.2), with $\check{i}_L = 0$, the output voltage transfer function is now derived as follows

$$\hat{i}_L = \frac{(E_i - v_o)}{L} t_T = \frac{v_o}{L}(\tau - t_T - t_x) \tag{17.15}$$

that is

$$\frac{v_o}{E_i} = \frac{\delta}{1 - \frac{t_x}{\tau}} \qquad 0 \le \delta < 1 \quad \text{and} \quad t_x \ge 0 \tag{17.16}$$

This voltage transfer function form may not be particularly useful since the dead time t_x is not expressed in term of circuit parameters. Accordingly, from equation (17.15)

$$\hat{i}_L = \frac{(E_i - v_o)}{L} t_T \tag{17.17}$$

and from the input current waveform in figure 17.2c:

$$\bar{I}_i = \tfrac{1}{2}\hat{i}_L \times \frac{t_T}{\tau} \tag{17.18}$$

Eliminating $\hat{i}_L$ yields

$$\frac{2\bar{I}_i}{\delta} = (1 - \frac{v_o}{E_i})\frac{\tau\delta E_i}{L} \tag{17.19}$$

that is

$$\frac{v_o}{E_i} = 1 - \frac{2L\bar{I}_i}{\delta^2 \tau E_i} \tag{17.20}$$

Assuming power-in equals power-out, that is, $E_i \bar{I}_i = v_o \bar{I}_o = v_o \bar{I}_L$, the input average current can be eliminated, and after re-arranging yields:

$$\frac{v_o}{E_i} = \frac{1}{1 + \frac{2L\bar{I}_o}{\delta^2 \tau E_i}} = \frac{1}{1 + \frac{2L\bar{I}_i}{\delta^2 \tau v_o}} \tag{17.21}$$

At a low output current or high input voltage, there is a likelihood of discontinuous inductor conduction. To avoid discontinuous conduction, larger inductance values are needed, which worsen transient response. Alternatively, with extremely low on-state duty cycles, a voltage-matching transformer can be used to increase δ. Once a transformer is used, any smps technique can be used to achieve the desired output voltage. Figures 17.2b and c show that the input current is always discontinuous.

17.1.3 Load conditions for discontinuous inductor current

As the load current decreases, the inductor average current also decreases, but the inductor ripple current magnitude is unchanged. If the load resistance is increased sufficiently, the bottom of the triangular inductor current, $\check{i}_L$, eventual reduces to zero. Any further increase in load resistance causes discontinuous inductor current and the linear voltage transfer function given by equation (17.4) is no longer valid and equations (17.16) and (17.20) are applicable. The critical load resistance for continuous inductor current is specified by

$$R_{crit} \le \frac{v_o}{\overline{I}_o} = \frac{v_o}{½\Delta i_L} \tag{17.22}$$

Substitution for v_o from equation (17.2) and using the fact that $\overline{I}_o = \overline{I}_L$, yields

$$R_{crit} \le \frac{v_o}{\overline{I}_o} = \frac{\Delta i_L L}{\overline{I}_L(\tau - t_T)} \tag{17.23}$$

Eliminating Δi_L by substituting the limiting condition given by equation (17.13) gives

$$R_{crit} \le \frac{v_o}{\overline{I}_o} = \frac{\Delta i_L L}{\overline{I}_L(\tau - t_T)} = \frac{2\overline{I}_L L}{\overline{I}_L(\tau - t_T)} = \frac{2L}{(\tau - t_T)} \tag{17.24}$$

Dividing throughout by τ and substituting $\delta = t_T / \tau$ yields

$$R_{crit} \le \frac{v_o}{\overline{I}_o} = \frac{2L}{(\tau - t_T)} = \frac{2L}{\tau(1-\delta)} \tag{17.25}$$

The critical resistance can be expressed in a number of forms. By substituting the switching frequency ($f_s = 1/\tau$) or the fundamental inductor reactance ($X_L = 2\pi f_s L$) the following forms result.

$$R_{crit} \le \frac{v_o}{\overline{I}_o} = \frac{2L}{\tau(1-\delta)} = \frac{v_o}{E_i} \times \frac{2L}{\tau\delta(1-\delta)} = \frac{2f_s L}{(1-\delta)} = \frac{X_L}{\pi(1-\delta)} \qquad (\Omega) \tag{17.26}$$

Notice that equation (17.26) is in fact equation (17.12), re-arranged.
If the load resistance increases beyond R_{crit}, the output voltage can no longer be maintained with duty cycle control according to the voltage transfer function in equation (17.4).

17.1.4 Control methods for discontinuous inductor current

Once the load current has reduced to the critical level as specified by equation (17.26), the input energy is in excess of the load requirement. Open loop load voltage regulation control is lost and the capacitor *C* tends to overcharge.
Hardware approaches can solve this problem – by producing continuous inductor current

- increase *L* thereby decreasing the inductor current ripple peak-to-peak magnitude
- step-down transformer impedance matching to effectively reduce the apparent load impedance

Two control approaches to maintain output voltage regulation when $R > R_{crit}$ are

- vary the switching frequency f_s, maintaining the switch on-time t_T constant so that Δi_L is fixed or
- reduce the switch on-time t_T, but maintain a constant switching frequency f_s, thereby reducing Δi_L.

If a fixed switching frequency is desired for all modes of operation, then reduced on-time control, using output voltage feedback, is preferred. If a fixed on-time mode of control is used, then the output voltage is control by varying inversely the frequency with output voltage. Alternatively, output voltage feedback can be used.

17.1.4i - fixed on-time t_T, variable switching frequency f_{var}

The operating frequency f_{var} is varied while the switch-on time t_T is maintained constant such that the ripple current remains unchanged. Operation is specified by equating the input energy and the output energy, thus maintaining a constant capacitor charge, hence output voltage. That is, equating energies

$$½\Delta i_L E_i t_T = \frac{v_o^2}{R}\frac{1}{f_{\text{var}}} \tag{17.27}$$

Isolating the variable switching frequency f_{var} gives

$$\begin{aligned} f_{\text{var}} &= \frac{v_o^2}{½\Delta i_L E_i t_T}\frac{1}{R} \\ f_{\text{var}} &= f_s R_{crit} \times \frac{1}{R} \\ f_{\text{var}} &\;\alpha\; \frac{1}{R} \end{aligned} \tag{17.28}$$

That is, once discontinuous inductor current occurs, if the switching frequency is varied inversely with load resistance and the switch on-state period is maintained constant, output voltage regulation can be maintained.

Load resistance *R* is not a directly or readily measurable parameter for feedback proposes. Alternatively, since $v_o = \overline{I}_o R$ substitution for *R* in equation (17.28) gives

$$\begin{aligned} f_{\text{var}} &= f_s \frac{R_{crit}}{v_o} \times \overline{I}_o \\ f_{\text{var}} &\;\alpha\; \overline{I}_o \end{aligned} \tag{17.29}$$

That is, for $\overline{I}_o < ½\Delta i_L$ or $\overline{I}_o < v_o / R_{crit}$, if t_T remains constant and f_{var} is varied proportionally with load current, then the required output voltage v_o will be maintained.

17.1.4ii - fixed switching frequency f_s, variable on-time t_{Tvar}

The operating frequency f_s remains fixed while the switch-on time t_{Tvar} is reduced, resulting in the ripple current being reduced. Operation is specified by equating the input energy and the output energy as in equation (17.27), thus maintaining a constant capacitor charge, hence voltage. That is

$$½\Delta i_L E_i t_{T\text{var}} = \frac{v_o^2}{R}\frac{1}{f_s} \tag{17.30}$$

Isolating the variable on-time t_{Tvar} yields

$$t_{T\text{var}} = \frac{v_o^2}{½\Delta i_L E_i f_s}\frac{1}{R}$$

Substituting Δi_L from equation (17.2) gives

$$\begin{aligned} t_{T\text{var}} &= t_T\sqrt{R_{crit}} \times \frac{1}{\sqrt{R}} \\ t_{T\text{var}} &\;\alpha\; \frac{1}{\sqrt{R}} \end{aligned} \tag{17.31}$$

That is, once discontinuous inductor current commences, if the switch on-time is varied inversely to the square root of the load resistance, maintaining the switching frequency constant, regulation of the output voltage can be maintained.
Again, load resistance *R* is not a directly or readily measurable parameter for feedback proposes and substitution of $v_o / \overline{I}_o$ for *R* in equation (17.31) gives

$$\begin{aligned} t_{T\text{var}} &= t_T\sqrt{\frac{R_{crit}}{v_o}} \times \sqrt{\overline{I}_o} \\ t_{T\text{var}} &\;\alpha\; \sqrt{\overline{I}_o} \end{aligned} \tag{17.32}$$

That is, if f_s is fixed and t_T is reduced proportionally to $\sqrt{\overline{I}_o}$, when $\overline{I}_o < ½\Delta i_L$ or $\overline{I}_o < v_o / R_{crit}$, then the required output voltage magnitude v_o will be maintained.

17.1.5 Output ripple voltage

Three components contribute to the output voltage ripple

- Ripple charging/discharging of the ideal output capacitor, *C*
- Capacitor equivalent series resistance, ESR
- Capacitor equivalent series inductance, ESL

The capacitor inductance and resistance parasitic series component values decrease as the quality of the capacitor increases. The output ripple voltage is the vectorial summation of the three components that are shown in figure 17.3 for the forward converter.

Ideal Capacitor: The ripple voltage for a capacitor is defined as

$$\Delta v_C = \tfrac{1}{C}\int i\,dt = \tfrac{1}{C}\Delta Q$$

Figures 17.2 and 17.3 show that for continuous inductor current, the inductor current which is the output current, swings by Δi around the average output current, $\overline{I}_o$, thus

$$\Delta v_C = \tfrac{1}{C}\int i\,dt = ½\,\tfrac{1}{C}\,\tfrac{\Delta i}{2}\,\tfrac{\tau}{2} \tag{17.33}$$

Substituting for Δi_L from equation (17.2)

$$\Delta v_C = \tfrac{1}{C}\int i\,dt = ½\,\tfrac{1}{C}\,\tfrac{\Delta i}{2}\,\tfrac{\tau}{2} = ⅛\,\tfrac{1}{C}\,\tfrac{v_o}{L} \times (\tau - t_T)\tau \tag{17.34}$$

If ESR and ESL are ignored, after rearranging, equation (17.34) gives the percentage voltage ripple (peak to peak) in the output voltage

$$\frac{\Delta v_C}{v_o} = \frac{\Delta v_o}{v_o} = \tfrac{1}{8}\frac{1}{LC}\times(1-\delta)\tau^2 = \tfrac{1}{2}\pi^2(1-\delta)\left(f_c/f_s\right)^{\frac{1}{2}} \qquad (17.35)$$

In complying with output voltage ripple requirements, from this equation, the switching frequency f_s=1/τ must be much higher that the cut-off frequency given by the forward converter low-pass, second-order LC output filter, $f_c = 1/2\pi\sqrt{LC}$. The voltage switching harmonics before filtering are the dc part δE_i and

$$V_n = \frac{\sqrt{2}E_i}{n\pi}\sqrt{1-\cos 2\pi n\delta} \qquad (17.36)$$

ESR: The equivalent series resistor voltage follows the ripple current, that is, it swings linearly about

$$V_{ESR} = \pm\tfrac{1}{2}\Delta i \times R_{ESR} \qquad (17.37)$$

ESL: The equivalent series inductor voltage is derived from $v = Ldi/dt$, that is, when the switch is on

$$V_{ESL}^{+} = L\Delta i / t_{on} = L\Delta i/\delta\tau \qquad (17.38)$$

When the switch is off

$$V_{ESL}^{-} = -L\Delta i / t_{off} = -L\Delta i/(1-\delta)\tau \qquad (17.39)$$

The total instantaneous ripple voltage is

$$\Delta v_o = \Delta v_C + V_{ESR} + V_{ESL} \qquad (17.40)$$

Forming a time domain solution for each component, then differentiating, gives a maximum ripple when

$$t = 2CR_{ESR}(1-\delta) \qquad (17.41)$$

This expression is independent of the equivalent series inductance, which is expected since it is constant during each operational state. If dominant, the inductor will affect the output voltage ripple at the switch turn-on and turn-off instants.

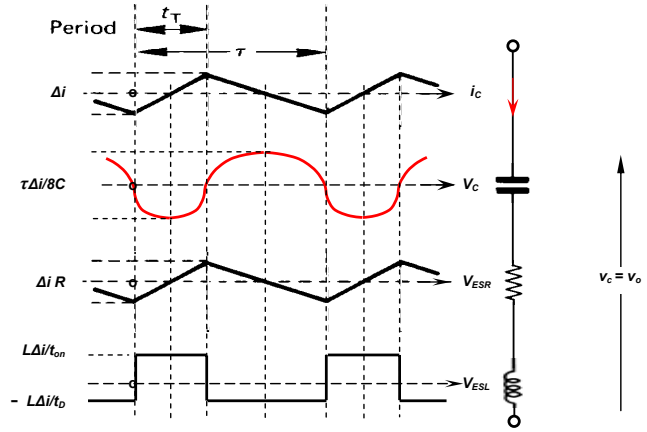


Figure 17.3. *Forward converter, three output ripple components, showing: left - voltage components; centre – waveforms; and right - capacitor model.*

Example 17.1: ***Buck (step-down forward) converter***

The step-down converter in figure 17.2a operates at a switching frequency of 10 kHz. The output voltage is to be fixed at 48 V dc across a 1 Ω resistive load. If the input voltage E_i =192 V and the choke L = 200μH:

i. calculate the switch T on-time duty cycle δ and switch on-time t_T.
ii. calculate the average load current $\bar{I}_o$, hence average input current $\bar{I}_i$.
iii. draw accurate waveforms for
- the voltage across, and the current through L; v_L and i_L
- the capacitor current, i_c
- the switch and diode voltage and current; v_T, v_D, i_T, i_D.

Hence calculate the switch utilisation ratio as defined by equation (17.11).

iv. calculate the mean and rms current ratings of diode D, switch T and L.
v. calculate the capacitor average and rms current, i_{Crms} and output ripple voltage if the capacitor has an internal equivalent series resistance of 20mΩ, assuming C = ∞.
vi. calculate the maximum load resistance R_{crit} before discontinuous inductor current. Calculate the output voltage and inductor non-conduction period, t_x, when the load resistance is triple the critical resistance R_{crit}.
vii. if the maximum load resistance is 1Ω, calculate
- the value the inductance L can be reduced, to be on the verge of discontinuous inductor current and for that L
- the peak-to-peak ripple and rms, inductor and capacitor currents.

viii. specify two control strategies for controlling the forward converter in a discontinuous inductor current mode.
ix. output ripple voltage hence percentage output ripple voltage, for C = 1,000μF and an equivalent series inductance of ESL = 0.5μH, assuming ESR = 0Ω.

Solution

i. From equation (17.4), assuming continuous inductor current, the duty cycle δ is

$$\delta = \frac{v_o}{E_i} = \frac{48\text{V}}{192\text{V}} = \tfrac{1}{4} = 25\%$$

Also, from equation (17.4), for a 10kHz switching frequency, the switching period τ is 100μs and the transistor on-time t_T is given by

$$\frac{v_o}{E_i} = \frac{t_T}{\tau} = \frac{48\text{V}}{192\text{V}} = \frac{t_T}{100\mu\text{s}}$$

whence the transistor on-time is 25μs and the diode conducts for 75μs.

ii. The average load current is $\bar{I}_o = \frac{v_o}{R} = \frac{48\text{V}}{1\Omega} = 48\text{A} = \bar{I}_L$

From power-in equals power-out, the average input current is

$$\bar{I}_i = v_o\bar{I}_o / E_i = 48\text{V}\times 48\text{A}/192\text{V} = 12\text{A}$$

iii. From equation (17.1) (or equation (17.2)) the inductor peak-to-peak ripple current is

$$\Delta i_L = \frac{E_i - v_o}{L}\times t_T = \frac{192\text{V-}48\text{V}}{200\mu\text{H}}\times 25\mu\text{s} = 18\text{A}$$

From part ii, the average inductor current is the average output current, 48A. The inductor current is continuous since $\check{i}_L = 39A$. Circuit voltage and current waveforms are shown in the figure to follow.

The circuit waveforms show that the maximum switch voltage and current are 192V and 57A respectively. The switch utilising ratio is given by equation (17.11), that is

$$SUR = \frac{P_{out}}{E_i\times\hat{i}_o} = \frac{v_o^2/R}{E_i\times\hat{i}_o} = \frac{48\text{V}^2/1\Omega}{192\text{V}\times 57\text{A}} \equiv 21\%$$

If the ripple current were assume small, the resulting SUR value of δ = 33% gives a misleading under-estimate indication.

iv. Current i_D through diode D is shown on the inductor current waveform. The average diode current is

$$\bar{I}_D = \frac{\tau - t_T}{\tau}\times\bar{I}_L = (1-\delta)\times\bar{I}_L = (1-\tfrac{1}{4})\times 48\text{A} = 36\text{A}$$

The rms diode current is given by

$$i_{Drms} = \sqrt{\frac{1}{\tau}\int_0^{\tau-t_T}(\hat{i}_L - \frac{\Delta i_L}{\tau - t_T}t)^2\,dt} = \sqrt{\frac{1}{100\mu\text{s}}\int_0^{75\mu\text{s}}(57\text{A-}\frac{18\text{A}}{75\mu\text{s}}t)^2\,dt} = 41.8\text{A}$$

Current i_T through the switch T is shown on the inductor current waveform. The average switch current is

$$\bar{I}_T = \frac{t_T}{\tau}\bar{I}_L = \delta\bar{I}_L = \tfrac{1}{4}\times 48\text{A} = 12\text{A}$$

Alternatively, from power-in equals power-out

$$\bar{I}_T = \bar{I}_i = v_o\bar{I}_o / E_i = 48\text{V}\times 48\text{A}/192\text{V} = 12\text{A}$$

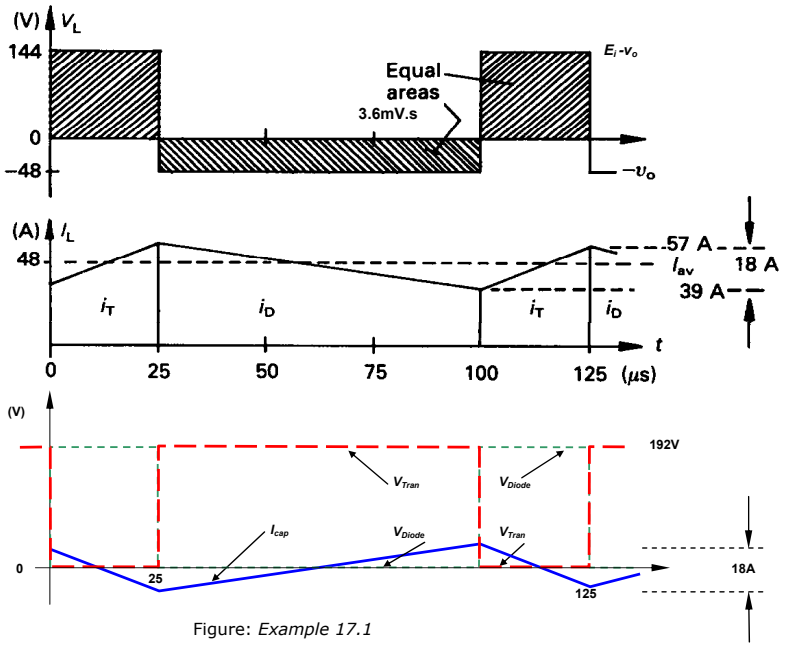


Figure: *Example 17.1*

The transistor rms current is given by

$$i_{T\,\mathrm{rms}} = \sqrt{\frac{1}{\tau}\int_0^{t_T}\left(\check{i}_L + \frac{\Delta i_L}{t_T}t\right)^2 dt} = \sqrt{\frac{1}{100\mu s}\int_0^{25\mu s}\left(39\mathrm{A} + \frac{18\mathrm{A}}{25\mu s}t\right)^2 dt}$$

$$= 24.1\mathrm{A}$$

The mean inductor current is the mean output current, $\bar{I}_o = \bar{I}_L = 48\mathrm{A}$.

The inductor rms current is given by equation (17.6), that is

$$I_{L\,rms} = \sqrt{\bar{I}_L^2 + \left(\frac{½\Delta i_L}{\sqrt{3}}\right)^2} = \sqrt{48\mathrm{A}^2 + \left(\frac{½\times 18\mathrm{A}}{\sqrt{3}}\right)^2} = 48.3\mathrm{A}$$

v. The average capacitor current $\bar{I}_C$ is zero and the rms ripple current is given by

$$i_{C\,\mathrm{rms}} = \sqrt{\frac{1}{\tau}\left[\int_0^{t_T}\left(-½\Delta i_L + \frac{\Delta i_L}{t_T}t\right)^2 dt + \int_0^{\tau - t_T}\left(½\Delta i_L - \frac{\Delta i_L}{\tau - t_T}t\right)^2 dt\right]}$$

$$= \sqrt{\frac{1}{100\mu s}\left[\int_0^{25\mu s}\left(-9\mathrm{A} + \frac{18\mathrm{A}}{25\mu s}t\right)^2 dt + \int_0^{75\mu s}\left(9\mathrm{A} - \frac{18\mathrm{A}}{75\mu s}t\right)^2 dt\right]}$$

$$= 5.2\mathrm{A} \quad \left(= \Delta i_L/2\sqrt{3}\right)$$

The capacitor voltage ripple (hence the output voltage ripple), assuming infinite output capacitance, is determined by the capacitor ripple current which is equal to the inductor ripple current, 18A p-p, that is

$$v_{o\,\mathrm{ripple}} = \Delta i_L \times R_{Cesr}$$

$$= 18\mathrm{A}\times 20\mathrm{m}\Omega = 360\mathrm{mV\ p\text{-}p}$$

and the rms output voltage ripple is

$$v_{o\,\mathrm{rms}} = i_{C\,\mathrm{rms}} \times R_{Cesr}$$

$$= 5.2\mathrm{A\ rms}\times 20\mathrm{m}\Omega = 104\mathrm{mV\ rms}$$

vi. Critical load resistance is given by equation (17.26), namely

$$R_{crit} \le \frac{v_o}{\bar{I}_o} = \frac{2L}{\tau(1-\delta)}$$

$$= \frac{2\times 200\mu\mathrm{H}}{100\mu s\times(1\text{-}¼)} = 16/3\Omega$$

$$= 5⅓\Omega \text{ when } \bar{I}_o = 9\mathrm{A}$$

Alternatively, the critical load current is 9A (½ Δi_L), thus from the equation immediately above, the load resistance must not be greater than $v_o / \bar{I}_o$ = 48V/9A = 5⅓Ω, if the inductor current is to be continuous.

When the load resistance is tripled to 16Ω the output voltage is given by equation (17.20), which is shown normalised in table 17.2. That is

$$v_o = E_i \times ¼k\delta^2\left[-1+\sqrt{1+\frac{8}{\delta^2 k}}\right] \text{ where } k = \frac{R\tau}{L} = \frac{16\Omega\times 100\mu s}{200\mu\mathrm{H}} = 8 \text{ thus}$$

$$v_o = 192\mathrm{V}\times ¼\times 8\times ¼^2\times\left[-1+\sqrt{1+\frac{8}{¼^2\times 8}}\right] = 75\mathrm{V} \qquad \left[\hat{i}_L = 14.625\mathrm{A}\right]$$

The inductor current is zero for an interval of the 100μs switching period, and the time is given by the appropriate normalised expression involving t_x for the forward converter in table 17.2 or by equation (17.16), which when re-arranged to isolate t_x becomes

$$t_x = \tau\left(1 - \frac{\delta}{v_o/E_i}\right) = 100\mu s\times\left(1 - \frac{¼}{75\mathrm{V}/50\mathrm{V}}\right) = 36\mu s \quad \left[t_T = 25\mu s \quad t_D = 39\mu s\right]$$

vii. The critical resistance formula given in equation (17.26) is valid for finding critical inductance when inductance is made the subject of the equation, that is, rearranging equation (17.26) gives

$$L_{crit} = ½\times R\times(1-\delta)\times\tau \qquad (\mathrm{H})$$

$$= ½\times 1\Omega\times(1\text{-}¼)\times 100\mu s = 37½\mu\mathrm{H}$$

This means the inductance can be reduced from 200μH with a 48A mean and 18A p-p ripple current, to 37½μH with the same 48A mean plus a superimposed 96A p-p $\left(2\bar{I}_L\right)$ ripple current. The rms capacitor current is given by

$$i_{C\mathrm{rms}} = \Delta i_L / 2\sqrt{3}$$

$$= 96\mathrm{A}/2\sqrt{3} = 27.2\mathrm{A\ rms}$$

The inductor rms current requires the following integration

$$i_{L\mathrm{rms}} = \sqrt{\frac{1}{\tau}\left[\int_0^{t_T}\left(\check{i}_L + \frac{\Delta i_L}{t_T}t\right)^2 dt + \int_0^{\tau - t_T}\left(\hat{i}_L - \frac{\Delta i_L}{\tau - t_T}t\right)^2 dt\right]}$$

$$= \sqrt{\frac{1}{100\mu s}}\times\left[\int_0^{25\mu s}\left(0 + \frac{96\mathrm{A}}{25\mu s}t\right)^2 dt + \int_0^{75\mu s}\left(96\mathrm{A} - \frac{96\mathrm{A}}{75\mu s}t\right)^2 dt\right]$$

$$= 96/\sqrt{3} = 55.4\ \mathrm{A\ rms}$$

or from equation (17.6)

$$i_{L\mathrm{rms}} = \sqrt{\bar{I}_L^2 + i_{L\mathrm{ripple}}^2}$$

$$= \sqrt{48^2 + (96/2\sqrt{3})^2} = 55.4\ \mathrm{A\ rms}$$

viii. For *R* >16/3Ω, or $\bar{I}_o < 9\mathrm{A}$, equations (17.29) or (17.32) can be used to develop a suitable control strategy.

(a) From equation (17.29), using a variable switching frequency of less than 10kHz,

$$f_{\mathrm{var}} = f_s\frac{R_{crit}}{v_o}\bar{I}_o = \ 10\mathrm{kHz}\ \frac{5⅓\Omega}{48\mathrm{V}}\ \bar{I}_o$$

$$f_{\mathrm{var}} \ = \ \frac{10}{9}\times\bar{I}_o \quad \mathrm{kHz}$$

(b) From equation (17.32), maintaining a fixed switching frequency of 10kHz, the on-time duty cycle is reduced (from 25μs) for $\bar{I}_o < 9\mathrm{A}$ according to

$$t_{T\,\mathrm{var}} = \ t_T\sqrt{\frac{R_{crit}}{v_o}}\sqrt{\bar{I}_o} = 25\mu s\sqrt{\frac{5⅓\Omega}{48\mathrm{V}}}\sqrt{\bar{I}_o}$$

$$t_{T\,\mathrm{var}} \ = \frac{25}{3}\times\sqrt{\bar{I}_o} \quad \mu s$$

ix. From equation (17.33) the output ripple voltage with an ideal 1,000μF capacitor is given by

$$\Delta v_C = \frac{\Delta i\,\tau}{8\,C}$$

$$= \frac{18\text{A}\times 100\mu\text{s}}{8\times 1000\mu\text{F}} = 225\text{mV p-p}$$

The voltage produced because of the equivalent series 0.5 μH inductance is

$$V_{ESL}^{+} = L\Delta i/\delta\tau$$

$$= 0.5\mu\text{H}\times 18\text{A}/0.25\times 100\mu\text{s} = 360\text{mV}$$

$$V_{ESL}^{-} = -L\Delta i/(1-\delta)\tau$$

$$= -0.5\mu\text{H}\times 18\text{A}/(1-0.25)\times 100\mu\text{s} = -120\text{mV}$$

Time domain summation of the capacitor and ESL inductor voltages show that the peak to peak output voltage swing is determined by the ESL inductor, giving

$$\Delta v_o = V_{ESL}^{+} - V_{ESL}^{-}$$

$$= 360\text{mV} + 120\text{mV} = 480\text{mV}$$

The percentage ripple in the output voltage is 480mV/48V = 1%.

♣

17.1.6 Underlying operational mechanisms of the forward converter

The inductor current is pivotal to the analysis and understanding of any smps. For analysis, the smps internal and external electrical conditions are in steady-state on a cycle-by-cycle basis and the input power is equal to the output power.

The first concept to appreciate is that the net capacitor charge change is zero over each switching cycle. That is, the average capacitor current is zero:

$$\overline{I}_c = \frac{1}{\tau}\int_t^{t+\tau} i_c(t)\,dt = 0$$

In so being, the output capacitor provides any load current deficit and stores any load current (inductor) surplus associated with the inductor current within each complete cycle. Thus, the capacitor is a temporary storage component where the capacitor voltage is fixed on a cycle-by-cycle basis, and because of its large capacitance does not vary significantly within a cycle.

The second concept involved is that the average inductor voltage is zero. Based on $v = L\,di/dt$, the equal area criteria in chapter 11.1.3i,

$$i_{t+\tau} - i_t = \frac{1}{L}\int_t^{t+\tau} v_L(t)\,dt = 0 \text{ since } i_{t+\tau} = i_t \text{ in steady-state}$$

Thus the average inductor voltage is zero:

$$\overline{V}_L = \frac{1}{\tau}\int_t^{t+\tau} v_L(t)\,dt = 0$$

The most enlightening way to appreciate the converter operating mechanisms is to consider how the inductor current varies with load resistance *R* and inductance *L*. The figure 17.4 shows the inductor current associated with the various parts of example 17.1.

For continuous inductor current operation, the two necessary and sufficient equations are $I_o = v_o/R$ and equation (17.2). Since the duty cycle and on-time are fixed for a given output voltage requirement, equation (17.2) can be simplified to show that the ripple current is inversely proportional to inductance, as follows

$$\Delta i_L = \frac{v_o}{L}\times(\tau - t_T)$$

$$\Delta i_L \,\alpha\, \frac{1}{L} \qquad (17.42)$$

Since the average inductor current is equal to the load current, then, at a given output voltage, the average inductor current is inversely proportional to the load resistance, that is

$$\overline{I}_L = \overline{I}_o = v_o/R$$

$$\overline{I}_L \,\alpha\, \frac{1}{R} \qquad (17.43)$$

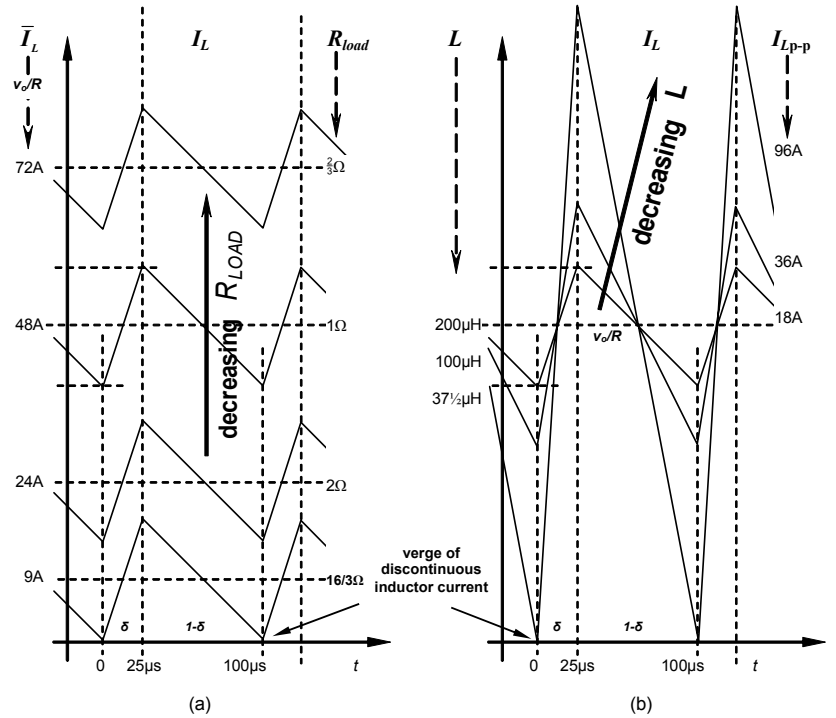


Figure 17.4. *Forward converter (buck converter) operational mechanisms showing that: (a) the average inductor current is inversely proportional to load resistance R (fixed L) and (b) the inductor ripple current magnitude is inversely proportional to inductance L (fixed load R).*

Equation (17.43) predicts that the average inductor current is inversely proportional to the load resistance, as shown in figure 17.4a. As the load is increased (load resistor decreased), the triangular inductor current moves vertically up, but importantly, from equation (17.42), the peak-to-peak ripple current is constant, that is the ripple current is independent of the load. As the load current is progressively decreased, by increasing *R*, the peak-to-peak current is unchanged; the inductor minimum current eventually reduces to zero, and discontinuous inductor current operation occurs.
Equation (17.42) indicates that the inductor ripple current is inversely proportional to inductance, as shown in figure 17.4b. As the inductance is varied the ripple current varies inversely, but importantly, from equation (17.43), the average current is constant, and specifically the average current value is not related to inductance *L* and is solely determined by the load current, v_o/R. As the inductance decreases the magnitude of the ripple current increases, the average is unchanged, and the minimum inductor current eventually reaches zero and discontinuous inductor current operation results.

17.1.7 Hysteresis voltage feedback control of the forward converter

The main function of a dc-to-dc converter is to provide a regulated output voltage, independent of input voltage or load changes, and it must respond quickly to maintain the output voltage due to any input voltage or load changes. Figure 17.5a shows a hysteresis controller for the buck and forward converters. The comparator compares the output voltage V_o to a reference voltage V_{ref}. If $V_o < V_{ref}$, the switch T_1 is turned off. If $V_o < V_{ref}$, the switch is turned on. This process is repeated continuously such that V_o is maintained at a value close to V_{ref}.
Undesirable high frequency switching action, chattering, is avoided by creating a dead band around V_{ref}. The dead band is created by using an upper boundary V_{upper} and a lower boundary V_{lower}. The region between the two boundaries is the dead band. Resistors $R_{f/b}$ and R_{ref} produce the required dead band and their values determine the upper and lower boundaries of the dead band.

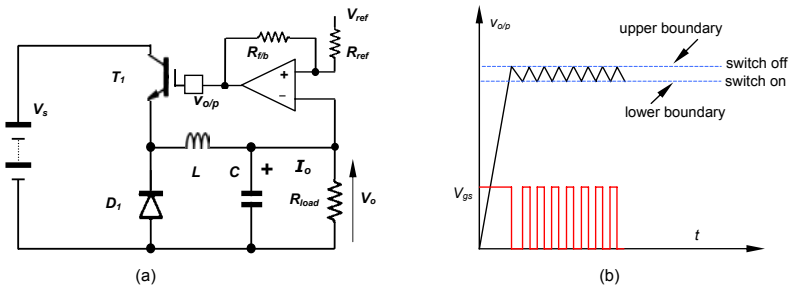


Figure 17.5: *The hysteresis controller with a dead band.*

The comparator in figure 17.5a is a Schmitt trigger. The input voltage v^+ of the positive op-amp input depends on V_{ref}, $v_{o/p}$, $R_{f/b}$, and R_{ref}. It switches from one boundary of the dead band to the other. During initial start-up of the buck converter, the op-amp negative input v^- is a small positive voltage and is less than v^+. The amplifier saturates such that $v_{o/p}$ attains the op-amp supply, thus the switch is turned on and v^+ is given by equation (17.44).

$$v^+ = \frac{R_{ref}}{R_{ref}+R_{f/b}}v_{o/p} + \frac{R_{f/b}}{R_{ref}+R_{f/b}}V_{ref} \qquad (17.44)$$

Progressively the output voltage V_o increases. The op-amp output voltage, $v_{o/p}$, remains unchanged until V_o is equal to v^+. The op-amp then enters its linear region and $v_{o/p}$ decreases, as does v^+. This continues until $v_{o/p}$ reaches zero and the op-amp saturates again. The voltage v^+ is now given by equation (17.45).

$$v^+ = \frac{R_{f/b}}{R_{ref}+R_{f/b}}V_{ref} \qquad (17.45)$$

Equations (17.44) and (17.45) represent the control boundaries of the control circuit where equation (17.44) defines the upper boundary and equation (17.45) gives the lower boundary of the dead band. The dead band is derived by subtracting equations (17.45) and (17.44), and is given by equation (17.46)

$$\Delta D_{band} = \frac{R_{f/b}}{R_{ref}+R_{f/b}}v_{o/p} \qquad (17.46)$$

$R_{f/b}$ and R_{ref} are chosen to give the required ΔD_{band} / $v_{o/p}$.

When the output voltage V_o is inside the dead band, the switch is off. Regardless of where the voltage starts, switching starts as soon as a boundary is traversed. In figure 17.5b the converter start-up process is illustrated. The switch T_1 turns on initially because the output voltage V_o is below the turn-on boundary, V_{lower}. The output voltage rises from zero to V_{upper} at a rate limited by the inductor L, the capacitor C, and the load. The switch then turns off as the output voltage crosses the upper boundary V_{upper} and remains off until the output voltage falls crosses below the lower boundary V_{lower} where the switch is turned on. Once the voltage is between the boundaries, within the hysteresis bounds, on-off switching action attempts to maintain the current within the boundaries under all conditions.
The operation of the system becomes independent of the input, the load, the inductor, and the capacitor values. The system tracks the desired voltage V_{ref} even if the component values or the load changes drastically. A drawback is that the controller gives rise to an overvoltage during start-up. This problem can be solved by the correct selection of the inductor and capacitor values that allow the output voltage to rise exponentially and settle somewhat close to the desired output voltage, while maintaining the desired ripple voltage. In power electronics terms, the major limitation is possible broadband EMC generation due to a widely varying switching frequency. The closer the hysteresis bounds, the higher the upper frequency, the wider the frequency variation.

Design Procedure
In the steady-state, the converter output voltage depends on the input voltage V_s, the switching frequency f_s, and the on duration of the switching period t_{on} and is given by equation (17.4):

$$V_o = t_{on} f_s V_s = \delta V_s$$

The product t_{on}xf_s is defined as the duty ratio δ. The output voltage V_o is regulated by changing δ while f_s is kept constant. This pulse width modulation method is widely used in dc-dc converters.
Another approach to regulate V_o is to vary f_s, keeping δ constant. However, this is undesirable because it is difficult to filter the wide bandwidth ripple in the input and output signals of the converter.
Hysteresis control of the buck converter is fixed boundary control. V_o is regulated by the switching action of the switch T_1 as the output V_o crosses the upper or the lower boundary of the hysteresis dead band. In

hysteresis control, f_s and δ are not fixed, but change with the converter conditions. For a given set of converter parameters, both f_s and δ are determined by the hysteresis boundaries, thus frequency f_s and the duty ratio δ are not control parameters in the design of hysteresis controllers. No matter what type of control is used, the basic operating principles of the buck converter do not change. The converter output voltage ripple depends on the dead band of the controller. As the dead band increases (or decreases), the output voltage ripple increases (or decreases) in conjunction with the switching frequency decreasing (or increasing). The voltage ripple specification can be ensured by setting the dead band of the controller at 50% of the ripple specifications with a suitable inductor value.
A properly designed hysteresis controller has excellent steady-state and dynamic properties. It responds quickly to step voltage set point changes. Fixed boundary controllers are stable under extreme disturbance conditions and can be chosen to guarantee ripple specifications or other converter operating constraints.

Example 17.2: ***Hysteresis controlled buck converter***

A dc-dc buck converter is to be regulated with voltage-based hysteresis control. The output voltage V_o = 5V and the load varies between 1 and 5Ω. The input voltage V_s also varies between 16V and 24V with a nominal value of 20V. The maximum ripple voltage is to be limited to ±1% and the nominal switching frequency is f_s = 100kHz.
Design to necessary controller and specify the converter L and C values.

Solution

The output voltage ripple is

$$\frac{\Delta V_o}{V_o} = 2\%$$
$$\Delta V_o = 0.02 \times 5\text{V} = 0.1\text{V}$$

To fulfil the required voltage ripple specification, the hysteresis band is chosen to be 50% of the output voltage ripple to account for an increase in the ripple magnitude due to the natural response of the converter RLC circuit, after the switch is turned off. The hysteresis band is

$$\Delta D_{band} = 0.5 \times 0.1\text{V} = 0.05\text{V}$$

Solving equation (17.46) for $R_{f/b}$:

$$R_{f/b} = R_{ref}\left(\frac{V_o}{\Delta D_b} - 1\right)$$

Let R_{ref} = 100Ω and v_o = 10V. Then $R_{f/b}$ is

$$R_{f/b} = 100\left(\frac{10\text{V}}{0.05\text{V}} - 1\right) = 19.9\text{k}\Omega$$

These resistances produce a hysteresis band that fulfils the output ripple requirement. The maximum and minimum values of the load current are

$$\hat{I}_o = \frac{V_o}{R_{\min}} = \frac{5\text{V}}{1\Omega} = 5\text{A}$$

and

$$\check{I}_o = \frac{V_o}{R_{\max}} = \frac{5\text{V}}{5\Omega} = 1\text{A}$$

Let the inductor current and the capacitor voltage swings be 10%. The inductor must limit the current swing at maximum load. The total current swing is as follows. Since

$$\frac{\Delta I_L}{\hat{I}_o} = 10\%$$
$$\Delta I_L = 0.1 \times 5\text{A} = 0.5\text{A}$$

The capacitor voltage swing is

$$\frac{\Delta V_c}{\Delta V_o} = 10\%$$
$$\Delta V_c = 0.1 \times 0.1\text{V} = 0.01\text{V}$$

Since the waveform of the output ripple voltage is approximately sinusoidal, accounting for the equivalent series resistance, ESR, of the capacitor C, the output ripple voltage is then

$$\Delta V_o = \sqrt{\Delta V_c^2 + \Delta V_{ESR}^2}$$

where ΔV_{ESR} is the voltage ripple across the capacitor resistance R_{ESR}. The ΔV_{ESR} is usually much greater than ΔV_c, thus a close approximation of the peak-to-peak output voltage ripple is:

$$\begin{aligned}\Delta V_o &\cong \Delta V_{ESR}\\ &\cong \Delta I_c R_{ESR}\\ &\cong (\Delta I_L - \Delta I_R) R_{ESR}\end{aligned}$$

With the result for ΔI_L, the inductor L is

$$L = \frac{V_o(V_s - V_o)}{V_s f_s \Delta I_L} = \frac{5\text{V}(20\text{V} - 5\text{V})}{20\text{V} \times 100\text{kHz} \times 0.5\text{A}} = 75\mu\text{H}$$

The capacitance is

$$C = \frac{\Delta I_L}{8 f_s \Delta V_c} = \frac{0.5\text{A}}{8 \times 100\text{kHz} \times 0.01\text{V}} = 62.5\mu\text{F} \cong 68\mu\text{F}$$

The ESR of the capacitor can be determined from Eq. (7.13):

$$R_{ESR} = \frac{\Delta V_o}{\Delta I_L - \Delta I_R} = \frac{0.1\text{V}}{0.5\text{A} - 0.1\text{A}} = ¼\Omega$$

Transient overshoot, undershoot, and recovery time to step load and input changes are important performance parameters in buck converters. Since the current in the inductor cannot change instantaneously, the transient response is inherently inferior to that of the linear regulators. The recovery time to step changes in the line and the load is controlled by the characteristic of the controller feedback loop. Transient overshoot and undershoot resulting from step load changes can be analyzed and calculated as follows. The ac output impedance is

$$Z_{out} = \frac{V_s - V_o}{\Delta I_{load}}$$

Since

$$V_L = -L\frac{di_L}{dt} \quad \text{and} \quad I_o = C\frac{dv}{dt}$$

thus

$$Z_{out} = \frac{LI_o}{(V_s - V_o)C}$$

As a result, for an increasing load current, from 1A to 5A, the change in the output voltage (transient undershoot) is:

$$\begin{aligned}\Delta\check{V}_o &= \Delta I_o Z_{out} = \frac{L\Delta I_o^2}{(V_s - V_o)C}\\ &= \frac{75\mu\text{H} \times (5\text{A} - 1\text{A})^2}{(20\text{V} - 5\text{V}) \times 68\mu\text{F}} = 1.231\text{V}\end{aligned}$$

and for a decreasing load current, from 5A to 1A, the change in the output voltage (transient overshoot) is

$$\begin{aligned}\Delta\hat{V}_o &= \frac{LI_o^2}{V_o C}\\ &= \frac{75\mu\text{H} \times (5\text{A} - 1\text{A})^2}{5\text{V} \times 68\mu\text{F}} = 3.692\text{V}\end{aligned}$$

♣

17.2 Flyback converters

Flyback converters store energy in an inductor, ('choke'), before transferring any energy to the load and output capacitor such that controllable output voltage magnitudes in excess of the input voltage are attainable. The key characteristic is that whilst energy is being transferred to the inductor, load energy is provided by the output capacitor. Such converters are also known as *ringing choke* converters.

Two basic (minimum component count and transformerless) versions of the flyback converter are possible, both are integral to the same underlying fundamental circuit configuration (see section 17.5).

- The step-up voltage flyback converter, called the *boost converter,* where the input and output voltage have the same polarity - non-inversion, and $v_o \geq E_i$.
- The step-up/step-down voltage flyback converter, called the *buck-boost converter,* where output voltage polarity inversion occurs, that is $v_o \geq 0$.

17.3 The boost converter

The *boost converter* transforms a dc voltage input to a dc voltage output that is greater in magnitude but has the same relative polarity as the input. The basic circuit configuration is shown in figure 17.6a. It will be seen that when the transistor is off, the output capacitor is charged to the input voltage E_i. Inherently, the output voltage v_o can never be less than the input voltage level.

When the transistor T is turned on, the supply voltage E_i is applied across the inductor L and the diode D is reverse-biased by the output voltage v_o. Energy is transferred from the supply to L and when the transistor is turned off this energy is transferred to the load and output capacitor through D. While the inductor is transferring its stored energy into C and the load, energy is also being provided from the input source.

The output current is always discontinuous, but the input current can be either continuous or discontinuous. For analysis, assume $v_o > E_i$ and a constant input and output voltage. Inductor currents are then linear and vary according to $v = L\,di/dt$.

17.3.1 Continuous inductor current

The circuit voltage and current waveforms for continuous inductor conduction are shown in figure 17.6b. The inductor current excursion, from $v = L\,di/dt$, which is the input current excursion, during the switch on-time t_T and switch off-time $\tau - t_T$, is given by

$$\Delta i_L = \frac{(v_o - E_i)}{L}(\tau - t_T) = \frac{E_i}{L}t_T \qquad (17.47)$$

After rearranging, the voltage and current transfer function is given by

$$\frac{v_o}{E_i} = \frac{\bar{I}_i}{\bar{I}_o} = \frac{1}{1-\delta} \qquad (17.48)$$

where $\delta = t_T/\tau$, t_T is the transistor on-time, and $P_{in} = P_{out}$, that is, $E_i\,I_i = v_o\,I_o$ is assumed.

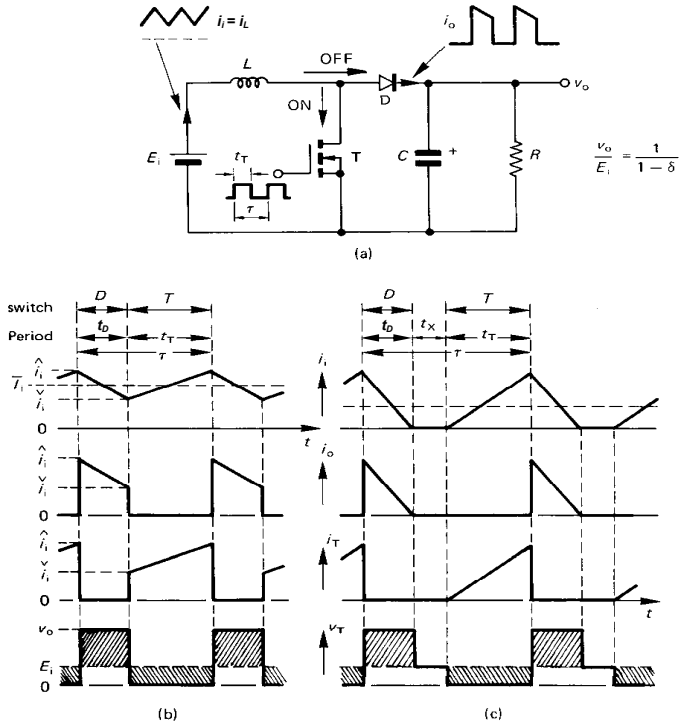


Figure 17.6. *Non-isolated, step-up, flyback converter (boost converter) where $v_0 \geq E_1$: (a) circuit diagram; (b) waveforms for continuous input current; and (c) waveforms for discontinuous input current.*

The maximum inductor current, which is the maximum input current, $\hat{i}_L$, using equation (17.47) and $v_o = I_o R$, is given by

$$\begin{aligned}\hat{i}_L &= \bar{I}_L + \tfrac{1}{2}\Delta i_L = \bar{I}_i + \tfrac{1}{2}\frac{E_i t_T}{L} \\ &= \frac{\bar{I}_o}{1-\delta} + \tfrac{1}{2}\frac{v_o}{L}(1-\delta)\delta\tau = v_o\left[\frac{1}{(1-\delta)R} + \frac{(1-\delta)\delta\tau}{2L}\right]\end{aligned} \tag{17.49}$$

while the minimum inductor current, $\check{i}_L$ is given by

$$\begin{aligned}\check{I}_L &= \bar{I}_L - \tfrac{1}{2}\Delta i_L = \bar{I}_i - \tfrac{1}{2}\frac{E_i t_T}{L} \\ &= \frac{\bar{I}_o}{1-\delta} - \tfrac{1}{2}\frac{v_o}{L}(1-\delta)\delta\tau = v_o\left[\frac{1}{(1-\delta)R} - \frac{(1-\delta)\delta\tau}{2L}\right]\end{aligned} \tag{17.50}$$

For continuous conduction $\check{i}_L \geq 0$, that is, from equation (17.50)

$$\bar{I}_L \geq \tfrac{1}{2}\frac{E_i t_T}{L} = \tfrac{1}{2}\frac{v_o(1-\delta)t_T}{L} \tag{17.51}$$

The inductor rms ripple current (and input ripple current in this case) is given by

$$i_{Lr} = \frac{\Delta i_L}{2\sqrt{3}} = \frac{1}{2\sqrt{3}}\frac{v_o}{L}(1-\delta)\delta\tau \tag{17.52}$$

The harmonic components in the input current are

$$I_{in} = \frac{\sqrt{2}\,E_i\tau\sin n\delta\pi}{2\pi^2 n^2(1-\delta)L} = \frac{\sqrt{2}\,v_o\tau\sin n\delta\pi}{2\pi^2 n^2 L} \tag{17.53}$$

while the inductor total rms current is

$$i_{L\text{rms}} = \sqrt{\bar{I}_L^2 + i_{Lr}^2} = \sqrt{\bar{I}_L^2 + \left(\frac{\tfrac{1}{2}\Delta i_L}{\sqrt{3}}\right)^2} = \sqrt{\frac{1}{3}\left(\hat{i}_L^2 + \hat{i}_L\times\check{i}_L + \check{i}_L^2\right)} \tag{17.54}$$

The switch and diode average and rms currents are given by

$$\begin{aligned}\bar{I}_T &= \bar{I}_i - \bar{I}_o = \delta\bar{I}_i = \delta\bar{I}_L & I_{T\text{rms}} &= \sqrt{\delta}\,i_{L\text{rms}} \\ \bar{I}_D &= (1-\delta)\bar{I}_i = \bar{I}_o & I_{D\text{rms}} &= \sqrt{1-\delta}\,i_{L\text{rms}}\end{aligned} \tag{17.55}$$

Switch utilisation ratio

The switch utilisation ratio, SUR, is a measure of how fully a switching device's power handling capabilities are utilised in any switching application. The ratio is defined as

$$SUR = P_{out} \Big/ p\hat{V}_T\hat{I}_T \tag{17.56}$$

where p is the number of power switches in the circuit; $p=1$ for the boost converter. The switch maximum instantaneous voltage and current are $\hat{V}_T$ and $\hat{I}_T$ respectively. As shown in figure 17.6b, the maximum switch voltage supported in the off-state is v_o, while the maximum current is the maximum inductor current $\hat{i}_L$ which is given by equation (17.49). If the inductance L is large such that the ripple current is small, the peak inductor current is approximated by the average inductor current such that $\hat{I}_T \approx \bar{I}_L = \bar{I}_o/1-\delta$, that is

$$SUR = \frac{v_o\bar{I}_o}{v_o\times\bar{I}_o/1-\delta} = 1-\delta \tag{17.57}$$

which assumes continuous inductor current. This result shows that the lower the duty cycle, that is the closer the step-up voltage v_o is to the input voltage E_i, the better the switch I-V ratings are utilised.

17.3.2 Discontinuous capacitor charging current in the switch off-state

It is possible that the input current (inductor current) falls below the output (resistor) current during a part of the cycle when the switch is off and the inductor is transferring energy to the output circuit. Under such conditions, towards the end of the off period, part of the load current requirement is provided by the capacitor even though this is the period during which its charge is replenished by inductor energy. The circuit independent transfer function in equation (17.48) remains valid. This *C* discontinuous charging condition commences when the minimum inductor current $\check{i}_L$ and the output current I_o are equal. That is

$$\begin{aligned}&\check{I}_L - \bar{I}_o \leq 0 \\ &\bar{I}_L - \tfrac{1}{2}\Delta i_L - \bar{I}_o \leq 0 \\ &\frac{\bar{I}_o}{1-\delta} - \tfrac{1}{2}\frac{E_i\delta\tau}{L} - \bar{I}_o \leq 0\end{aligned} \tag{17.58}$$

which yields

$$\delta \leq 1 - \sqrt{\frac{2L}{\tau R}} \tag{17.59}$$

17.3.3 Discontinuous inductor current

If the inequality in equation (17.51) is not satisfied, the input current, which is also the inductor current, reaches zero and discontinuous inductor conduction occurs during the switch off period. Various circuit voltage and current waveforms for discontinuous inductor conduction are shown in figure 17.6c.
The onset of discontinuous inductor current operation occurs when the minimum inductor current $\check{i}_L$, reaches zero. That is, with $\check{i}_L = 0$ in equation (17.50), the last equality

$$\frac{1}{(1-\delta)R} - \frac{(1-\delta)\delta\tau}{2L} = 0 \tag{17.60}$$

relates circuit component values (R and L) and operating conditions (f and δ) at the verge of discontinuous inductor current.

With $\check{i}_L = 0$, the output voltage is determined as follows

$$\hat{i}_L = \frac{E_i t_T}{L} = \frac{(v_o - E_i)}{L}(\tau - t_T - t_x) \tag{17.61}$$

yielding

$$\frac{v_o}{E_i} = \frac{1 - \frac{t_x}{\tau}}{1 - \frac{t_x}{\tau} - \delta} \tag{17.62}$$

Alternatively, using

$$\hat{i}_L = \frac{E_i t_T}{L}$$

and

$$\bar{I}_L - \bar{I}_o = \tfrac{1}{2}\delta\hat{i}_L$$

yields

$$\frac{2}{\delta}(\bar{I}_L - \bar{I}_o) = \frac{E_i t_T}{L}$$

Assuming power-in equals power-out and $\bar{I}_L = \bar{I}_i$

$$\frac{2}{\delta}\bar{I}_o\left(\frac{v_o}{E_i} - 1\right) = \frac{E_i t_T}{L}$$

that is

$$\frac{v_o}{E_i} = 1 + \frac{E_i\tau\delta^2}{2L\bar{I}_o} = 1 + \frac{v_o\tau\delta^2}{2L\bar{I}_i} \tag{17.63}$$

or

$$\frac{v_o}{E_i} = \frac{1}{1 - \frac{E_i\tau\delta^2}{2L\bar{I}_i}} \tag{17.64}$$

On the verge of discontinuous conduction (when equation (17.48) is valid), these equations can be rearranged to give

$$\bar{I}_o = \frac{E_i}{2L}\tau\delta(1-\delta) \tag{17.65}$$

At a low output current or low input voltage, there is a likelihood of discontinuous inductor current conduction. (See appendix 17.11.) To avoid discontinuous conduction, larger inductance values are needed, which worsen the transient response. Alternatively, with extremely high on-state duty cycles, (because of a low input voltage E_i) a voltage-matching step-up transformer can be used to decrease δ. Figures 17.6b and c show that the output current is always discontinuous.

17.3.4 Load conditions for discontinuous inductor current

As the load current decreases, the inductor average current also decreases, but the inductor ripple current magnitude is unchanged. If the load resistance is increased sufficiently, the bottom of the triangular inductor current, $\check{i}_L$, eventually reduces to zero. Any further increase in load resistance causes discontinuous inductor current and the voltage transfer function given by equation (17.48) is no longer valid and equations (17.62) and (17.63) are applicable. (Certain circuit parameter values - L, R, and τ - can avoid discontinuous conduction for all δ. See appendix 17.11.) The critical load resistance for continuous inductor current is specified by

$$R_{crit} \le \frac{v_o}{\overline{I}_o} \tag{17.66}$$

Eliminating the output current by using the fact that power-in equals power-out and $\overline{I}_i = \overline{I}_L$, yields

$$R_{crit} \le \frac{v_o}{\overline{I}_o} = \frac{v_o^2}{E_i \overline{I}_L} \tag{17.67}$$

Using $\overline{I}_L = ½\Delta i_L$ then substituting with the right hand equality of equation (17.47), halved, gives

$$R_{crit} \le \frac{v_o}{\overline{I}_o} = \frac{v_o^2}{E_i \overline{I}_L} = \frac{v_o^2 2L}{E_i^2 t_T} = \frac{2L}{\tau\delta(1-\delta)^2} \tag{17.68}$$

The critical resistance can be expressed in a number of forms. By substituting the switching frequency ($f_s = 1/\tau$) or the fundamental inductor reactance ($X_L = 2\pi f_s L$) the following forms result.

$$R_{crit} \le \frac{v_o}{\overline{I}_o} = \frac{2L}{\tau\delta(1-\delta)^2} = \frac{v_o}{E_i} \times \frac{2L}{\tau\delta(1-\delta)} = \frac{2f_s L}{\delta(1-\delta)^2} = \frac{X_L}{\pi\delta(1-\delta)^2} \quad (\Omega) \tag{17.69}$$

Equation (17.69) is equation (17.60), re-arranged.
If the load resistance increases beyond R_{crit}, generally the output voltage can no longer be maintained with purely duty cycle control according to the voltage transfer function in equation (17.48).

17.3.5 Control methods for discontinuous inductor current

Once the load current has reduced to the critical level as specified by equation (17.69), the input energy is in excess of the load requirement. Open loop load voltage regulation control is lost and the capacitor C tends to overcharge, thereby increasing v_o.

Hardware approaches can be used to solve this problem – by ensuring continuous inductor current

- increase L thereby decreasing the inductor current ripple p-p magnitude
- step-down transformer impedance matching to effectively reduce the apparent load impedance

Two control approaches to maintain output voltage regulation when $R > R_{crit}$ are

- vary the switching frequency f_s, maintaining the switch on-time t_T constant so that Δi_L is fixed or
- reduce the switch on-time t_T, but maintain a constant switching frequency f_s, thereby reducing Δi_L.

If a fixed switching frequency is desired for all modes of operation, then reduced on-time control, using output voltage feedback, is preferred. If a fixed on-time mode of control is used, then the output voltage is control by inversely varying the frequency with output voltage. Alternatively, output voltage feedback can be used.

17.3.5i - fixed on-time t_T, variable switching frequency f_{var}

The operating frequency f_{var} is varied while the switch-on time t_T is maintained constant such that the ripple current remains unchanged. Operation is specified by equating the input energy and the output energy, thus maintaining a constant capacitor charge, hence output voltage. That is, equating energies

$$½\Delta i_L E_i \tau = \frac{v_o^2}{R} \frac{1}{f_{var}} \tag{17.70}$$

Isolating the variable switching frequency f_{var} gives

$$f_{var} = \frac{v_o^2}{½\Delta i_L E_i \tau} \frac{1}{R} = f_s R_{crit} \times \frac{1}{R}$$

$$f_{var} \;\alpha\; \frac{1}{R} \tag{17.71}$$

Load resistance R is not a directly or readily measurable parameter for feedback proposes. Alternatively, since $v_o = \overline{I}_o R$, substitution for R in equation (17.71) gives

$$f_{var} = f_s \frac{R_{crit}}{v_o} \times \overline{I}_o$$

$$f_{var} \;\alpha\; \overline{I}_o \tag{17.72}$$

That is, for discontinuous inductor current, namely $\overline{I}_i < ½\Delta i_L$ or $\overline{I}_o < v_o / R_{crit}$, if the switch on-state period t_T remains constant and f_{var} is either varied proportionally with load current or varied inversely with load resistance, then the required output voltage v_o will be maintained.

17.3.5ii - fixed switching frequency f_s, variable on-time t_{Tvar}

The operating frequency f_s remains fixed while the switch-on time t_{Tvar} is reduced such that the ripple current can be reduced. Operation is specified by equating the input energy and the output energy as in equation (17.70), thus maintaining a constant capacitor charge, hence voltage. That is

$$½\Delta i_L E_i t_{T\,var} = \frac{v_o^2}{R} \frac{1}{f_s} \tag{17.73}$$

Isolating the variable on-time t_{Tvar} gives

$$t_{T\,var} = \frac{v_o^2}{½\Delta i_L E_i f_s} \frac{1}{R}$$

Substituting Δi_L from equation (17.47) gives

$$t_{T\,var} = t_T \sqrt{R_{crit}} \times \frac{1}{\sqrt{R}}$$

$$t_{T\,var} \;\alpha\; \frac{1}{\sqrt{R}} \tag{17.74}$$

Again, load resistance R is not a directly or readily measurable parameter for feedback proposes and substitution of $v_o / \overline{I}_o$ for R in equation (17.74) gives

$$t_{T\,var} = t_T \sqrt{\frac{R_{crit}}{v_o}} \times \sqrt{\overline{I}_o}$$

$$t_{T\,var} \;\alpha\; \sqrt{\overline{I}_o} \tag{17.75}$$

That is, if the switching frequency f_s is fixed and switch on-time t_T is reduced proportionally to $\sqrt{\overline{I}_o}$ or inversely to $\sqrt{R}$, when discontinuous inductor current commences, namely $\overline{I}_i < ½\Delta i_L$ or $\overline{I}_o < v_o / R_{crit}$, then the required output voltage magnitude v_o will be maintained.

17.3.6 Output ripple voltage

The output ripple voltage is the capacitor ripple voltage. The ripple voltage for a capacitor is defined as

$$\Delta v_o = \frac{1}{C}\int i\,dt = \frac{1}{C}\Delta Q$$

Figure 17.6 shows that for continuous inductor current, the constant output current $\overline{I}_o$ is provided solely from the capacitor during the period t_T when the switch is on, thus

$$\Delta v_o = \frac{1}{C}\int i\,dt = \frac{1}{C} t_T \overline{I}_o$$

Substituting for $\overline{I}_o = v_o / R$ gives

$$\Delta v_o = \frac{1}{C}\int i\,dt = \frac{1}{C} t_T \overline{I}_o = \frac{1}{C} t_T \frac{v_o}{R}$$

Rearranging gives the percentage voltage ripple (peak to peak) in the output voltage

$$\frac{\Delta v_o}{v_o} = \frac{\delta\tau}{RC} \tag{17.76}$$

The capacitor equivalent series resistance and inductance can be account for, as with the forward converter, 17.1.4. When the switch conducts, the output current is constant and is provided from the capacitor. Thus no ESL voltage effects result during this constant capacitor current portion of the cycle.

Example 17.3: ***Boost (step-up flyback) converter***

The boost converter in figure 17.6 is to operate with a 50µs transistor fixed on-time in order to convert the 50 V input up to 75 V at the output. The inductor is 250µH and the resistive load is 2.5Ω.

i. Calculate the switching frequency, hence transistor off-time, assuming continuous inductor current.
ii. Calculate the mean input and output current.
iii. Draw the inductor current, showing the minimum and maximum values.

iv. Calculate the capacitor rms ripple current.
v. Derive general expressions relating the operating frequency to varying load resistance.
vi. At what load resistance does the instantaneous input current fall below the output current.

Solution

i. From equation (17.48), which assumes continuous inductor current

$$\frac{v_o}{E_i} = \frac{1}{1-\delta} \quad \text{where} \quad \delta = \frac{t_T}{\tau}$$

that is

$$\frac{75\text{V}}{50\text{V}} = \frac{1}{1-\delta} \quad \text{where} \quad \delta = \frac{50\mu\text{s}}{\tau} = \tfrac{1}{3}$$

That is, τ = 150 µs or $f_s = 1/\tau$ = 6.66 kHz, with a 100µs switch off-time.

ii. The mean output current $\bar{I}_o$ is given by

$$\bar{I}_o = v_o / R = 75\text{V}/2.5\Omega = 30\text{A}$$

From power transfer considerations

$$\bar{I}_i = \bar{I}_L = v_o \bar{I}_o / E_i = 75\text{V}\times 30\text{A}/50\text{V} = 45\text{A}$$

iii. From $v = L\, di/dt$, the ripple current $\Delta i_L = E_i t_T / L$ = 50V x 50µs /250 µH = 10 A

that is

$$\hat{i}_L = \bar{I}_L + \tfrac{1}{2}\Delta i_L = 45\text{A} + \tfrac{1}{2}\times 10\text{A} = 50\text{A}$$

$$\check{i}_L = \bar{I}_L - \tfrac{1}{2}\Delta i_L = 45\text{A} - \tfrac{1}{2}\times 10\text{A} = 40\text{A}$$

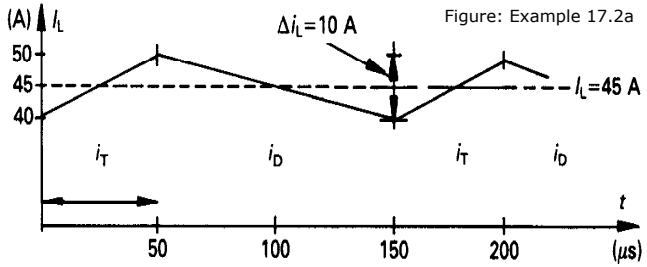


Figure: Example 17.2a

iv. The capacitor current is derived by using Kirchhoff's current law such that at any instant in time, the diode current, plus the capacitor current, plus the 30A constant load current into R, all sum to zero.

$$i_{C\text{rms}} = \sqrt{\frac{1}{\tau}\left[\int_0^{t_T} \bar{I}_o^2 dt + \int_0^{\tau - t_T} \left(\frac{\Delta i_L}{\tau - t_T} t - \hat{i}_L + \bar{I}_o\right)^2 dt\right]}$$

$$= \sqrt{\frac{1}{150\mu\text{s}}\left[\int_0^{50\mu\text{s}} 30\text{A}^2 dt + \int_0^{100\mu\text{s}} \left(\frac{10\text{A}}{100\mu\text{s}} t - 20\text{A}\right)^2 dt\right]} = 21.3\text{A}$$

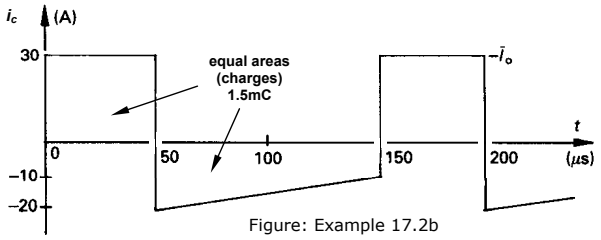


Figure: Example 17.2b

v. The critical load resistance, R_{crit}, produces an input current with Δi_L = 10 A ripple. Since the energy input equals the energy output

$$\tfrac{1}{2}\Delta i \times E_i \times \tau = v_o \times v_o / R_{crit} \times \tau$$

that is

$$R_{crit} = \frac{2v_o^2}{E_i \Delta i} = \frac{2\times 75\text{V}^2}{50\text{V}\times 10\text{A}} = 22\tfrac{1}{2}\Omega$$

Alternatively, equation (17.69) or equation (17.51) can be rearranged to give R_{crit}.
For a load resistance of less than 22½ Ω, continuous inductor current flows and the operating frequency is fixed at 6.66 kHz with $\delta = \frac{1}{3}$, that is

f_s = 6.66 kHz for all $R \le 22\tfrac{1}{2}\ \Omega$

For load resistance greater than 22½ Ω, (< v_o / R_{crit} = 3⅓A), the energy input occurs in 150 µs burst whence from equation (17.70)

$$\tfrac{1}{2}\Delta i_L E_i \times 150\mu\text{s} = \frac{v_o^2}{R}\frac{1}{f_{\text{var}}}$$

that is

$$f_{\text{var}} = \frac{R_{crit}}{\tau}\frac{1}{R} = \frac{22\tfrac{1}{2}\Omega}{150\mu\text{s}}\frac{1}{R}$$

$$f_{\text{var}} = \frac{150}{R}\ \text{kHz} \quad \text{for} \quad R \ge 22\tfrac{1}{2}\Omega$$

vi. The ±5A inductor ripple current is independent of the load, provided the critical resistance is not exceeded. When the average inductor current (input current) is less than 5A more than the output current, the capacitor must provide load current not only when the switch is on but also when the switch is off. The transition is given by equation (17.59), that is

$$\delta \le 1 - \sqrt{\frac{2L}{\tau R}}$$

$$\frac{1}{3} \le 1 - \sqrt{\frac{2\times 250\mu\text{H}}{150\mu\text{s}\times R}}$$

This yields $R \ge 7\tfrac{1}{2}\Omega$ and a load current of 10A. The average inductor current is 15A, with a minimum value of 10A, the same as the load current. That is, for $R < 7\tfrac{1}{2}\Omega$ all the load requirement is provided from the input inductor when the switch is off, with excess energy charging (replenishing) the output capacitor. For $R > 7\tfrac{1}{2}\Omega$ insufficient energy is available from the inductor to provide the load energy throughout the whole of the period when the switch is off. The capacitor supplements the load requirement towards the end of the off period. When $R > 22\tfrac{1}{2}\Omega$ (the critical resistance), discontinuous inductor current occurs, and the duty cycle dependent transfer function is no longer valid.

♣

Example 17.4: ***Alternative boost (step-up flyback) converter***

The alternative boost converters (producing a dc supply either above E_i (left) or below 0V (right) – see figure 17.9b) shown in the following figure are to operate under the same conditions as the boost converter in example 17.3, namely, with a 50µs transistor fixed on-time in order to convert the 50 V input up to 75 V at the output. The energy transfer inductor is 250µH and the resistive load is 2.5Ω.

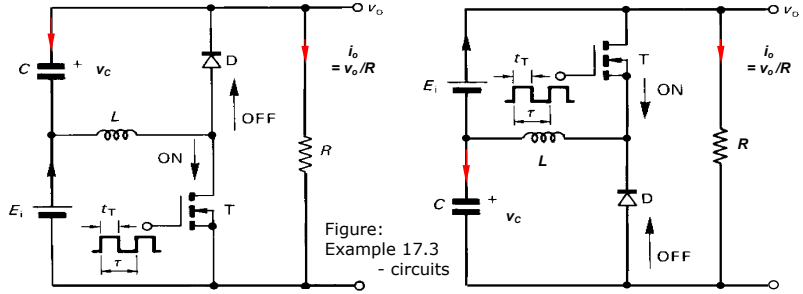


Figure: Example 17.3 - circuits

i. Derive the voltage transfer ratio and critical resistance expression for the alternative boost converter, hence showing the control performance is identical to the boost converter shown in figure 17.6.

ii. By considering circuit voltage and current waveforms, identify how the two boost converters differ from the conventional boost circuit in figure 17.6. Use example 17.3 for a comparison basis.

Solution

i. Assuming non-zero, continuous inductor current, the inductor current excursion, from $v = L di/dt$, which for this boost converter is not the input current excursion, during the switch on-time t_T and switch off-time $\tau - t_T$, is given by

$$L\Delta i_L = E_i t_T = v_C(\tau - t_T)$$

but $v_C = v_o - E_i$, thus substitution for v_C gives

$$E_i t_T = (v_o - E_i)(\tau - t_T)$$

and after rearranging

$$\frac{v_o}{E_i} = \frac{\bar{I}_i}{\bar{I}_o} = \frac{1}{1-\delta} \quad \left(= 1 + \frac{\delta}{1-\delta}\text{: that is } v_o \geq E_i \text{ alternately } E_i + \delta v_o = v_o\right)$$

where $\delta = t_T/\tau$ and t_T is the transistor on-time. This is the same voltage transfer function as for the conventional boost converter, equation (17.48). This result would be expected since both converters have the same ac equivalent circuit. Similarly, the critical resistance would be expected to be the same for each boost converter variation.

Examination of the switch on and off states shows that during the switch on-state, energy is transfer to the load from the input supply, independent of switching action. This mechanism is analogous to ac autotransformer action where the output current is due to both transformer action and the input current being directed to the load.

The critical load resistance for continuous inductor current is specified by $R_{crit} \leq v_o / \bar{I}_o$.

By equating the capacitor net charge flow, the inductor current is related to the output current by $\bar{I}_L = \bar{I}_o/(1-\delta)$. At minimum inductor current, $\bar{I}_L = \tfrac{1}{2}\Delta i_L$ and substituting with $\Delta i_L = E_i t_T / L$, gives

$$R_{crit} \leq \frac{v_o}{\bar{I}_o} = \frac{v_o}{(1-\delta)\bar{I}_L} = \frac{v_o}{(1-\delta)\tfrac{1}{2}\Delta i_L} = \frac{v_o}{(1-\delta)\tfrac{1}{2}E_i t_T / L} = \frac{2L}{\tau\delta(1-\delta)^2}$$

Thus for a given energy throughput, some energy is provided from the supply to the load when providing the inductor energy, hence the discontinuous inductor current threshold occurs at the same load level for each boost converter, including the basic converter in figure 17.6.

ii. Since the boost circuits have the same ac equivalent circuit, the inductor and capacitor, currents and voltages would be expected to be basically the same for each circuit, as shown by the waveforms in example 17.3. Consequently, the switch and diode voltages and currents are also the same for each boost converter.

The two principal differences are the supply current and the capacitor voltage rating. The capacitor voltage rating for the alternative boost converter is lower, $v_o - E_i$, as opposed to v_o for the conventional converter. The supply current for the alternative converter is discontinuous (although always non-zero), as shown in the following waveforms. This will negate the desirable continuous current feature exploited in boost converters that are controlled so as to produce sinusoidal input current or draw continuous input power.

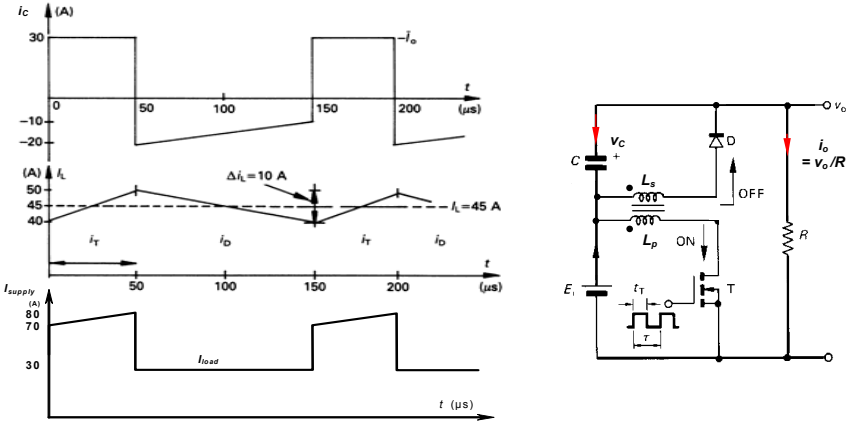


Figure: Example 17.3 – *waveforms and transformer coupled version.*

An isolated version, with the input supply isolated from the load, is not possible. But the couple inductor version shown in the example figure can be useful in avoiding very short (or long) switch duty cycles and help control (both avoiding or ensuring) continuous inductor current conduction conditions.

♣

17.4 The buck-boost converter

The basic *buck-boost flyback converter* circuit is shown in figure 17.7a. When transistor T is on, energy is transferred to the inductor and the load current is provided solely from the output capacitor. When the transistor turns off, inductor current is forced through the diode. Energy stored in *L* is transferred to *C* and the load *R*. This transfer action results in an output voltage of opposite polarity to that of the input. Neither the input nor the output current is continuous, although the inductor current may be continuous or discontinuous.

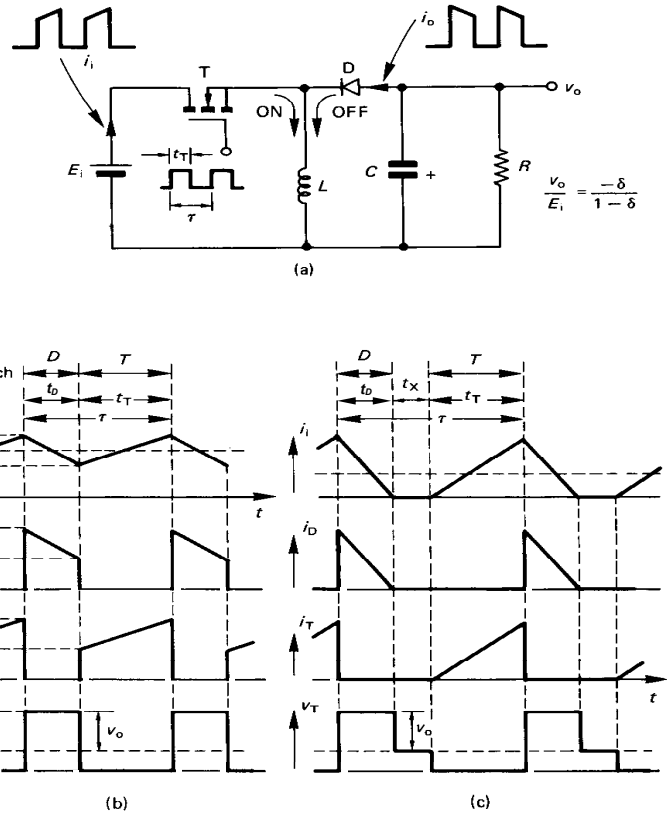


Figure 17.7. *Non-isolated, step up/down flyback converter (buck-boost converter) where $v_o \leq 0$: (a) circuit diagram; (b) waveforms for continuous inductor current; and (c) discontinuous inductor current.*

17.4.1 Continuous choke (inductor) current

Various circuit voltage and current waveforms for the buck-boost flyback converter operating in a continuous inductor conduction mode are shown in figure 17.7b.

Assuming a constant input and output voltage, from $v = L di/dt$, the change in inductor current is given by

$$\Delta i_L = \frac{E_i}{L} t_T = \frac{-v_o}{L}(\tau - t_T) \qquad (17.77)$$

Thus assuming $P_{in} = P_{out}$, that is $v_o \bar{I}_o = E_i \bar{I}_i$

$$\frac{v_o}{E_i} = \frac{\bar{I}_i}{\bar{I}_o} = -\frac{\delta}{1-\delta} \tag{17.78}$$

where δ = t_T /τ. For $\delta < ½$ the output magnitude is less than the input voltage magnitude, while for $\delta > ½$ the output voltage is greater in magnitude (but as for $\delta < ½$, opposite in polarity) than the input voltage.
The maximum and minimum inductor current is given by

$$\hat{i}_L = \frac{\bar{I}_o}{1-\delta} + ½\frac{v_o}{L}(1-\delta)\tau = v_o\left[\frac{1}{(1-\delta)R} + \frac{(1-\delta)\tau}{2L}\right] \tag{17.79}$$

and

$$\check{i}_L = \frac{\bar{I}_o}{1-\delta} - ½\frac{v_o}{L}(1-\delta)\tau = v_o\left[\frac{1}{(1-\delta)R} - \frac{(1-\delta)\tau}{2L}\right] \tag{17.80}$$

The inductor rms ripple current is given by

$$i_{L\tau} = \frac{\Delta i_L}{2\sqrt{3}} = \frac{1}{2\sqrt{3}}\frac{v_o}{L}(1-\delta)\delta\tau \tag{17.81}$$

while the inductor total rms current is

$$i_{L\text{rms}} = \sqrt{\bar{I}_L^2 + i_{L\tau}^2} = \sqrt{\bar{I}_L^2 + \left(\frac{½\Delta i_L}{\sqrt{3}}\right)^2} = \sqrt{\frac{1}{3}\left(\hat{i}_L^2 + \hat{i}_L \times \check{i}_L + \check{i}_L^2\right)} \tag{17.82}$$

The switch and diode average and rms currents are given by

$$\begin{aligned} \bar{I}_T &= \bar{I}_i = \delta \bar{I}_L & I_{T\text{rms}} &= \sqrt{\delta}\, i_{L\text{rms}} \\ \bar{I}_D &= (1-\delta)\bar{I}_L = \bar{I}_o & I_{D\text{rms}} &= \sqrt{1-\delta}\, i_{L\text{rms}} \end{aligned} \tag{17.83}$$

Switch utilisation ratio
The switch utilisation ratio, SUR, is a measure of how fully a switching device's power handling capabilities are utilised in any switching application. The ratio is defined as

$$SUR = P_{out} \Big/ p\hat{V}_T \hat{I}_T \tag{17.84}$$

where p is the number of power switches in the circuit; $p=1$ for the buck-boost converter. The switch maximum instantaneous voltage and current are $\hat{V}_T$ and $\hat{I}_T$ respectively. As shown in figure 17.7b, the maximum switch voltage supported in the off-state is $E_i + v_o$, while the maximum current is the maximum inductor current $\hat{i}_L$ which is given by equation (17.79). If the inductance L is large such that the ripple current is small, the peak inductor current is approximated by the average inductor current which yields $\hat{I}_T \approx \bar{I}_L = \bar{I}_o / 1-\delta$, that is

$$SUR = \frac{v_o \bar{I}_o}{(E_i + v_o) \times \bar{I}_o / 1-\delta} = \delta(1-\delta) \tag{17.85}$$

which assumes continuous inductor current. This result shows that the closer the output voltage v_o is in magnitude to the input voltage E_i, that is $\delta = ½$, the better the switch *I-V* ratings are utilised.

17.4.2 Discontinuous capacitor charging current in the switch off-state

It is possible that the inductor current falls below the output (resistor) current during a part of the cycle when the switch is off and the inductor is transferring (replenishing) energy to the output circuit. Under such conditions, towards the end of the off period, some of the load current requirement is provided by the capacitor even though this is the period during which its charge is replenished by inductor energy. The circuit independent transfer function in equation (17.78) remains valid. This discontinuous capacitor charging condition occurs when the minimum inductor current and the output current are equal. That is

$$\begin{aligned} &\check{I}_L - \bar{I}_o \le 0 \\ &\bar{I}_L - ½\Delta i_L - \bar{I}_o \le 0 \\ &\frac{\bar{I}_o}{1-\delta} - ½\frac{\bar{I}_o R}{L}(1-\delta)\tau - \bar{I}_o \le 0 \end{aligned} \tag{17.86}$$

which yields

$$\delta \le 1 + \frac{L}{\tau R} - \sqrt{\left(1+\frac{L}{\tau R}\right)^2 - 1} \tag{17.87}$$

17.4.3 Discontinuous choke current

The onset of discontinuous inductor operation occurs when the minimum inductor current $\check{i}_L$, reaches zero. That is, with $\check{i}_L = 0$ in equation (17.80), the last equality

$$\frac{1}{(1-\delta)R} - \frac{(1-\delta)\tau}{2L} = 0 \tag{17.88}$$

relates circuit component values (R and L) and operating conditions (f and δ) at the verge of discontinuous inductor current.
The change from continuous to discontinuous inductor current conduction occurs when

$$\bar{I}_L = ½\hat{i}_L = ½\Delta i_L \tag{17.89}$$

where from equation (17.77) $\hat{i}_L = v_o(\tau - t_T)/L$

The circuit waveforms for discontinuous conduction are shown in figure 17.7c. The output voltage for discontinuous conduction is evaluated from

$$\hat{i}_L = \frac{E_i}{L}t = -\frac{v_o}{L}(\tau - t_T - t_x) \tag{17.90}$$

which yields

$$\frac{v_o}{E_i} = -\frac{\delta}{1-\delta-\frac{t_x}{\tau}} \tag{17.91}$$

Alternatively, using equation (17.90) and

$$\bar{I}_i = ½\delta\hat{i}_L \tag{17.92}$$

yields

$$\bar{I}_i = \frac{E_i \tau \delta^2}{2L} \tag{17.93}$$

The inductor current is neither the input current nor the output current, but is comprised of separate components (in time) of each of these currents. Examination of figure 17.7b, reveals that these currents are a proportion of the inductor current dependant on the duty cycle, and that on the verge of discontinuous conduction:

$\bar{I}_i = ½\delta\hat{i}_L$ and $\bar{I}_o = ½\delta_{off}\hat{i}_L = ½(1-\delta)\hat{i}_L$ where $\hat{i}_L = \Delta i_L$

Thus using power in equals power out, that is $E_i\bar{I}_i = v_o\bar{I}_o$, equation (17.93) becomes

$$\frac{v_o}{E_i} = \frac{E_i\tau\delta^2}{2L\bar{I}_o} = \frac{v_o\tau\delta^2}{2L\bar{I}_i} = \delta\sqrt{\frac{\tau R}{2L}} \tag{17.94}$$

On the verge of discontinuous conduction, these equations can be rearranged to give

$$\bar{I}_o = \frac{E_i}{2L}\tau\delta(1-\delta) = \frac{v_o}{2L}\tau(1-\delta)^2 \tag{17.95}$$

At a low output current or low input voltage there is a likelihood of discontinuous conduction. To avoid this condition, a larger inductance value is needed, which worsen the transient response. Alternatively, with extremely low on-state duty cycles, a voltage-matching transformer can be used to increase δ. Once a transformer is employed, any smps technique can be used to achieve the desired output voltage. Figures 17.7b and c show that both the input and output current are always discontinuous.

17.4.4 Load conditions for discontinuous inductor current

As the load current decreases, the inductor average current also decreases, but the inductor ripple current magnitude is unchanged. If the load resistance is increased sufficiently, the bottom of the triangular inductor current, $\check{i}_L$, eventually reduces to zero. Any further increase in load resistance causes discontinuous inductor current and the voltage transfer function given by equation (17.78) is no longer valid and equations (17.90) and (17.94) are applicable. The critical load resistance for continuous inductor current is specified by

$$R_{crit} \le \frac{v_o}{\bar{I}_o} \tag{17.96}$$

Substituting for, the average input current in terms of $\hat{i}_L$ and v_o in terms of Δi_L from equation (17.77), yields

$$R_{crit} \le \frac{v_o}{\overline{I}_o} = \frac{2L}{\tau(1-\delta)^2} \qquad (17.97)$$

By substituting the switching frequency ($f_s = 1/\tau$) or the fundamental inductor reactance ($X_L = 2\pi f_s L$) the following critical resistance forms result.

$$R_{crit} \le \frac{v_o}{\overline{I}_o} = \frac{2L}{\tau(1-\delta)^2} = \frac{v_o}{E_i} \times \frac{2L}{\tau\delta(1-\delta)} = \frac{2f_s L}{(1-\delta)^2} = \frac{X_L}{\pi(1-\delta)^2} \quad (\Omega) \qquad (17.98)$$

Equation (17.98) is equation (17.88), re-arranged.

If the load resistance increases beyond R_{crit}, the output voltage can no longer be maintained with duty cycle control according to the voltage transfer function in equation (17.78).

17.4.5 Control methods for discontinuous inductor current

Once the load current has reduced to the critical level as specified by equation (17.98), the input energy is in excess of the load requirement. Open loop load voltage regulation control is lost and the capacitor *C* tends to overcharge.

Hardware approaches can solve this problem – by ensuring continuous inductor current

- increase *L* thereby decreasing the inductor current ripple p-p magnitude
- step-down transformer impedance matching to effectively reduce the apparent load impedance

Two control approaches to maintain output voltage regulation when $R > R_{crit}$ are

- vary the switching frequency f_s, maintaining the switch on-time t_T constant so that Δi_L is fixed or
- reduce the switch on-time t_T, but maintain a constant switching frequency f_s, thereby reducing Δi_L.

If a fixed switching frequency is desired for all modes of operation, then reduced on-time control, using output voltage feedback, is preferred. If a fixed on-time mode of control is used, then the output voltage is control by inversely varying the frequency with output voltage. Alternatively, output voltage feedback can be used.

17.4.5i - fixed on-time t_T, variable switching frequency f_{var}

The operating frequency f_{var} is varied while the switch-on time t_T is maintained constant such that the ripple current remains unchanged. Operation is specified by equating the input energy and the output energy, thus maintaining a constant capacitor charge, hence output voltage. That is, equating energies

$$½\Delta i_L E_i t_T = \frac{v_o^2}{R}\frac{1}{f_{\text{var}}} \qquad (17.99)$$

Isolating the variable switching frequency f_{var} gives

$$f_{\text{var}} = \frac{v_o^2}{½\Delta i_L E_i t_T}\frac{1}{R} = f_s R_{crit} \times \frac{1}{R}$$

$$f_{\text{var}} \quad \alpha \quad \frac{1}{R} \qquad (17.100)$$

Load resistance *R* is not a directly or readily measurable parameter for feedback proposes. Alternatively, since $v_o = \overline{I}_o R$, substitution for *R* in equation (17.100) gives

$$f_{\text{var}} = f_s \frac{R_{crit}}{v_o} \times \overline{I}_o$$

$$f_{\text{var}} \quad \alpha \quad \overline{I}_o \qquad (17.101)$$

That is, for discontinuous inductor current, namely $\overline{I}_L < ½\Delta i_L$ or $\overline{I}_o < v_o / R_{crit}$, if the switch on-state period t_T remains constant and f_{var} is either varied proportionally with load current or varied inversely with load resistance, then the required output voltage v_o will be maintained.

17.4.5ii - fixed switching frequency f_s, variable on-time $t_{T\text{var}}$

The operating frequency f_s remains fixed while the switch-on time $t_{T\text{var}}$ is reduced such that the ripple current can be reduced. Operation is specified by equating the input energy and the output energy as in equation (17.99), thus maintaining a constant capacitor charge, hence voltage. That is

$$½\Delta i_L E_i t_{T\,\text{var}} = \frac{v_o^2}{R}\frac{1}{f_s} \qquad (17.102)$$

Isolating the variable on-time $t_{T\text{var}}$ gives

$$t_{T\,\text{var}} = \frac{v_o^2}{½\Delta i_L E_i f_s}\frac{1}{R}$$

Substituting Δi_L from equation (17.77) gives

$$t_{T\,\text{var}} = t_T\sqrt{R_{crit}} \times \frac{1}{\sqrt{R}}$$

$$t_{T\,\text{var}} \quad \alpha \quad \frac{1}{\sqrt{R}} \qquad (17.103)$$

Again, load resistance *R* is not a directly or readily measurable parameter for feedback proposes and substitution of $v_o / \overline{I}_o$ for *R* in equation (17.74) gives

$$t_{T\,\text{var}} = t_T\sqrt{\frac{R_{crit}}{v_o}} \times \sqrt{\overline{I}_o}$$

$$t_{T\,\text{var}} \quad \alpha \quad \sqrt{\overline{I}_o} \qquad (17.104)$$

That is, if the switching frequency f_s is fixed and switch on-time t_T is reduced proportionally to $\sqrt{\overline{I}_o}$ or inversely to $\sqrt{R}$, when discontinuous inductor current commences, namely $\overline{I}_L < ½\Delta i_L$ or $\overline{I}_o < v_o / R_{crit}$, then the required output voltage magnitude v_o will be maintained.

Alternatively the output voltage is related to the duty cycle by $v_o = -\delta E_i\sqrt{R\tau/2L}$. See table 17.2.

17.4.6 Output ripple voltage

The output ripple voltage is the capacitor ripple voltage. Ripple voltage for a capacitor is defined as

$$\Delta v_o = \tfrac{1}{C}\int i\,dt$$

Figure 17.7 shows that the constant output current $\overline{I}_o$ is provided solely from the capacitor during the on period t_T when the switch conducting, thus

$$\Delta v_o = \tfrac{1}{C}\int i\,dt = \tfrac{1}{C} t_T \overline{I}_o$$

Substituting for $\overline{I}_o = v_o / R$ gives

$$\Delta v_o = \tfrac{1}{C}\int i\,dt = \tfrac{1}{C} t_T \overline{I}_o = \tfrac{1}{C} t_T \,{v_o}\!\!\diagup\!\!_R$$

Rearranging gives the percentage peak-to-peak voltage ripple in the output voltage

$$\frac{\Delta v_o}{v_o} = \frac{1}{RC} t_T = \frac{\delta\tau}{RC} \qquad (17.105)$$

The capacitor equivalent series resistance and inductance can be account for, as with the forward converter, 17.1.5. When the switch conducts, the output current is constant and is provided solely from the capacitor. Thus no ESL voltage effects result during this constant capacitor current portion of the switching cycle.

17.4.7 Buck-boost, flyback converter design procedure

The output voltage of the buck-boost converter can be regulated by operating at a fixed frequency and varying the transistor on-time t_T. However, the output voltage diminishes while the transistor is on and increases when the transistor is off. This characteristic makes the converter difficult to control on a fixed frequency basis.

A simple approach to control the flyback regulator in the discontinuous mode is to fix the peak inductor current, which specifies a fixed diode conduction time, t_D. Frequency then varies directly with output current and transistor on-time varies inversely with input voltage.

With discontinuous inductor conduction, the worst-case condition exists when the input voltage is low while the output current is at a maximum. Then the frequency is a maximum and the dead time t_x is zero because the transistor turns on as soon as the diode stops conducting.

Given			Worst case	
$E_{i(\min)}$	$\overline{I}_{o(\max)}$		$E_i = E_{i(\min)}$	$t_x = 0$
V_o	$f_{(\max)}$	Δe_o	$\overline{I}_o = \overline{I}_{o(\max)}$	

Assuming a fixed peak inductor current $\hat{i}_i$ and output voltage v_o, the following equations are valid

$$E_{i(\min)} t_T = v_o t_D = \hat{i}_i \times L \qquad (17.106)$$

$$\tau_{(\min)} = 1/f_{(\max)} \qquad (17.107)$$

Equation (17.106) yields

$$t_D = \frac{1}{f_{(\max)}\left(\frac{v_o}{E_{i(\min)}} + 1\right)} \qquad (17.108)$$

Where the diode conduction time t_D is constant since in equation (17.106), v_o, $\hat{i}_i$, and L are all constants. The average output capacitor current is given by

$$\overline{I}_o = \tfrac{1}{2}\hat{i}_i(1-\delta)$$

and substituting equation (17.108) yields

$$\overline{I}_{o(\max)} = \tfrac{1}{2}\hat{i}_i \times f_{(\max)} \times \frac{1}{f_{(\max)}\left(\frac{v_o}{E_{i(\min)}} + 1\right)}$$

therefore

$$\hat{i}_i = 2 \times \overline{I}_{o(\max)} \times \left(\frac{v_o}{E_{i(\min)}} + 1\right)$$

and upon substitution into equation (17.106)

$$L = \frac{t_D v_o}{2\,\overline{I}_{o(\max)}\left(\frac{v_o}{E_{i(\min)}} + 1\right)} \qquad (17.109)$$

The minimum capacitance is specified by the maximum allowable ripple voltage, that is

$$\breve{C} = \frac{\Delta Q}{\Delta e_o} = \frac{\hat{i}_i\, t_D}{2\Delta e_o}$$

that is

$$\breve{C} = \frac{\overline{I}_{o(\max)} t_D}{\Delta e_o\left(\frac{v_o}{E_{i(\min)}} + 1\right)} \qquad (17.110)$$

For large output capacitance, the ripple voltage is dropped across the capacitor equivalent series resistance, which is given by

$$ESR_{(\max)} = \frac{\Delta e_o}{\hat{i}_i} \qquad (17.111)$$

The frequency varies as a function of load current. Equation (17.107) gives

$$\frac{\overline{I}_o}{f} = \tfrac{1}{2}\hat{i}_i\, t_T = \frac{\overline{I}_{o(\max)}}{f_{(\max)}}$$

therefore

$$f = f_{(\max)} \times \frac{\overline{I}_o}{\overline{I}_{o(\max)}} \qquad (17.112)$$

and

$$f_{(\min)} = f_{(\max)} \times \frac{\overline{I}_{o(\min)}}{\overline{I}_{o(\max)}} \qquad (17.113)$$

Example 17.5: ***Buck-boost flyback converter***

The 10kHz flyback converter in figure 17.7 is to operate from a 50V input and produces an inverted non-isolated 75V output. The inductor is 300μH and the resistive load is 2.5Ω.

i. Calculate the duty cycle, hence transistor off-time, assuming continuous inductor current.
ii. Calculate the mean input and output current.
iii. Draw the inductor current, showing the minimum and maximum values.
iv. Calculate the capacitor rms ripple current and output p-p ripple voltage if C = 10,000μF.
v. Determine
- the critical load resistance.
- the minimum inductance for continuous inductor conduction with 2.5 Ω load

vi. At what load resistance does the instantaneous inductor current fall below the output current?
vii. What is the output voltage if the load resistance is increased to four times the critical resistance?

Solution

i. From equation (17.91), which assumes continuous inductor current

$$\frac{v_o}{E_i} = -\frac{\delta}{1-\delta} \quad \text{where} \quad \delta = t_T/\tau$$

that is

$$\frac{75\text{V}}{50\text{V}} = \frac{\delta}{1-\delta} \quad \text{thus} \quad \delta = \tfrac{3}{5}$$

That is, τ = 1/ f_s = 100 μs with a 60μs switch on-time.

ii. The mean output current $\overline{I}_o$ is given by

$$\overline{I}_o = v_o / R = 75\text{V}/2.5\Omega = 30\text{A}$$

From power transfer considerations

$$\overline{I}_i = v_o \overline{I}_o / E_i = 75\text{V} \times 30\text{A}/50\text{V} = 45\text{A}$$

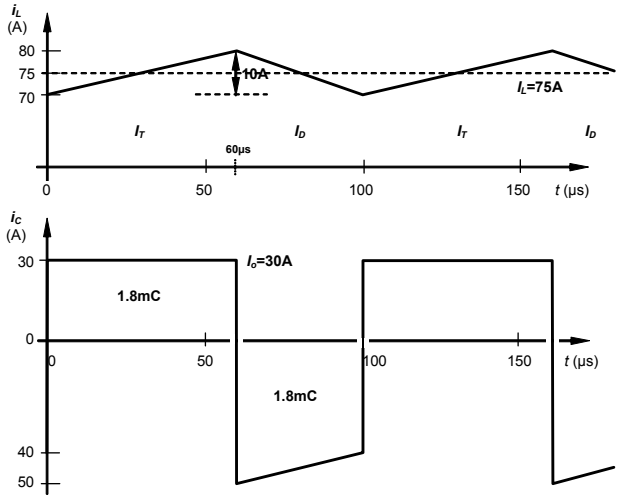


Figure: Example 17.4

iii. The average inductor current can be derived from

$$\overline{I}_i = \delta \overline{I}_L \quad \text{or} \quad \overline{I}_o = (1-\delta)\overline{I}_L$$

That is

$$\overline{I}_L = \overline{I}_i / \delta = \overline{I}_o / (1-\delta)$$
$$= 45\text{A}/\tfrac{3}{5} = 30\text{A}/\tfrac{2}{5} = 75\text{A}$$

From $v = L\, di/dt$, the ripple current $\Delta i_L = E_i t_T / L$ = 50V x 60μs /300 μH = 10 A, that is

$$\hat{i}_L = \overline{I}_L + \tfrac{1}{2}\Delta i_L = 75\text{A} + \tfrac{1}{2}\times 10\text{A} = 80\text{A}$$

$$\breve{i}_L = \overline{I}_L - \tfrac{1}{2}\Delta i_L = 75\text{A} - \tfrac{1}{2}\times 10\text{A} = 70\text{A}$$

Since $\breve{i}_L$ = 70A ≥ 0A, rhe inductor current is continuous, thus the analysis in parts *i*, *ii*, and *iii*, is valid.

iv. The capacitor current is derived by using Kirchhoff's current law such that at any instant in time, the diode current, plus the capacitor current, plus the 30A constant load current into R, all sum to zero.

$$i_{C\text{rms}} = \sqrt{\frac{1}{\tau}\left[\int_0^{t_T} \overline{I}_o^2\, dt + \int_0^{\tau - t_T} \left(\frac{\Delta i_L}{\tau - t_T} t - \hat{i}_L + \overline{I}_o\right)^2 dt\right]}$$

$$= \sqrt{\frac{1}{100\mu\text{s}}\left[\int_0^{60\,\mu\text{s}} 30\text{A}^2\, dt + \int_0^{40\,\mu\text{s}} \left(\frac{10\text{A}}{40\mu\text{s}} t - 50\text{A}\right)^2 dt\right]} = 36.8\text{A}$$

The output ripple voltage is given by equation (17.105), that is

$$\frac{\Delta v_o}{v_o} = \frac{\delta\tau}{CR} = \frac{⅗\times 100\mu s}{10{,}000\mu F\times 2½\Omega} \equiv 0.24\%$$

The output ripple voltage is therefore

$$\Delta v_o = 0.24\times 10^{-2}\times 75V = 180mV$$

v. The critical load resistance, R_{crit}, produces an inductor current with Δi_L = 10 A ripple. From equation (17.98)

$$R_{crit} = \frac{2L}{\tau(1-\delta)^2} = \frac{2\times 300\mu H}{100\mu s\times(1-⅗)^2} = 37½\Omega$$

The minimum inductance for continuous inductor current operation, with a 2½Ω load, can be found by rearranging the critical resistance formula, as follows:

$$L_{crit} = ½R\tau(1-\delta)^2 = ½\times 2.5\Omega\times 100\mu s\times(1-⅗)^2 = 20\mu H$$

vi. The ± 5A inductor ripple current is independent of the load, provided the critical resistance of 37½Ω is not exceeded. When the average inductor current is less than 5A more than the output current, the capacitor must provide load current not only when the switch is on but also for a portion of the time when the switch is off. The transition is given by equation (17.87), that is

$$\delta \le 1+\frac{L}{\tau R}-\sqrt{\left(1+\frac{L}{\tau R}\right)^2-1}$$

Alternately, when

$$\bar{I}_L - \bar{I}_o = 5A$$

$$\frac{\bar{I}_o}{1-\delta} - \bar{I}_o = 5A$$

For δ = ⅗, $\bar{I}_o$ = 3⅓A. whence

$$R = \frac{v_o}{\bar{I}_o} = \frac{75V}{11⅔\,A} = 22½\Omega$$

The average inductor current is 8⅓A, with a minimum value of 3⅓A, the same as the load current. That is, for R < 22½Ω all the load requirement is provided from the inductor when the switch is off, with excess energy charging the output capacitor. For R > 22½Ω insufficient energy is available from the inductor to provide the load energy throughout the whole of the period when the switch is off. The capacitor supplements the load requirement towards the end of the off period. When R > 37½Ω (the critical resistance), discontinuous inductor current occurs, and the purely duty cycle dependent transfer function (circuit parameter independent) is no longer valid.

vii. When the load resistance is increased to 150Ω, four times the critical resistance, the output voltage is given by equation (17.94):

$$v_o = E_i\,\delta\sqrt{\frac{\tau R}{2L}} = 50V\times ⅗\times\sqrt{\frac{100\mu s\times 150\Omega}{2\times 300\mu H}} = 150V$$

♣

17.5 Flyback converters – a conceptual assessment

In section 17.2, the boost and buck-boost converters were both introduced as flyback or ringing choke converters. This is not the traditional approach adopted to the classification of these two converters. This text has classified both as flyback converters since they are in fact the same converter. A converter is considered a two port network – an input E_i and an output v_o – that are related by a transfer function which is expressed in terms of the switch on-state duty cycle δ.

$$\frac{V_o}{E_i} = f(\delta)$$

A second output v_1 exists between the input E_i and the output v_o, as shown in figure 17.8. By Kirchhoff's voltage law, this auxiliary output is

$$v_1 = E_i - v_o$$

$$\frac{V_1}{E_i} = 1 - \frac{V_o}{E_i} = 1 - f(\delta)$$

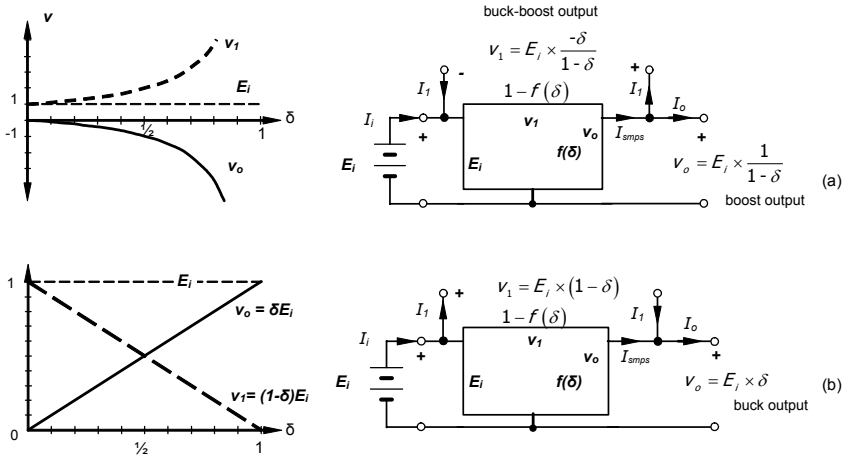


Figure 17.8. *Basic converters shown as a three-port block diagrams for: (a) the flyback converter and (b) the forward converter.*

The flyback converter – figure 17.8a

If $f(\delta)$ represents a boost converter, with a voltage transfer function $1/1-\delta$, then $1 - f(\delta) = -\delta/1-\delta$, which is the buck-boost converter transfer function. The converse is also true. Thus if a boost converter output exists, a buck-boost output is inherently available, independent of the connection position of the output capacitor C_o. In terms of dc circuit theory, the output capacitor can be connected across v_o (as in figure 17.2), v_1 (as in example 17.3), or apportioned between both outputs. The circuit permutations in figure 17.9 show how the boost converter, using ac and dc circuit theory, can be systematically translated to the buck-boost converter, and vice versa. The schematic of an auto-transformer (variac) is interposed since it too can provide the equivalent two ac output possibilities. Whether a dc converter or an ac variac, power can be drawn from either output separately or from both outputs simultaneously. The output ports of both converters, when an extra switch and diode are added, are bidirectional reversible as considered in section 17.7.2.

The forward converter – figure 17.8b

Just as the boost and buck-boost outputs are complementary, the buck converter has a complementary output possibility. If the output v_o is defined by the buck converter transfer function δ then the supplementary output v_1 is defined by 1-δ. The output 1-δ cannot exist independently of the output δ. In order to maintain output voltage transfer function integrity according to the duty cycle dependant transfer functions, the current sourced from port v_o (the buck output) must exceed the current sunk by port v_1.
That is, if the outputs v_o and v_1 are resistively loaded, in figure 17.8a

$$I_{smps} > 0$$

$$I_o \ge I_1$$

$$\frac{V_o}{R_o} \ge \frac{V_1}{R_1}$$

or

$$R_1 \ge R_o\frac{1-\delta}{\delta}$$

Notice in figure 17.8a, in the flyback converter case, I_{smps} is always positive. Therefore no load resistance restrictions exist for the two outputs, save the inductor current is continuous.

Thus only two fundamental single-switch, single-inductor converters (flyback and forward) exist, each offering two output voltage transfer function possibilities. One of the four output possibilities, $1-\delta$, cannot uniquely exist.

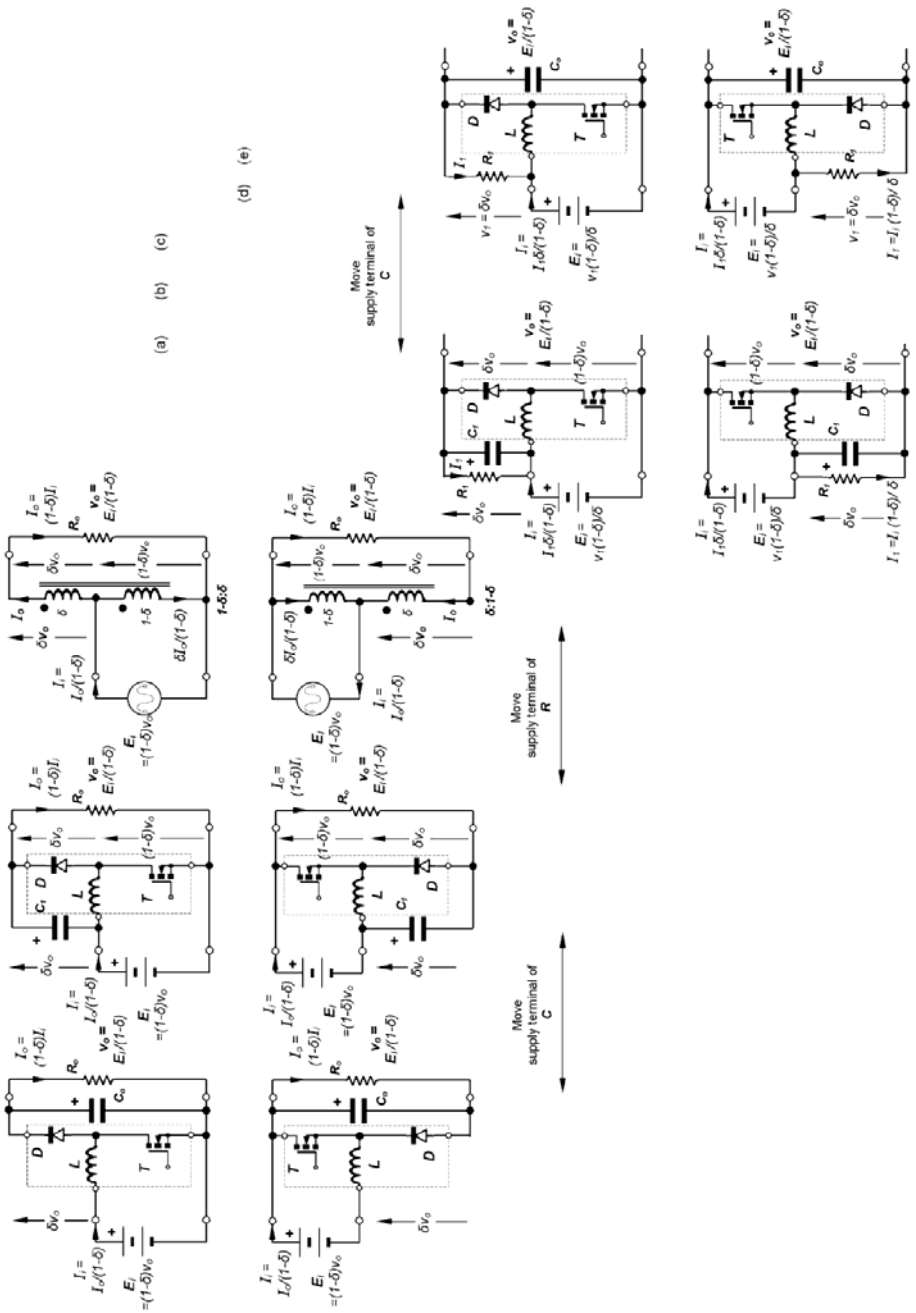


Figure 17.9. *Basic: (a) boost to (e) buck-boost converter systematic translations.*

17.6 The output reversible converter

The basic *reversible converter,* sometimes called an *asymmetrical half bridge converter* (see chapter 14.5), shown in figure 17.10a allows two-quadrant output voltage operation. Operation is characterised by both switches operating simultaneously, being either both on or both off.

The input voltage E_i is chopped by switches T_1 and T_2, and because the input voltage is greater than the load voltage v_o, energy is transferred from the dc supply E_i to L, C, and the load R. When the switches are turned off, energy stored in L is transferred via the diodes D_1 and D_2 to C and the load R but in a path involving energy being returned to the supply, E_i. This connection feature allows energy to be transferred from the load back into E_i when used with an appropriate load and the correct duty cycle.

Parts b and c respectively of figure 17.10 illustrate reversible converter circuit current and voltage waveforms for continuous and discontinuous conduction of L, in a forward converter mode, when $\delta > ½$.

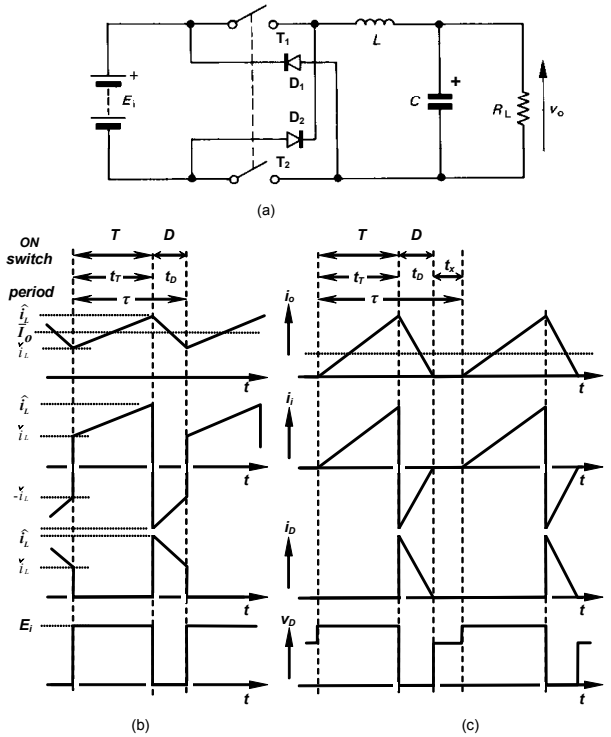


Figure 17.10. *Basic reversible converter with δ>½: (a) circuit diagram; (b) waveforms for continuous inductor current; and (c) discontinuous inductor current.*

For analysis it is assumed that components are lossless and the output voltage v_o is maintained constant because of the large capacitance magnitude of the capacitor C across the output. The input voltage E_i is also assumed constant, such that $E_i \geq v_o > 0$, as shown in figure 17.10a.

17.6.1 Continuous inductor current

When the switches are turned on for period t_T, the difference between the supply voltage E_i and the output voltage v_0 is impressed across L. From $V = Ldi/dt$, the rising current change through the inductor will be

$$\Delta i_L = \hat{i}_L - \check{i}_L = \frac{E_i - v_o}{L} \times t_T \qquad (17.114)$$

When the two switches are turned off for the remainder of the switching period, $\tau - t_T$, the two freewheel diodes conduct in series and $E_i + v_o$ is impressed across L. Thus, assuming continuous inductor conduction the inductor current fall is given by

$$\Delta i_L = \frac{E_i + v_o}{L} \times (\tau - t_T) \tag{17.115}$$

Equating equations (17.114) and (17.115) yields

$$\frac{v_o}{E_i} = \frac{\bar{I}_i}{\bar{I}_o} = \frac{2t_T - \tau}{\tau} = 2\delta - 1 \qquad 0 \le \delta \le 1 \tag{17.116}$$

The voltage transfer function is independent of circuit inductance L and capacitance C.
Equation (17.116) shows that for a given input voltage, the output voltage is determined by the transistor conduction duty cycle δ and the output voltage $|v_o|$ is always less than the input voltage. This confirms and validates the original analysis assumption that $E_i \ge |v_o|$. The linear transfer function varies between -1 and 1 for $0 \le \delta \le 1$, that is, the output can be varied between $v_o = -E_i$, and $v_o = E_i$. The significance of the change in transfer function polarity at $\delta = ½$ is that

- for $\delta > ½$ the converter acts as a forward converter, but
- for $\delta < ½$, if the output is a negative source, the converter acts as a boost converter with energy transferred to the supply E_i, from the negative output source.

Thus the transfer function can be expressed as follows

$$\frac{v_o}{E_i} = \frac{\bar{I}_i}{\bar{I}_o} = 2\delta - 1 = 2\,(\delta - ½) \qquad ½ \le \delta \le 1 \tag{17.117}$$

and

$$\frac{E_i}{v_o} = \frac{\bar{I}_o}{\bar{I}_i} = \frac{1}{2\delta - 1} = \frac{1}{2\,(\delta - ½)} \qquad 0 \le \delta \le ½ \tag{17.118}$$

where equation (17.118) is in the boost converter transfer function form.

17.6.2 Discontinuous inductor current

In the forward converter mode, $\delta \ge ½$, the onset of discontinuous inductor current operation occurs when the minimum inductor current $\check{i}_L$, reaches zero. That is,

$$\bar{I}_L = ½\Delta i_L = \bar{I}_o \tag{17.119}$$

If the transistor on-time t_T is reduced or the load resistance increases, the discontinuous condition dead time t_x appears as indicated in figure 17.10c. From equations (17.114) and (17.115), with $\check{i}_L = 0$, the following output voltage transfer function can be derived

$$\Delta i_L = \hat{i}_L - 0 = \frac{E_i - v_o}{L} \times t_T = \frac{E_i + v_o}{L} \times (\tau - t_T - t_x) \tag{17.120}$$

which after rearranging yields

$$\frac{v_o}{E_i} = \frac{2\delta - 1 - \frac{t_x}{\tau}}{1 - \frac{t_x}{\tau}} \qquad 0 \le \delta < 1 \tag{17.121}$$

17.6.3 Load conditions for discontinuous inductor current

In the forward converter mode, $\delta \ge ½$, as the load current decreases, the inductor average current also decreases, but the inductor ripple current magnitude is unchanged. If the load resistance is increased sufficiently, the trough of the triangular inductor current, $\check{i}_L$, eventual reduces to zero. Any further increase in load resistance causes discontinuous inductor current and the linear voltage transfer function given by equation (17.116) is no longer valid. Equation (17.121) is applicable. The critical load resistance for continuous inductor current is specified by

$$R_{crit} \le \frac{v_o}{\bar{I}_o} \tag{17.122}$$

Substituting $\bar{I}_o = \bar{I}_L$ and using equations (17.114) and (17.119), yields

$$R_{crit} \le \frac{v_o}{\bar{I}_o} = \frac{v_o}{½\Delta i_L} = \frac{2v_o L}{(E_i - v_o)t_T} \tag{17.123}$$

Dividing throughout by E_i and substituting $\delta = t_T / \tau$ yields

$$R_{crit} \le \frac{v_o}{\bar{I}_o} = \frac{(2\delta - 1)L}{(1 - \delta)\delta\tau} \tag{17.124}$$

By substituting the switching frequency ($f_s = 1/\tau$) or the fundamental inductor reactance ($X_L = 2\pi f_s L$), critical resistance can be expressed in the following forms.

$$R_{crit} \le \frac{v_o}{\bar{I}_o} = \frac{2(\delta - ½)L}{(1 - \delta)\delta\tau} = \frac{2(\delta - ½)f_s L}{(1 - \delta)\delta} = \frac{(\delta - ½)X_L}{\pi(1 - \delta)\delta} \qquad (\Omega) \tag{17.125}$$

If the load resistance increases beyond R_{crit}, the output voltage can no longer be maintained with duty cycle control according to the voltage transfer function in equation (17.116).

17.6.4 Control methods for discontinuous inductor current

Once the load current has reduced to the critical level as specified by equation (17.120) the input energy is in excess of the load requirement. Open loop load voltage regulation control is lost and the capacitor C tends to overcharge.
As with the other converters considered, hardware and control approaches can mitigate this overcharging problem. The specific control solutions for the forward converter in section 17.3.4, are applicable to the reversible converter. The two time domain control approaches offer the following operational modes.

17.6.4i - fixed on-time t_T, variable switching frequency f_{var}

The operating frequency f_{var} is varied while the switch-on time t_T is maintained constant such that the magnitude of the ripple current remains unchanged. Operation is specified by equating the input energy and the output energy, thus maintaining a constant capacitor charge, hence output voltage. That is, equating energies

$$½\Delta i_L E_i t_T = \frac{v_o^2}{R}\frac{1}{f_{var}} \tag{17.126}$$

Isolating the variable switching frequency f_{var} and using $v_o = \bar{I}_o R$ to eliminate R yields

$$f_{var} = f_s R_{crit} \times \frac{1}{R} = f_s \frac{R_{crit}}{v_o} \times \bar{I}_o$$
$$f_{var} \;\alpha\; \frac{1}{R} \text{ or } f_{var} \;\alpha\; \bar{I}_o \tag{17.127}$$

That is, once discontinuous inductor current occurs at $\bar{I}_o < ½\Delta i_L$ or $\bar{I}_o < v_o / R_{crit}$, a constant output voltage v_o can be maintained if the switch on-state period t_T remains constant and the switching frequency is varied

- proportionally with load current, $\bar{I}_o$
- inversely with the load resistance, R_{crit}
- inversely with the output voltage, v_o.

17.6.4ii - fixed switching frequency f_s, variable on-time t_{Tvar}

The operating frequency f_s remains fixed while the switch-on time t_{Tvar} is reduced, resulting in the ripple current magnitude being reduced. Equating input energy and output energy as in equation (17.27), thus maintaining a constant capacitor charge, hence voltage, gives

$$½\Delta i_L E_i t_{T\,var} = \frac{v_o^2}{R}\frac{1}{f_s} \tag{17.128}$$

Isolating the variable on-time t_{Tvar}, substituting for Δi_L, and using $v_o = \bar{I}_o R$ to eliminate R, gives

$$t_{T\,var} = t_T\sqrt{R_{crit}} \times \frac{1}{\sqrt{R}} = t_T\sqrt{\frac{R_{crit}}{v_o}} \times \sqrt{\bar{I}_o}$$
$$t_{T\,var} \;\alpha\; \frac{1}{\sqrt{R}} \text{ or } t_{T\,var} \;\alpha\; \sqrt{\bar{I}_o} \tag{17.129}$$

That is, once discontinuous inductor current commences, if the switching frequency f_s remains constant, regulation of the output voltage v_o can be maintained if the switch on-state period t_T is varied

- proportionally with the square root of the load current, $\sqrt{\bar{I}_o}$
- inversely with the square root of the load resistance, $\sqrt{R_{crit}}$
- inversely with the square root of the output voltage, $\sqrt{v_o}$.

Example 17.6: ***Reversible forward converter***

The step-down reversible converter in figure 17.10a operates at a switching frequency of 10 kHz. The output voltage is to be fixed at 48 V dc across a 1 Ω resistive load. If the input voltage E_i = 192 V and the choke L = 200μH:

i. calculate the switch T on-time duty cycle δ and switch on-time t_T
ii. calculate the average load current $\bar{I}_o$, hence average input current $\bar{I}_i$
iii. draw accurate waveforms for
- the voltage across, and the current through L; v_L and i_L
- the capacitor current, i_c
- the switch and diode voltage and current; v_T, v_D, i_T, i_D

iv. calculate
- the maximum load resistance R_{crit} before discontinuous inductor current with L=200μH and
- the value to which the inductance L can be reduced before discontinuous inductor current, if the maximum load resistance is 1Ω.

Solution

i. The switch on-state duty cycle δ can be calculate from equation (17.116), that is

$$2\delta - 1 = \frac{v_o}{E_i} = \frac{48\text{V}}{192\text{V}} = ¼ \quad \Rightarrow \quad \delta = ⅜$$

Also, from equation (17.116), for a 10kHz switching frequency, the switching period τ is 100μs and the transistor on-time t_T is given by

$$\delta = \frac{t_T}{\tau} = \frac{t_T}{100\mu\text{s}} = ⅜$$

whence the transistor on-time is 62½μs and the diodes conduct for 37½μs.

ii. The average load current is $\bar{I}_o = \frac{v_o}{R} = \frac{48\text{V}}{1\Omega} = 48\text{A} = \bar{I}_L$

From power-in equals power-out, the average input current is

$$\bar{I}_i = v_o \bar{I}_o / E_i = 48\text{V} \times 48\text{A}/192\text{V} = 12\text{A}$$

iii. The average output current is the average inductor current, 48A. The ripple current is given by equation (17.116), that is

$$\Delta i_L = \hat{i}_L - \check{i}_L = \frac{E_i - v_o}{L} \times t_T$$

$$= \frac{192\text{V} - 48\text{V}}{200\mu\text{H}} \times 62.5\mu\text{s} = 45\text{A p-p}$$

iv. Critical load resistance is given by equation (17.125), namely

$$R_{crit} \le \frac{v_o}{\bar{I}_o} = \frac{(2\delta - 1)L}{\tau\delta(1-\delta)}$$

$$= \frac{(2 \times ⅜ - 1) \times 200\mu\text{H}}{100\mu\text{s} \times ⅜ \times (1 - ⅜)} = 32/15\Omega$$

$$= 2\frac{2}{15}\Omega \text{ when } \bar{I}_o = ½\Delta i_L = 22½\text{A}$$

Alternatively, the critical load current is 22½A (½Δi_L), thus the load resistance must not be greater than $v_o / \bar{I}_o$ = 48V/22.5A = 32/15Ω, if the inductor current is to be continuous.
The critical resistance formula given in equation (17.125) is valid for finding critical inductance when inductance is made the subject of the equation, that is, rearranging equation (17.125) gives

$$L_{crit} = R \times (1-\delta) \times \delta \times \tau / (2\delta - 1) \qquad \text{(H)}$$

$$= 1\Omega \times (1 - ⅜) \times ⅜ \times 100\mu\text{s} / (2 \times ⅜ - 1)$$

$$= 93¾\mu\text{H}$$

That is, the inductance can be decreased from 200μH to 93¾μH when the load is 1Ω and continuous inductor current will flow.

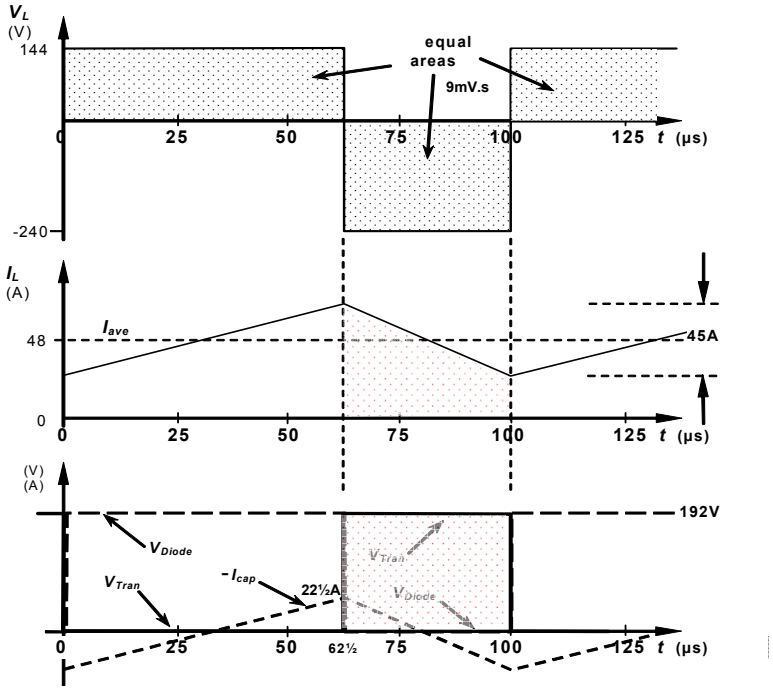


Figure: Example 17.5

♣

17.6.5 Comparison of the reversible converter with alternative converters

The reversible converter provides the full functional output range of the forward converter when δ > ½ and provides part of the voltage function of the buck-boost converter when δ < ½ but with energy transferring in the opposite direction.

Comparison of example 17.1 and 17.5 shows that although the same output voltage range can be achieved, the inductor ripple current is much larger for a given inductance L. A similar result occurs when compared with the buck-boost converter. Thus in each case, the reversible converter has a narrower output resistance range before discontinuous inductor conduction occurs. It is therefore concluded that the reversible converter should only be used if two quadrant operation is needed.

The ripple current $\bar{I}_f$ given by equation (17.2) for the forward converter and equation (17.114) for the reversible converter when $v_o > 0$, yield the following current ripple relationship.

$$\bar{I}_f = (2 - 1/\delta_r) \times \bar{I}_r$$

where $2\delta_r - 1 = \delta_f$ for $0 \le \delta_f \le 1$ and $½ \le \delta_r \le 1$ (17.130)

This equation shows that the ripple current of the forward converter $\bar{I}_f$ is never greater than the ripple current $\bar{I}_r$ for the reversible converter, for the same output voltage.
In the voltage inverting mode, from equations (17.77) and (17.114), the relationship between the two corresponding ripple currents is given by

$$\bar{I}_{fly} = \frac{2(\delta_r - 1)}{2\delta_r - 1} \times \bar{I}_r$$

where $\frac{2(\delta_r - 1)}{2\delta_r - 1} = \delta_{fly}$ for $0 \le \delta_{fly} \le ½$ and $0 \le \delta_r \le ½$ (17.131)

Again the reversible converter always has the higher inductor ripple current. Essentially the higher ripple current results in each mode because the inductor energy release phase involving the diodes, occurs back into the supply, which is effectively in cumulative series with the output capacitor voltage.

The reversible converter offers some functional flexibility, since it can operate as a conventional forward converter, when only one of the two switches is turned off. (In fact, in this mode, switch turn-off is alternated between T_1 and T_2 so as to balance switch and diode losses.)

17.7 The Ćuk converter

The Ćuk converter in figure 17.11 performs an inverting boost converter function with inductance in the input and the output. As a result, both the input and output currents can be continuous. A capacitor is used in the process of transferring energy from the input to the output and ac couples the input boost converter stage (L_1, T) to the output forward converter (D, L_2). Specifically, the capacitor C_1 ac couples the switch T in the boost converter stage into the output forward converter stage.

17.7.1 Continuous inductor current

When the switch T is on and the diode D is reversed biased

$$i_{C1(on)} = -\bar{I}_{L2} = \bar{I}_o \tag{17.132}$$

When the switch is turned off, inductor currents i_{L1} and i_{L2} are divert through the diode and

$$i_{C1(off)} = \bar{I}_i \tag{17.133}$$

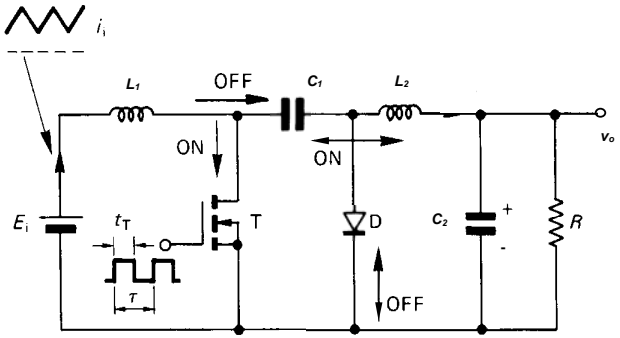


Figure 17.11. *Basic Ćuk converter.*

Over one steady-state cycle the average capacitor charge is zero, that is

$$i_{C1(on)}\delta\tau + i_{C1(off)}(1-\delta)\tau = 0 \tag{17.134}$$

which gives

$$\frac{i_{C1(on)}}{i_{C1(off)}} = \frac{\delta}{(1-\delta)} = \frac{\bar{I}_i}{\bar{I}_o} \tag{17.135}$$

From power-in equals power-out

$$\frac{v_o}{E_i} = \frac{\bar{I}_i}{\bar{I}_o} = \frac{\bar{I}_{L1}}{\bar{I}_{L2}} \tag{17.136}$$

Thus equation (17.135) becomes

$$\frac{v_o}{E_i} = \frac{\bar{I}_i}{\bar{I}_o} = \frac{\bar{I}_{L1}}{\bar{I}_{L2}} = -\frac{\delta}{(1-\delta)} \tag{17.137}$$

17.7.2 Discontinuous inductor current

The current rise in L_1 occurs when the switch is on, that is

$$\Delta i_{L1} = \frac{\delta\tau E_i}{L_1} \tag{17.138}$$

For continuous current in the input inductor L_1,

$$\bar{I}_i = \bar{I}_{L1} \geq \tfrac{1}{2}\Delta i_{L1} \tag{17.139}$$

which yields a maximum allowable load resistance, for continuous inductor current, of

$$R_{crit} \leq \frac{v_o}{\bar{I}_o} = \frac{2\delta L_1}{\tau(1-\delta)^2} = \frac{2f_s L_1\delta}{(1-\delta)^2} = \frac{\delta X_{L1}}{\pi(1-\delta)^2} \tag{17.140}$$

This is the same expression as that obtained for the boost converter, equation (17.69), which can be re-arranged to give the minimum inductance for continuous input inductor current, namely

$$\breve{L}_1 = \frac{(1-\delta)^2 R\tau}{2\delta} \tag{17.141}$$

The current rise in L_2 occurs when the switch is on and the inductor voltage is E_i, that is

$$\Delta i_{L2} = \frac{\delta\tau E_i}{L_2} \tag{17.142}$$

For continuous current in the output inductor L_2,

$$\bar{I}_o = \bar{I}_{L2} \geq \tfrac{1}{2}\Delta i_{L2} \tag{17.143}$$

which yields

$$R_{crit} \leq \frac{v_o}{\bar{I}_o} = \frac{2L_2}{\tau(1-\delta)} = \frac{2f_s L_2}{(1-\delta)} = \frac{X_{L2}}{\pi(1-\delta)} \tag{17.144}$$

This is the same expression as that obtained for the forward converter, equation (17.26) which can be re-arranged to give the minimum inductance for continuous output inductor current, namely

$$\breve{L}_2 = \tfrac{1}{2}(1-\delta)R\tau \tag{17.145}$$

17.7.3 Optimal inductance relationship

Optimal inductor conditions are that both inductors should both simultaneous reach the verge of discontinuous conduction. The relationship between inductance and ripple current is given by equations (17.138) and (17.142).

$$\Delta i_{L1} = \frac{\delta\tau E_i}{L_1} \text{ and } \Delta i_{L2} = \frac{\delta\tau E_i}{L_2}$$

After diving these two equations

$$\frac{L_2}{L_1} = \frac{\Delta i_{L1}}{\Delta i_{L2}} \tag{17.146}$$

Critical inductance is given by equations (17.141) and (17.145), that is

$$\breve{L}_2 = \tfrac{1}{2}(1-\delta)R\tau \quad \text{and} \quad \breve{L}_1 = \frac{(1-\delta)^2 R\tau}{2\delta}$$

After dividing

$$\frac{\breve{L}_2}{\breve{L}_1} = \frac{\delta}{1-\delta} \tag{17.147}$$

At the verge of simultaneous discontinuous inductor conduction

$$\frac{\breve{L}_2}{\breve{L}_1} = \frac{\delta}{1-\delta} = \frac{\Delta i_{L1}}{\Delta i_{L2}} = \left|\frac{v_o}{E_i}\right| \tag{17.148}$$

That is, the voltage transfer ratio uniquely specifies the ratio of the minimum inductances and their ripple current.

17.7.4 Output voltage ripple

The output stage (L_2, C_2, and R) is the forward converter output stage; hence the per unit output voltage ripple on C_2 is given by equation (17.35), that is

$$\frac{\Delta v_{C2}}{v_o} = \frac{\Delta v_o}{v_o} = \tfrac{1}{8} \times \frac{(1-\delta)\tau^2}{L_2 C_2} \tag{17.149}$$

If the ripple current in L_1 is assumed constant, the per unit voltage ripple on the ac coupling capacitor C_1 is approximated by

$$\frac{\Delta v_{C1}}{v_o} = \frac{\delta\tau}{RC_1} \tag{17.150}$$

Example 17.7: ***Ćuk converter***

The Ćuk converter in figure 17.11 is to operate at 10kHz from a 50V battery input and produces an inverted non-isolated 75V output. The load power is 1.8kW.

i. Calculate the duty cycle hence switch on and off times, assuming continuous current in both inductors.
ii. Calculate the mean input and output, hence inductor, currents.
iii. At the 1.8kW load level, calculate the inductances L_1 and L_2 such that the ripple current is 1A p-p in each.
iv. Specify the capacitance for C_1 and C_2 if the ripple voltage is to be a maximum of 1% of the output voltage.
v. Determine the critical load resistance for which the purely duty cycle dependant voltage transfer function becomes invalid.
vi. At the critical load resistance value, determine the inductance value to which the non-critically operating inductor can be reduced.
vii. Determine the necessary conditions to ensure that both inductors operate simultaneously on the verge of discontinuous conduction, and the relative ripple currents for that condition.

Solution

i. The voltage transfer function is given by equation (17.137), that is

$$\frac{v_o}{E_i} = -\frac{\delta}{(1-\delta)} = -\frac{75\text{V}}{50\text{V}} = -1\tfrac{1}{2}$$

from which $\delta = \tfrac{3}{5}$. For a 10kHz switching frequency the period is 100µs, thus the switch on-time is 60µs and the off-time is 40µs.

ii. The mean output current is determined by the load and the mean input current is related to the output current by assuming 100% efficiency, that is

$$\bar{I}_o = \bar{I}_{L2} = P_o / v_o = 1800\text{W} / 75\text{V} = 24\text{A}$$
$$\bar{I}_i = \bar{I}_{L1} = P_o / E_i = 1800\text{W} / 50\text{V} = 36\text{A}$$

The load resistance is therefore $R = v_o/I_o$ = 75V/24A = 3⅛Ω.

iii. The inductor ripple current for each inductor is given by the same expression, that is equations (17.138) and (17.142). Thus for the same ripple current of 1A pp

$$\Delta i_{L1} = \frac{\delta\tau E_i}{L_1} = \Delta i_{L2} = \frac{\delta\tau E_i}{L_2}$$

which gives

$$L_1 = L_2 = \frac{\delta\tau E_i}{\Delta i} = \frac{\tfrac{3}{5}\times 100\mu\text{s}\times 50\text{V}}{1\text{A}} = 3\text{mH}$$

iv. The capacitor ripple voltages are given by equations (17.150) and (17.149), which after re-arranging gives

$$C_1 = \frac{v_o}{\Delta v_{C1}}\times\frac{\delta\tau}{R} = \frac{100}{1}\times\frac{\tfrac{3}{5}\times 100\mu\text{s}}{\tfrac{25}{8}\Omega} = 1.92\text{mF}$$
$$C_2 = \frac{v_o}{\Delta v_{C2}}\times\tfrac{1}{8}\times\frac{(1-\delta)\tau^2}{L_2} = \frac{100}{1}\times\tfrac{1}{8}\times\frac{(1-\tfrac{3}{5})\times 100\mu\text{s}^2}{3\text{mH}} = 16.6\mu\text{F}$$

v. The critical load resistance for each inductor is given by equations (17.140) and (17.144). When both inductors are 3mH:

$$R_{crit} \le \frac{2\delta L_1}{\tau(1-\delta)^2} = \frac{2\times\tfrac{3}{5}\times 3\text{mH}}{100\mu\text{s}\times(1-\tfrac{3}{5})^2} = 225\Omega$$
$$R_{crit} \le \frac{2L_2}{\tau(1-\delta)} = \frac{2\times 3\text{mH}}{100\mu\text{s}\times(1-\tfrac{3}{5})} = 150\Omega$$

The limiting critical load resistance is 150Ω or for $I_o = v_o/R$ = 75V/150Ω = ½A, when a lower output current results in the current in L_2 becoming discontinuous although the current in L_1 is still continuous.

vi. From equation (17.140), rearranged

$$L_{1\,crit} \ge \frac{\tau R(1-\delta)^2}{2\delta} = \frac{100\mu\text{s}\times 100\Omega\times(1-\tfrac{3}{5})^2}{2\times\tfrac{3}{5}} = 2\text{mH}$$

That is, if L_1 is reduced from 3mH to 2mH, then both L_1 and L_2 enter discontinuous conduction at the same load condition, 75V, ½A, and 150Ω.

vii. For both converter inductors to be simultaneously on the verge of discontinuous conduction, equation (17.148) gives

$$\frac{\check{L}_2}{\check{L}_1} = \frac{\delta}{1-\delta} = \frac{\Delta i_{L1}}{\Delta i_{L2}} = \left|\frac{v_o}{E_i}\right|$$
$$\frac{3\text{mH}}{2\text{mH}} = \frac{\tfrac{3}{5}}{1-\tfrac{3}{5}} = \frac{1\text{A}}{\tfrac{2}{3}\text{A}} = \left|\frac{75\text{V}}{50\text{V}}\right| = \frac{3}{2}$$

♣

17.8 Comparison of basic converters

The converters considered employ an inductor to transfer energy from one dc voltage level to another dc voltage level. The basic converters comprise a switch, diode, inductor, and a capacitor. The reversible converter is a two-quadrant converter with two switches and two diodes, while the Ćuk converter uses two inductors and two capacitors.

Table 17.1 summarises the main electrical features and characteristics of each basic converter. Figure 17.12 shows a plot of the voltage transformation ratios and the switch utilisation ratios of the converters considered. With reference to figure 17.12, it should be noted that the flyback step-up/step-down converter and the Ćuk converter both invert the input polarity.

Every converter can operate in any one of three inductor current modes:

- discontinuous
- continuous
- both continuous and discontinuous

The main converter operational features of continuous conduction compared with discontinuous inductor conduction are

- The voltage transformation ratio (transfer function) is independent of the load.
- Larger inductance but lower core hysteresis losses and saturation less likely.
- Higher converter costs with increased volume and weight.
- Worse transient response *(L/R)*.
- Power delivered is inversely proportional to load resistance, $P = V_o^2/R$. In the discontinuous conduction mode, power delivery is inversely dependent on inductance.

17.8.1 Critical load current

Examination of Table 17.1 shows no obvious commonality between the various converters and their performance factors and parameters. One common feature is the relationship between critical average output current $\bar{I}_o$ and the input voltage E_i at the boundary of continuous and discontinuous conduction. Equations (17.14), (17.65), and (17.95) are identical, (for all smps), that is

$$\bar{I}_{o_{critical}} = \frac{E_i\tau}{2L}\delta(1-\delta) \quad \text{(A)} \qquad (17.151)$$

This quadratic expression in δ shows that the critical mean output current reduces to zero as the on-state duty cycle δ tends to zero or unity. The maximum critical load current condition, for a given input voltage E_i, is when δ = ½ and

$$\hat{\bar{I}}_{o_c} = E_i\tau/8L \qquad (17.152)$$

Since power-in equals power-out, then from equation (17.151) the input average current and output voltage at the boundary of continuous conduction for all smps are related by

$$\bar{I}_{i_{critical}} = \frac{v_o\tau}{2L}\delta(1-\delta) \quad \text{(A)} \qquad (17.153)$$

The maximum output current at the boundary (at δ = ½), for a given output voltage, v_o, is

$$\hat{\bar{I}}_{i_c} = v_o\tau/8L \qquad (17.154)$$

The smps commonality factor reduces to $R_{crit} = \frac{v_o}{E_i} \times \frac{2L}{\tau\,\delta(1-\delta)}$.

The reversible converter, using the critical resistance equation (17.125) derived in section 17.6.3, yields twice the critical average output current given by equation (17.151). This is because its duty cycle range is restricted to half that of the other converters considered. Converter normalised equations for discontinuous conduction are shown in table 17.2.
A detailed analysis summary of discontinuous inductor current operation is given in Appendix 17.11.

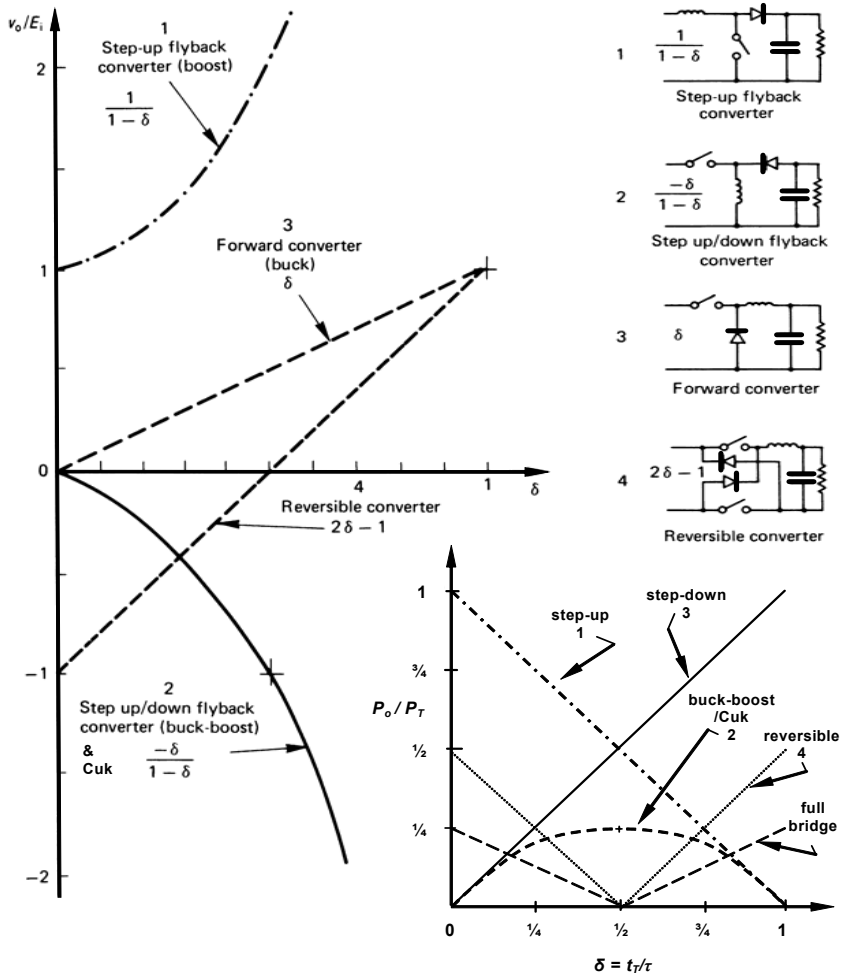


Figure 17.12. *Transformation voltage ratios and switch utilisation ratios for five converters when operated in the continuous inductor conduction mode.*

Table 17.1: Converter characteristics comparison with continuous inductor current

			converter			
			Forward Step-down	*Flyback Step-up*	*Flyback Step-up/down*	*Reversible*
Output voltage continuous *I*	v_o/E_i		δ	$\frac{1}{1-\delta}$	$-\frac{\delta}{1-\delta}$	$2\delta - 1$
Output voltage discontinuous *I*	v_o/E_i		$1 - \frac{2L\bar{I}_i}{E_i\delta^2\tau}$	$1 + \frac{E_i\delta^2 t_T}{2L\bar{I}_o}$	$\frac{E_i\delta^2\tau}{2L\bar{I}_o}$	
Output polarity with respect to input			Non-inverted	Non-inverted	inverted	any
Current sampled from the supply			discontinuous	continuous	discontinuous	bi-directional
Load current			continuous	discontinuous	discontinuous	continuous
Maximum transistor voltage	*V*	V	E_i	v_o	$E_i + v_o$	E_i
Maximum diode voltage	*V*	V	E_i	v_o	$E_i + v_o$	E_i
Ripple current	Δi	A	$E_i\delta\tau(1-\delta)/L$	$E_i\delta\tau/L$	$E_i\delta\tau/L$	$2E_i\delta\tau(1-\delta)/L$
Maximum transistor current	$\hat{i}_T$	A	$\bar{I}_o + \frac{v_o\tau(1-\delta)}{2L}$	$\bar{I}_i + \frac{E_i\tau\delta}{2L}$	$\bar{I}_L + \frac{E_i\tau\delta}{2L}$	$\bar{I}_o + \frac{(E_i - v_o)\tau\delta}{2L}$
switch utilisation ratio	*SUR*		δ	1-δ	δ (1-δ)	**½δ**
Transistor rms current			low	high	high	low
Critical load resistance	R_{crit}	Ω	$\frac{2L}{\tau(1-\delta)}$	$\frac{2L}{\tau\delta(1-\delta)^2}$	$\frac{2L}{\tau(1-\delta)^2}$	$\frac{2(\delta - ½)L}{\tau\delta(1-\delta)}$
Critical inductance	L_{crit}	H	$½R(1-\delta)\tau$	$½R\tau\delta(1-\delta)^2$	$½R\tau(1-\delta)^2$	$\frac{½R(1-\delta)\delta\tau}{(\delta - ½)}$
o/p ripple voltage p-p	Δv_o	V	$\frac{\tau^2(1-\delta)}{8LC}v_o$	$\frac{\tau\delta}{RC}v_o$	$\frac{\tau\delta}{RC}v_o$	$\frac{\tau\delta}{RC}v_o$

Table 17.2: Comparison of characteristics when the inductor current is discontinuous, $\delta < \delta_{critical}$

$k = \frac{R\tau}{L};\quad \delta = \frac{t_T}{\tau}$	converter		
	Forward step-down	*Flyback step-up*	*Flyback step-up/down*
$\delta_{critical}(k) =$	$\delta \le 1 - \frac{2}{k}$	$k > \frac{27}{2}$ then $\delta(1-\delta)^2 \le \frac{2}{k}$	$\delta \le 1 - \sqrt{\frac{2}{k}}$
$\frac{v_o}{E_i}(k,\delta) = \frac{\bar{I}_o R}{E_i}$	$¼k\delta^2\left[-1 + \sqrt{1 + \frac{8}{k\delta^2}}\right]$	$½\left[1 + \sqrt{1 + 2k\delta^2}\right]$	$-\delta\sqrt{½k}$
$\delta_D = \frac{t_D}{\tau}(k,\delta)$	$\delta \times \left(1 - \frac{v_o}{E_i}\right) / \frac{v_o}{E_i}$	$\delta / \left(\frac{v_o}{E_i} - 1\right)$	$\delta / \left\lvert\frac{v_o}{E_i}\right\rvert$
$\delta_x = \frac{t_x}{\tau}(k,\delta)$ $= 1 - \delta - \delta_D$	$1 - \delta / \frac{v_o}{E_i}$	$1 - \delta \times \frac{v_o}{E_i} / \frac{v_o}{E_i} - 1$	$1 - \delta\left(1 + \left\lvert\frac{v_o}{E_i}\right\rvert\right) / \left\lvert\frac{v_o}{E_i}\right\rvert$
$\hat{I}_L \times \frac{R}{E_i}(k,\delta)$	$k\delta \times \left[1 - \frac{v_o}{E_i}\right]$	$k\delta$	$k\delta$

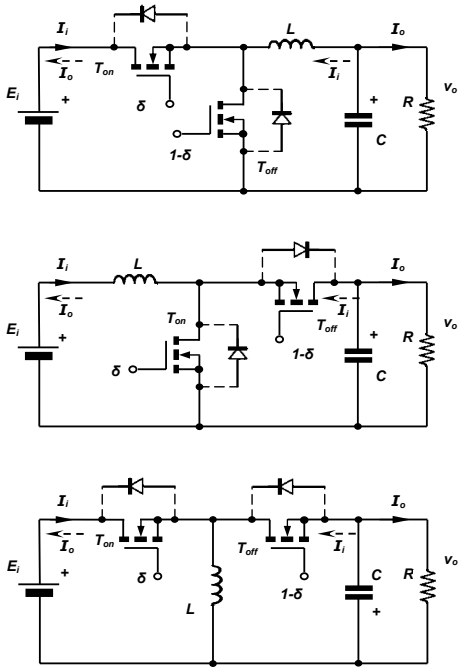


Figure 17.13. *The three basic bidirectional current converter configurations: (a) the forward converter; (b) step-up flyback converter; and (c) step up/down flyback converter.*

17.8.2 Bidirectional converters

Discontinuous inductor current can be avoided if the smps diode is parallel connected with a shunt switch as shown in figure 17.13. If the switch has bipolar conduction properties, as with the MOSFET, then it can perform three functions

- *Synchronised rectification*: If the shunting switch conducts when the diode conducts, during period δ_D, then the diode is bypassed and losses are reduced to those of the MOSFET, which can be less than those of a Schottky diode.
- *Guaranteed continuous inductor current conduction*: If the shunting switch conducts for the period 1- δ_D, (complement to the main smps switch) then if the inductor current falls to zero, that current can reverse with energy taken from the output capacitor. Seamless, continuous inductor current results and importantly, the voltage transfer function is then that for continuous inductor current, independent of the load resistance.
- *Bidirectional energy transfer:* If the output diode has a shunting switch and an inverse parallel diode is added across the converter main switch (or both switches have bidirectional conduction properties, as with the MOSFET) then power can be efficiently and seamlessly transferred in either direction, between E_i and v_o. The voltage polarities are unchanged – it is the current direction that reverses. The buck and boost converters interchange transfer functions when operating in the reverse direction, while the buck/boost converter has the same transfer function in both current directions of operation.

17.8.3 Isolation

In each converter, the output is not electrically isolated from the input and a transformer can be used to provide isolation. Figure 17.14 shows isolated versions of the three basic converters. The transformer turns ratio provides electrical isolation as well as matching to obtain the required output voltage range.

- Figure 17.14a illustrates an isolated version of the forward converter shown in figure 17.2. When the transistor is turned on, diode D_1 conducts and *L* in the transformer secondary stores energy. When the transistor turns off, the diode D_3 provides a current path for the release of the energy stored in *L*. However when the transistor turns off and D_1 ceases to conduct, the stored transformer magnetising energy must be released. The winding incorporating D_2 provides a path to reset the core flux. A maximum possible duty cycle exists, depending on the turns ratio of the primary winding and freewheel winding. If a 1:1 ratio (as shown) is employed, a 50 per cent duty cycle limit will ensure the required volts-second for core reset.
- The step-up flyback isolated converter in part b of figure 17.14 is little used. The two transistors must be driven by complementary signals. Core leakage and reset functions (and no-load operation) are facilitated by a third winding and blocking diode D_2.

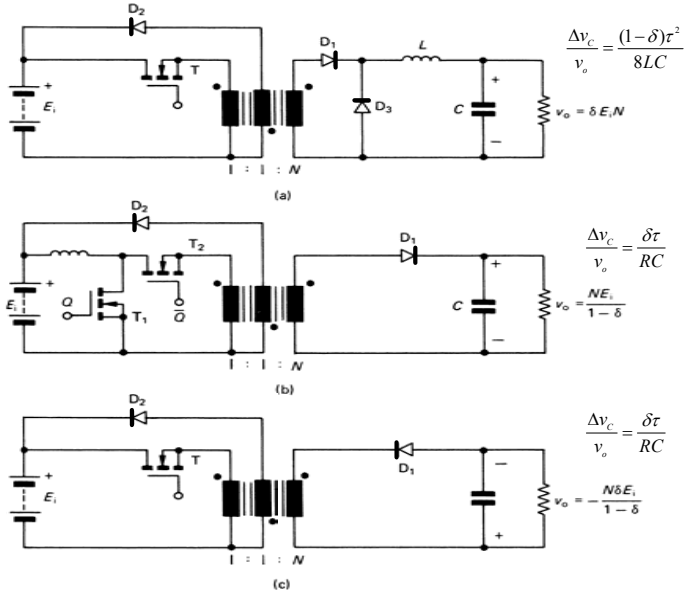


Figure 17.14. *Isolated output versions of the three basic converter configurations: (a) the forward converter; (b) step-up flyback converter; and (c) step up/down flyback converter.*

- The magnetic core in the buck-boost converter of part c of figure 17.14 performs a bifilar inductor function. When the transistor is turned on, energy is stored in the core. When the transistor is turned off, the core energy is released via the secondary winding into the capacitor. A core air gap is necessary to prevent magnetic saturation and an optional clamping winding can be employed, which operates at zero load.

The converters in parts a and c of figure 17.14 provide an opportunity to compare the main features and attributes of forward and flyback isolated converters. In the comparison it is assumed that the transformer turns ratio is 1:1:1.

17.8.3i - *The isolated output,* ***forward converter*** *– figure 17.14a:*

- $v_o = n_T \delta E_i$ or $I_i = n_T \delta I_o$
- The magnetic element acts as a transformer, that is, because of the relative voltage polarities of the windings, energy is transferred from the input to the output, and not stored in the core, when the switch is on. A small amount of magnetising energy, due to the magnetising current to flux the core, is built up in the core.
- The magnetising flux is reset by the current through the catch (feedback) winding and D_3, when the switch is off. The magnetising energy is recovered and returned to the supply E_i.
- The necessary transformer Vμs balance requirement (core energy-in equals core energy-out) means the maximum duty cycle is limited to $0 \le \delta \le 1/\left(1+n_{f/b}\right) < 1$ for 1:$n_{f/b}$:n_{sec} turns ratio. For example, the duty cycle is limited to 50%, 0 ≤ δ ≤ ½, with a 1:1:1 turns ratio.

- Because of the demagnetising winding, the off-state switch supporting voltage is $E_i + v_o$.
- The blocking voltage requirement of diode D_3 is E_i, v_o for D_1, and $2E_i$ for D_2.
- The critical load resistance for continuous inductor current is independent of the transformer:

$$R_{crit} \le \frac{4L}{\tau(1-2\delta)} \qquad (17.155)$$

17.8.3ii - *The isolated output,* ***flyback converter*** *– figure 17.14c:*

- $v_o = n_T E_i \delta/(1-\delta)$ or $I_i = n_T I_o \delta/(1-\delta)$
- The magnetic element acts as a magnetic energy storage inductor. Because of the relative voltage polarities of the windings (dot convention), when the switch is on, energy is stored in the core and no current flows in the secondary.
- The stored energy, which is due to the core magnetising flux is released (reset) as current into the load and capacitor C when the switch is off. (Unlike the forward converter, where magnetising energy is returned to E_i, not the output, v_o.) Therefore there is no flyback converter duty cycle restriction, $0 \le \delta \le 1$.
- The third winding turns ratio is configured such that energy is only returned to the supply E_i under no load conditions.
- The switch supporting off-state voltage is $E_i + v_o$.
- The diode blocking voltage requirements are $E_i + v_o$ for D_1 and $2E_i$ for D_2.
- The critical load resistance for continuous inductor current is independent of the transformer turns ratio when the magnetising inductance is referenced to the secondary:

$$R_{crit} \le \frac{4L_{msec}}{\tau(1-2\delta)^2} = \frac{4\eta_T^2 L_{mprim}}{\tau(1-2\delta)^2} \qquad (17.156)$$

The operational characteristics of each converter change considerably when the flexibility offered by tailoring the turns ratio is exploited. A multi-winding magnetic element design procedure is outlined in section 9.1.1, where the transformer turns ratio (n_p:n_s) is not necessarily 1:1.

The basic approach to any transformer (coupled circuit) problem is to transfer, or refer, all components and variables to either the transformer primary or secondary circuit, whilst maintaining power and time invariance. Thus, maintaining power-in equals power-out, and assuming a secondary to primary turns ratio of n_T is to one (n_T:1), gives

$$\frac{v_s}{v_p} = \frac{n_s}{n_p} = n_T \qquad \frac{i_p}{i_s} = \frac{n_s}{n_p} = n_T \qquad \frac{Z_s}{Z_p} = \left(\frac{n_s}{n_p}\right)^2 = n_T^2 \qquad (17.157)$$

Time, that is switching frequency, power, and per unit values (δ, $\Delta v_o/v_o$), are invariant. The circuit is then analysed without a transformer. Subsequently, the appropriate parameters are referred back to their original side of the magnetically coupled circuit.
If the coupled circuit is used as a transformer, magnetising current (flux) builds, which must be reset to zero each cycle. Consider the transformer coupled forward converter in figure 17.14a. From Faraday's equation, $v = Nd\phi/dt$, and for maximum on-time duty cycle $\hat{\delta}$ the conduction V-μs of the primary must equal the conduction V-μs of the feedback winding which is returning the magnetising energy to the supply E_i.

$$E_i t_{on} = \frac{E_i}{n_{f/b}} t_{off} \quad \text{and} \quad t_{on} + t_{off} = \tau \qquad (17.158)$$

That is

$$E_i \hat{\delta} = \frac{E_i}{n_{f/b}}\left(1-\hat{\delta}\right)$$

$$\hat{\delta} = \frac{1}{1+n_{f/b}} \qquad (17.159)$$

$$0 \le \delta \le \frac{1}{1+n_{f/b}}$$

From Faraday's Law, the magnetizing current starts from zero and increases linearly to

$$\hat{I}_M = E_i t_{on} / L_M \qquad (17.160)$$

where L_M is the magnetizing inductance referred to the primary. During the switch off period, this current falls linearly, as energy is returned to E_i. The current must reach zero before the switch is turned on again, whence the energy taken from E_i and stored as magnetic fluxing energy in the core, has been returned to the supply.

Two examples illustrate the features of magnetically coupled circuit converters. Example 17.8 illustrates how the coupled circuit in the flyback converter acts as an inductor, storing energy from the primary source, and subsequently releasing that energy in the secondary circuit. In example 17.9, the forward converter coupled circuit acts as a transformer where energy is transferred through the core under transformer action, but in so doing, self-inductance (magnetising) energy is built up in the core, which must be periodically released if saturation is to be avoided. Relative orientation of the windings, according to the flux dot convention shown in figure 17.14, is thus important, not only the primary relative to the secondary, but also relative to the feedback winding.

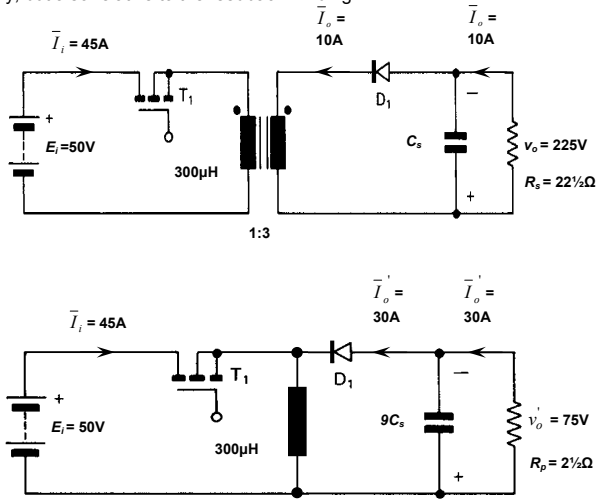


Figure 17.15. *Isolated output step up/down flyback converter and its equivalent circuit when the secondary output is referred to the primary.*

Example 17.8: ***Transformer coupled flyback converter***

The 10kHz flyback converter in figure 17.14c operates from a 50V input and produces a 225V dc output from a 1:1:3 (1:n_{fb}:n_{sec}) step-up transformer loaded with a 22½Ω resistor. The transformer magnetising inductance is 300μH, referred to the primary (or 300uH×3^2 = 2.7mH referred to the secondary):

i. Calculate the switch duty cycle, hence transistor off-time, assuming continuous inductor current.
ii. Calculate the mean input and output current.
iii. Draw the transformer currents, showing the minimum and maximum values.
iv. Calculate the capacitor rms ripple current and p-p voltage ripple if C = 1100μF.
v. Determine
- the critical load resistance
- the minimum inductance for continuous inductor conduction for a 22½ Ω load.

Solution

The feedback winding does not conduct during normal continuous inductor current operation. This winding can therefore be ignored for analysis during normal operation.
Figure 17.15 shows secondary parameters referred to the primary, specifically

$$v_o = 225\text{V} \quad v_o' = v_o / n_T = 225\text{V}/3 = 75\text{V}$$

$$R_s = 225\Omega \quad R_p = R_s / n_T^2 = 225\Omega / 3^2 = 22\tfrac{1}{2}\Omega$$

Note that the output capacitance is transferred by a factor of nine, n_T^2, since capacitive reactance is inversely proportion to capacitance ($X = 1/\omega C$).

It will be noticed that the equivalent circuit parameter values to be analysed, when referred to the primary, are the same as in example 17.5. The circuit is analysed as in example 17.5 and the essential results from example 17.5 are summarised in Table 17.3 and transferred to the secondary where appropriate. The waveform answers to part iii are shown in figure 17.16.

Table 17.3: Transformer coupled flyback converter analysis

parameter		value for primary analysis	transfer factor $n_T = 3$ →	value for secondary analysis
E_i	V	50	3	150
v_o	V	75	3	225
R_L	Ω	2½	3^2	22½
C_o	μF	10,000	3^{-2}	1100
L_M	μH	300	3^{-2}	2700
$I_{o(ave)}$	A	30	⅓	10
P_o	W	2250	invariant	2250
$I_{i(ave)}$	A	45	⅓	15
δ	p.u.	⅗	invariant	⅗
τ	μs	100	invariant	100
t_{on}	μs	60	invariant	60
t_D	μs	40	invariant	40
f_s	kHz	10	invariant	10
Δi_L	A	10	⅓	10/3
$\bar{I}_L$	A	75	⅓	25
$\hat{I}_L$	A	80	⅓	80/3
$\check{I}_L$	A	70	⅓	70/3
i_{Crms}	A rms	36.8	⅓	13.3
R_{crit}	Ω	37½	3^2	337½
L_{crit}	μH	20	3^2	180
V_{Dr}	V	125	3	375
Δv_o	mV	180	3	540
$\Delta v_o / v_o$	p.u.	0.24%	invariant	0.24%

Note the invariance of power, P_o; normalised parameters δ, and $\Delta v_o/v_o$; and time t_{on}, t_D, τ, and $1/f$.

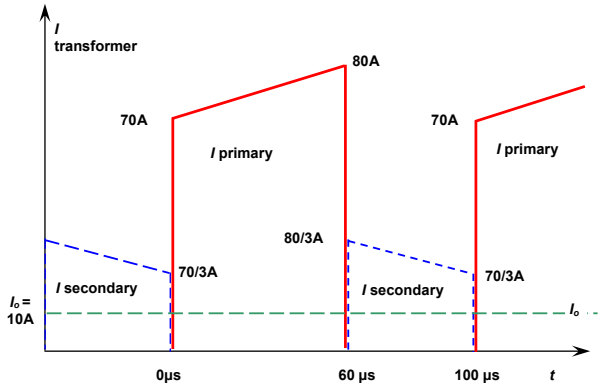


Figure 17.16. *Currents for the transformer windings in example 17.8.*

♣

Example 17.9: ***Transformer coupled forward converter***

The 10kHz forward converter in figure 17.14a operates from a 192V dc input and a 1:3:2 (1:$n_{f/b}$:n_{sec}) step-up transformer loaded with a 4Ω resistor. The transformer magnetising inductance is 1.2mH, referred to the primary. The secondary smps inductance is 800μH.

i. Calculate the maximum switch duty cycle, hence transistor off-time, assuming continuous inductor current.

At the maximum duty cycle:

ii. Calculate the mean input and output current.

iii. Draw the transformer currents, showing the minimum and maximum values.

iv. Determine

- the critical load resistance
- the minimum inductance for continuous inductor conduction for a 4 Ω load

Solution

i. The maximum duty cycle is determined solely by the transformer turns ratio between the primary and the feedback winding which resets the core flux. From equation (17.159)

$$\hat{\delta} = \frac{1}{1+n_{f/b}}$$

$$= \frac{1}{1+3} = ¼$$

The maximum conduction time is 25% of the 100μs period, namely 25μs. The secondary output voltage is therefore

$$v_{sec} = \delta n_f E_i$$

$$= ¼ \times 2 \times 192 = 96\text{V}$$

The load current is therefore 96V/4Ω = 24A, as shown in figure 17.17a.

Figure 17.17b shows secondary parameters referred to the primary, specifically

$$R_s = 4\Omega \quad R_p = R_s / n_T^2 = 4\Omega / 2^2 = 1\Omega$$

$$v_o = 96\text{V} \quad v_o' = v_o / n_T = 96\text{V}/2 = 48\text{V}$$

$$L_o = 800\mu\text{H} \quad L_o' = L_o / n_T^2 = 800\mu\text{H}/2^2 = 200\mu\text{H}$$

Note that the output capacitance is transferred by a factor of four, n_T^2, since capacitive reactance is inversely proportion to capacitance, $X = 1/\omega C$.

Inspection of example 17.1 will show that the equivalent circuit in figure 17.17b is the same as the circuit in example 17.1, except that a magnetising branch has been added. The various operating conditions and values in example 17.1 are valid for example 17.9.

ii. The mean output current is the same for both circuits (example 17.1), 48A, or 24 A when referred to the secondary circuit. The mean input current from E_i remains 12A, but the switch mean current is not 12A. Magnetising current is provided from the supply E_i through the switch, but returned to the supply E_i through diode D2, which bypasses the switch. The net magnetising energy flow is zero. The magnetising current maximum value is given by equation (17.160)

$$\hat{I}_M = E_i t_{on} / L_M$$

$$= 192\text{V} \times 25\mu\text{s}/1.2\text{mH} = 4\text{A}$$

This current increases the switch mean current from 12A to

$$\bar{I}_T = 12\text{A} + ½ \times \delta \times 4\text{A} = 12½\text{A}$$

Figure 17.17c show the equivalent circuit when the switch is off. The output circuit functions independently of the input circuit, which is returning stored core energy to the supply E_i via the feedback winding and diode D2. Parameters have been referred to the feedback winding which has three times the turns of the primary, $n_{f/b}$ =3. The 192V input voltage remains the circuit reference. Equation (17.160)-Faraday's law, referred to the feedback winding, must be satisfied during the switch off period, that is

$$\frac{\hat{I}_M}{n_{f/b}} = \frac{E_i t_{off}}{n_{f/b}^2 L_M}$$

$$\frac{4}{3} = \frac{192\text{V} \times 75\mu\text{s}}{3^2 \times 1.2\text{mH}}$$

The diode D2 voltage rating is $(n_{f/b}+1)\times E_i$, 768V and its mean current is

$$\bar{I}_{D2} = \tfrac{1}{2}(1-\delta)\frac{I_M}{n_{f/b}} = \tfrac{1}{2}\times(1-0.25)\times\frac{4\text{A}}{3} = \tfrac{1}{2}\text{A}$$

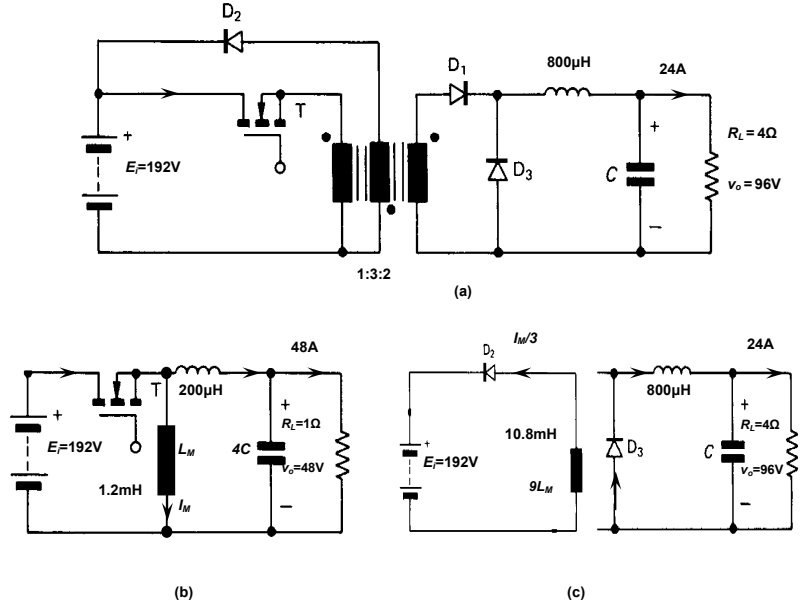


Figure 17.17. *Isolated output forward converter and its equivalent circuits when the output is referred to the primary.*

iii. The three winding currents for the transformer are shown in figure 17.18.

iv. The critical resistance and inductance, referred to the primary, from example 17.1 are 5⅓Ω and 37½μH. Transforming into secondary quantities, by multiplying by 2^2, give critical values of R_L = 21⅓Ω and L = 150μH.

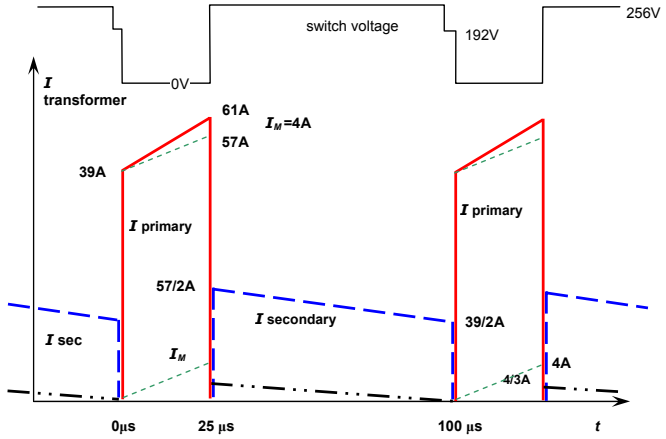


Figure 17.18. *Currents for the three transformer windings in example 17.9.*

♣

17.9 Multiple-switch, balanced, isolated converters

The basic single-switch converters considered have the limitation of using their magnetic components (whether as an inductor or transformer) only in a unipolar flux mode. Since only one quadrant of the *B-H* characteristic is employed, these converters are generally restricted to lower powers because of the limited flux swing, which is reduced by the core remanence flux.

The high-power forward converter circuits shown in figure 17.19 operate the magnetic transformer component in the bipolar or push-pull flux mode and require two or four switches. Because the transformers are fully utilised magnetically, they tend to be almost half the size of the equivalent single transistor isolated converter at power levels above 100 W. Also core saturation due to the magnetising current (flux) not being fully reset to zero each cycle, is not a major issue, since with balanced bidirectional fluxing, the average magnetising current (flux) is zero.
In each case, the transformer can be simplified to an auto-transformer, if isolation is not a requirement.

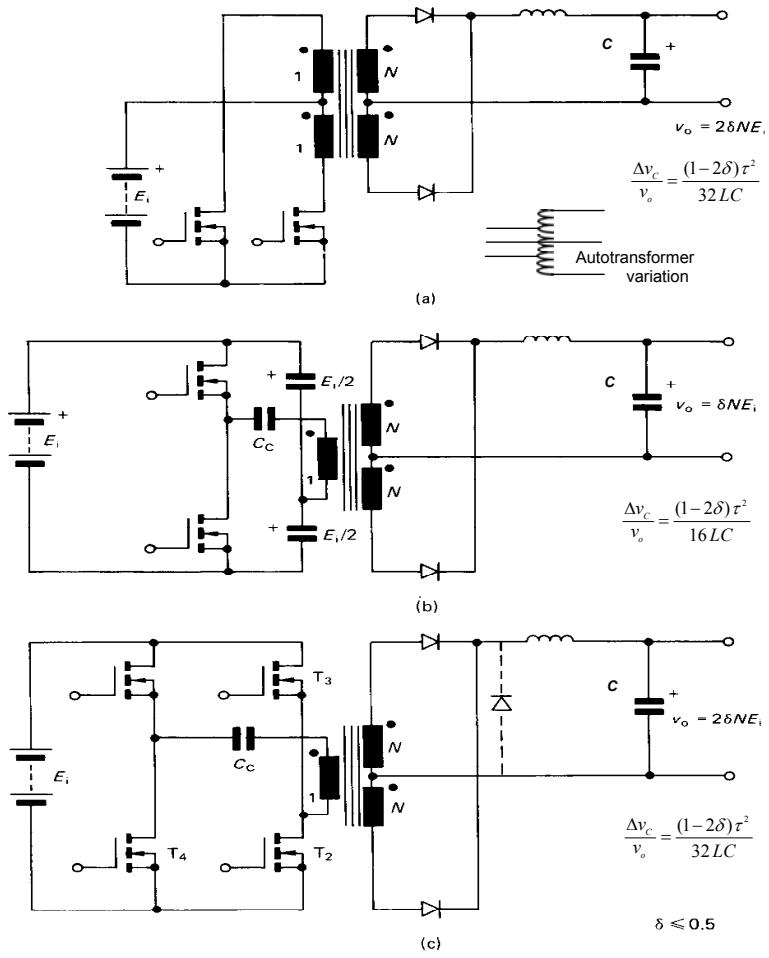


Figure 17.19. *Multiple-switch, isolated output, pulse-width modulated converters: (a) push-pull plus autotransformer option; (b) half-bridge; and (c) full-bridge.*

17.9.1 The push-pull converter

Figure 17.19a illustrates a push-pull forward converter circuit which employs two switches and a centre-tapped transformer. Each switch must have the same duty cycle in order to prevent unidirectional core saturation. Because of transformer coupling action, the off switch supports twice the input voltage, $2E_i$, plus any voltage associated with leakage inductance stored energy. Advantageously, no floating gate drives are required and importantly, no switch shoot through (simultaneous conduction) can occur.
The voltage transfer function, for continuous inductor current conduction, is based on the equivalent secondary output circuit show in figure 17.20. Because of transformer action, the input voltage is $N{\times}E_i$ where N is the transformer turns ratio. When a primary switch is on, current flows in the outer loop shown in figure 17.20. That is

$$\Delta i_L = \hat{i}_L - \check{i}_L = \frac{N \times E_i - v_o}{L} \times t_T \qquad (17.161)$$

When the primary switches are off, the secondary voltage falls to zero and current continues to flow through the secondary winding due to the energy stored in L. Efficiency is increased if the diode D_f is used to bypass the transformer winding, as shown in figure 17.20. The secondary winding i^2R losses are decreased and minimal voltage is coupled from the secondary back into the primary circuit. The current in the inner off loop shown in figure 17.20 is given by

$$\Delta i_L = \frac{v_o}{L} \times (\tau - t_T) \qquad (17.162)$$

Equating equations (17.161) and (17.162) gives the following voltage and current transfer function

$$\frac{v_o}{E_i} = \frac{\bar{I}_i}{\bar{I}_o} = 2N\frac{t_T}{\tau} = 2N\delta \qquad 0 \le \delta \le ½ \qquad (17.163)$$

The output voltage ripple is similar to that of the forward converter

$$\frac{\Delta v_C}{v_o} = \frac{\Delta v_o}{v_o} = \frac{(1-2\delta)\tau^2}{32LC} \qquad (17.164)$$

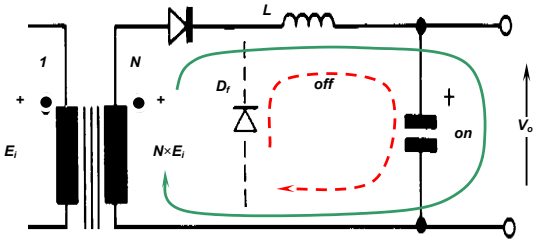


Figure 17.20. *Equivalent circuit for transformer bridge converters based on a forward converter in the secondary.*

17.9.2 Bridge converters

Figures 17.19b and c show half and full-bridge isolated forward converters respectively.

i. Half-bridge

In the half-bridge the transistors are switched alternately and must have the same conduction period. This ensures the core volts-second balance requirement to prevent saturation due to bias in one flux direction.

Using similar analysis as for the push-pull converter in 17.9.1, the voltage transfer function of the half bridge with a forward converter output stage, for continuous inductor conduction, is given by

$$\frac{v_o}{E_i} = \frac{\bar{I}_i}{\bar{I}_o} = N\frac{t_T}{\tau} = N\delta \qquad 0 \le \delta \le ½ \qquad (17.165)$$

A floating base drive is required. Although the maximum winding voltage is $½E_i$, the switches must support E_i in the off-state, when the complementary switch conducts.

The output ripple voltage is given by

$$\frac{\Delta v_C}{v_o} = \frac{\Delta v_o}{v_o} = \frac{(1-2\delta)\tau^2}{16LC} \qquad (17.166)$$

ii. Full-bridge

The full bridge in figure 17.19c replaces the capacitor supplies of the half-bridge converter with switching devices. In the off-state each switch must support the rail voltage E_i and two floating gate drive circuits are required. This bridge converter is usually reserved for high-power applications.

Using similar analysis as for the push-pull converter in 17.9.1, the voltage transfer function of the full bridge with a forward converter output stage, with continuous inductor conduction is given by

$$\frac{v_o}{E_i} = \frac{\bar{I}_i}{\bar{I}_o} = 2N\frac{t_T}{\tau} = 2N\delta \qquad 0 \le \delta \le ½ \qquad (17.167)$$

Any volts-second imbalance (magnetising flux build-up) can be minimised by using dc blocking capacitance C_c, as shown in figures 17.19b and c.

The output ripple voltage is given by

$$\frac{\Delta v_C}{v_o} = \frac{\Delta v_o}{v_o} = \frac{(1-2\delta)\tau^2}{32LC} \qquad (17.168)$$

Output stage variations

In each forward converter in figure 17.19, a single secondary transformer winding and full-wave rectifier can be used. Better copper utilisation results. If the output diode shown dashed in figure 17.19c is used, the off state loop voltage is decreased from two diode voltage drops to one. The core magnetising current conducts through the secondary winding into the load circuit.
The three converters in figure 17.19 all employ the same forward converter output stage, so the critical load resistance for continuous inductor current is the same for each case, viz.,

$$R_{crit} = \frac{4L}{\tau(1-2\delta)} \qquad (17.169)$$

Re-arrangement of this equation gives an expression for minimum inductance in terms of the load resistance.
If the output inductor is not used, conventional unregulated transformer square-wave voltage ratio action occurs for each transformer based smps, where, independent of δ:

$$\frac{v_o}{E_i} = \frac{\bar{I}_i}{\bar{I}_o} = \frac{n_s}{n_o} = N \qquad (17.170)$$

Zero voltage switching (ZVS) of the H-bridge semiconductors

The H-bridge load circuit in figure 17.19 parts b and c, is a transformer, and all transformers have leakage inductance. This leakage inductance can be utilised as a turn-on snubber, producing H-bridge zero voltage switching ZVS conditions, which eliminate both switch turn-on losses and diode reverse recovery current injection problems. A consequence of ZVS is purely capacitive snubbers (no snubber diode or reset resistor) also become lossless.
The sequence of circuit diagrams in figure 17.21 illustrate how the transformer leakage inductance is used to achieve ZVS.
When any switch that is conducting current is turned off, current associated with the leakage inductance diverts to a diode, as shown in the off-loops in figures 17.21 parts b, c, and d. The switch in anti-parallel with that conducting diode in figure 17.21 can be turned on, while the diode conducts, without any switch turn-on losses, ZVS. The magnetising current circulates in a zero volt loop created in the secondary, as shown in figure 17.21. The zero volt loops, figures 17.21 b and c, are alternated on a cycle-by-cycle basis. At a maximum duty cycle, the negative voltage sequence in figure 17.21d is used, where the leakage inductance current falls rapidly to zero.
An inherent consequence of ZVS is that lossless capacitive turn-off snubbers can be employed across the bridge switches, as highlighted in chapter 18.1.1ii. The snubber capacitance can be optimally designed if the converter is operated in a constant current mode.

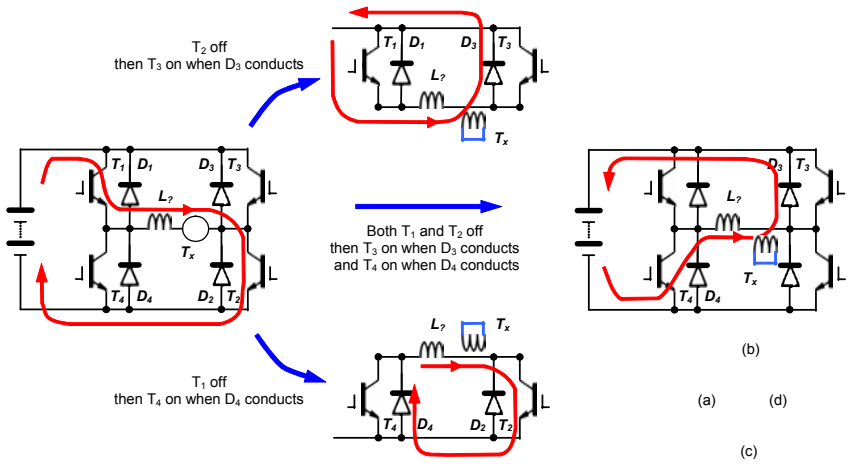


Figure 17.21. *H-bridge current conduction paths:*
(a) switches T_1 and T_2 conducting; (b) switch T_2 off and then T_3 on;
(c) switch T_1 off and then T_4 on; and (d) switches T_1 and T_2 off, then T_3 and T_4 on.

17.10 Basic generic smps transfer function mapping

The three basic smps, viz., the buck, boost and buck-boost converters, utilise a switch, diode and inductor, as shown in figure 17.22a, to perform their fundamental dc to dc conversion function.
Figure 17.22b shows a general form of the circuit in figure 17.22a, where the function of the two switching elements have not be prejudged to be a diode and a unidirectional voltage and current switch.
If the switch T_1 in the configuration of the circuit in figure 17.22a is controlled with an on-state duty cycle of δ, then the transfer functions associated with the buck, boost and buck boost converters are realised. Although each transfer function is fixed, the output function can be modified by mapping the input parameter. For example, if the complement of the duty cycle δ is used to control T_1, namely 1-δ, then in the case of the buck converter, the output voltage tracks 1-δ. The mapped transfer functions of the three basic converters, when controlled by the duty cycle complement 1-δ, are shown in table 17.4 and are plotted in figure 17.23. Practically, the same result is obtained if switch T_2 in the generalised case in figure 17.22 part b and part c is controlled by δ and switch T_1 is controlled by the complement, 1- δ.

Generally, if the duty cycle is encoded by $f(\delta)$, any effective transfer function can be generated within the voltage range of the basic converter. For example, in the case of the buck converter, any monotonically increasing output voltage profile can be produced in the range between zero volts and the input voltage magnitude. A lookup table mapping approach provides total flexibility.

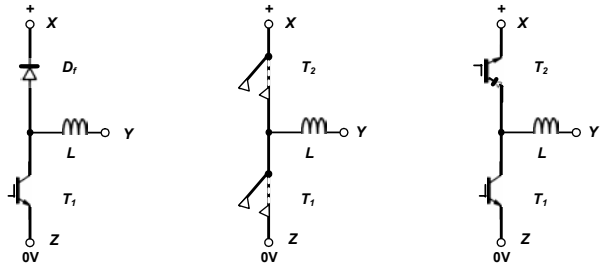


Figure 17.22. *Circuit elements of basic smps:*
(a) circuit diagram; (b) generalised functional circuit; and (c) specific circuit components.

Table 17.4: Mapped transfer functions

duty cycle mapping	$0<\delta<1$	$0<1-\delta<1$	$0<f(\delta)<1$
buck	δ	$1-\delta$	$1-f(\delta)$
boost	$\frac{1}{1-\delta}$	$\frac{1}{\delta}$	$\frac{1}{1-f(\delta)}$
buck-boost	$-\frac{\delta}{1-\delta}$	$-\frac{1-\delta}{\delta}$	$-\frac{f(\delta)}{1-f(\delta)}$

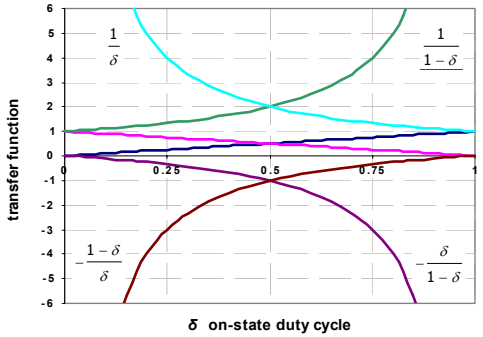


Figure 17.23. *The transfer functions of the three basic converters in terms of δ and their complementary transfer functions in terms of 1- δ.*

17.11 Appendix: Analysis of non-continuous inductor current operation

Operation with constant input voltage, E_i

In applications were the input voltage E_i is fixed, as with rectifier ac voltage input circuits and battery supplies, the output voltage v_o can be controlled by varying the duty cycle.

In the continuous inductor conduction region, the transfer function for the three basic converters is determined solely in terms of the on-state duty cycle, δ. Operation in the discontinuous inductor current region, for a constant input voltage, can be characterised for each converter in terms of duty cycle and the normalised output or input current, as shown in figure 17.28. Key region and boundary equations, for a constant input voltage E_i, are summarised in tables 17.5 and 17.6.

Operation with constant output voltage, v_o

In applications were the output voltage v_o is fixed, as required with regulated dc power supplies, the effects of varying input voltage E_i can be controlled and compensated by varying the duty cycle.

In the inductor continuous current conduction region, the transfer function is determined solely in terms of the on-state duty cycle, δ. Operation in the discontinuous region, for a constant output voltage, can be characterised in terms of duty cycle and the normalised output or input current, as shown in figure 17.29. Key region and boundary equations, for a constant output voltage v_o, are summarised in tables 17.7 and 17.8.

Because of the invariance of power, the output current $\bar{I}_o$ characteristics for each converter with a constant input voltage E_i , shown in figure 17.28, are the same as those for the input current $\bar{I}_i$ when the output voltage v_o is maintained constant, as shown in figure 17.29. [That is the right hand side of each plot in figures 17.28 and 17.29 (or figures 17.24 and 17.27) are the same.]

Generalised characteristics, with operating condition k (=$R\tau/L$), for the three basic converters, are summarised in Table 17.9. The associated monographs in figures 17.30, 17.31, and 17.32, with a specific load condition, k, for each converter, yield the inductor current waveforms for any on-state duty cycle δ. The three graphs illustrate operational boundaries between continuous inductor current at high δ and discontinuous inductor current at lower δ.

The graphs for the boost converter in figure 17.31 highlight a little appreciated feature that, if k > 13½, then discontinuous inductor current having appeared, disappears at lower and higher duty cycles. Specifically, continuous inductor current occurs for low duty cycles, where the same theoretical equation is interpreted to the contrary. That is, from table 17.2, the roots of

$$\delta(1-\delta)^2 \le \frac{2}{k} \qquad (17.171)$$

are not interpreted correctly. The correct interpretation of δ and k (= $R\tau/L$) gives:

- for k < 13½, discontinuous inductor current never occurs, independent of δ (equation (17.171) has two imaginary roots)
- for k = 13½, discontinuous inductor current occurs at only δ = ⅓ (equation (17.171) has three roots, two of which are coincident at δ = ⅓)
- for k > 13½, discontinuous inductor current occurs for δ around ⅓ as given by the two (of the three) real roots of equation (17.171) associated with the local minimum turning point of the cubic equation (17.171).

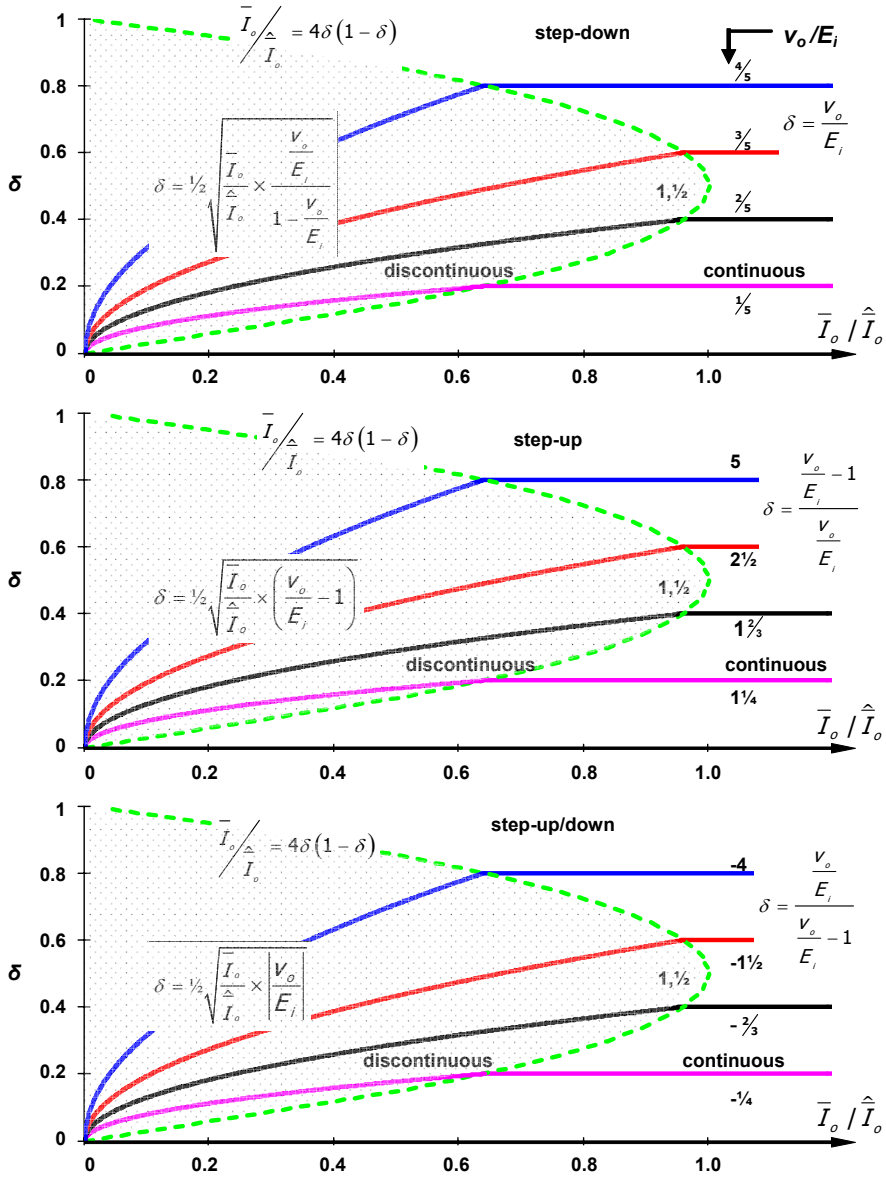


Figure 17.24. *Characteristics for three dc-dc converters with respect to $\bar{I}_o$, when the input voltage E_i is held constant. See table 17.5.*

Table 17.5: Transfer functions with constant input voltage, E_i, with respect to $\bar{I}_o$

E_i constant	converter		
	step-down	step-up	step-up/down
reference equation	(17.4)	(17.48)	(17.78)
continuous inductor current conduction (and change of variable)	$\frac{v_o}{E_i}=\delta$	$\frac{v_o}{E_i}=\frac{1}{1-\delta}$	$\frac{v_o}{E_i}=\frac{-\delta}{1-\delta}$
	$\delta=\frac{v_o}{E_i}$	$\delta=\frac{\frac{v_o}{E_i}-1}{\frac{v_o}{E_i}}$	$\delta=\frac{\frac{v_o}{E_i}}{\frac{v_o}{E_i}-1}$
reference equation	(17.21)	(17.63)	(17.94)
discontinuous inductor current conduction	$\frac{v_o}{E_i}=\frac{1}{1+\frac{2L\bar{I}_o}{\delta^2\tau E_i}}$	$\frac{v_o}{E_i}=1+\frac{\delta^2 E_i\tau}{2L\bar{I}_o}$	$\frac{v_o}{E_i}=-\frac{\delta^2 E_i\tau}{2L\bar{I}_o}$
normalised $\frac{v_o}{E_i}=$ where $\hat{\bar{I}}_o=\frac{E_i\tau}{8L}$	$\frac{v_o}{E_i}=\frac{1}{1+\frac{1}{4\delta^2}\times\frac{\bar{I}_o}{\hat{\bar{I}}_o}}$	$\frac{v_o}{E_i}=1+4\delta^2/\frac{\bar{I}_o}{\hat{\bar{I}}_o}$	$\frac{v_o}{E_i}=-4\delta^2/\frac{\bar{I}_o}{\hat{\bar{I}}_o}$
$\bar{I}_o=\hat{\bar{I}}_o=1$pu @	$\delta=½;\ \frac{v_o}{E_i}=½$	$\delta=½;\ \frac{v_o}{E_i}=2$	$\delta=½;\ \frac{v_o}{E_i}=-1$
change of variable $\frac{\bar{I}_o}{\hat{\bar{I}}_o}=$	$\frac{\bar{I}_o}{\hat{\bar{I}}_o}=4\delta^2\times\frac{\left(1-\frac{v_o}{E_i}\right)}{\frac{v_o}{E_i}}$	$\frac{\bar{I}_o}{\hat{\bar{I}}_o}=4\delta^2\times\frac{1}{\frac{v_o}{E_i}-1}$	$\frac{\bar{I}_o}{\hat{\bar{I}}_o}=-4\delta^2\times\frac{1}{\frac{v_o}{E_i}}$
change of variable $\delta=$ all with a boundary $\delta=½+½\sqrt{1-\frac{\bar{I}_o}{\hat{\bar{I}}_o}}$	$\delta=½\sqrt{\frac{\bar{I}_o}{\hat{\bar{I}}_o}\times\frac{\frac{v_o}{E_i}}{1-\frac{v_o}{E_i}}}$	$\delta=½\sqrt{\frac{\bar{I}_o}{\hat{\bar{I}}_o}\times\left(\frac{v_o}{E_i}-1\right)}$	$\delta=½\sqrt{\frac{\bar{I}_o}{\hat{\bar{I}}_o}\times\left\lvert\frac{v_o}{E_i}\right\rvert}$
conduction boundary $\delta=½+½\sqrt{1-\frac{\bar{I}_o}{\hat{\bar{I}}_o}}$	$\frac{\bar{I}_o}{\hat{\bar{I}}_o}=4\times\frac{v_o}{E_i}\left(1-\frac{v_o}{E_i}\right)$ $=4\delta(1-\delta)$	$\frac{\bar{I}_o}{\hat{\bar{I}}_o}=4\times\frac{\left(\frac{v_o}{E_i}-1\right)}{\left(\frac{v_o}{E_i}\right)^2}$ $=4\delta(1-\delta)$	$\frac{\bar{I}_o}{\hat{\bar{I}}_o}=-4\times\frac{\frac{v_o}{E_i}}{\left(1-\frac{v_o}{E_i}\right)^2}$ $=4\delta(1-\delta)$

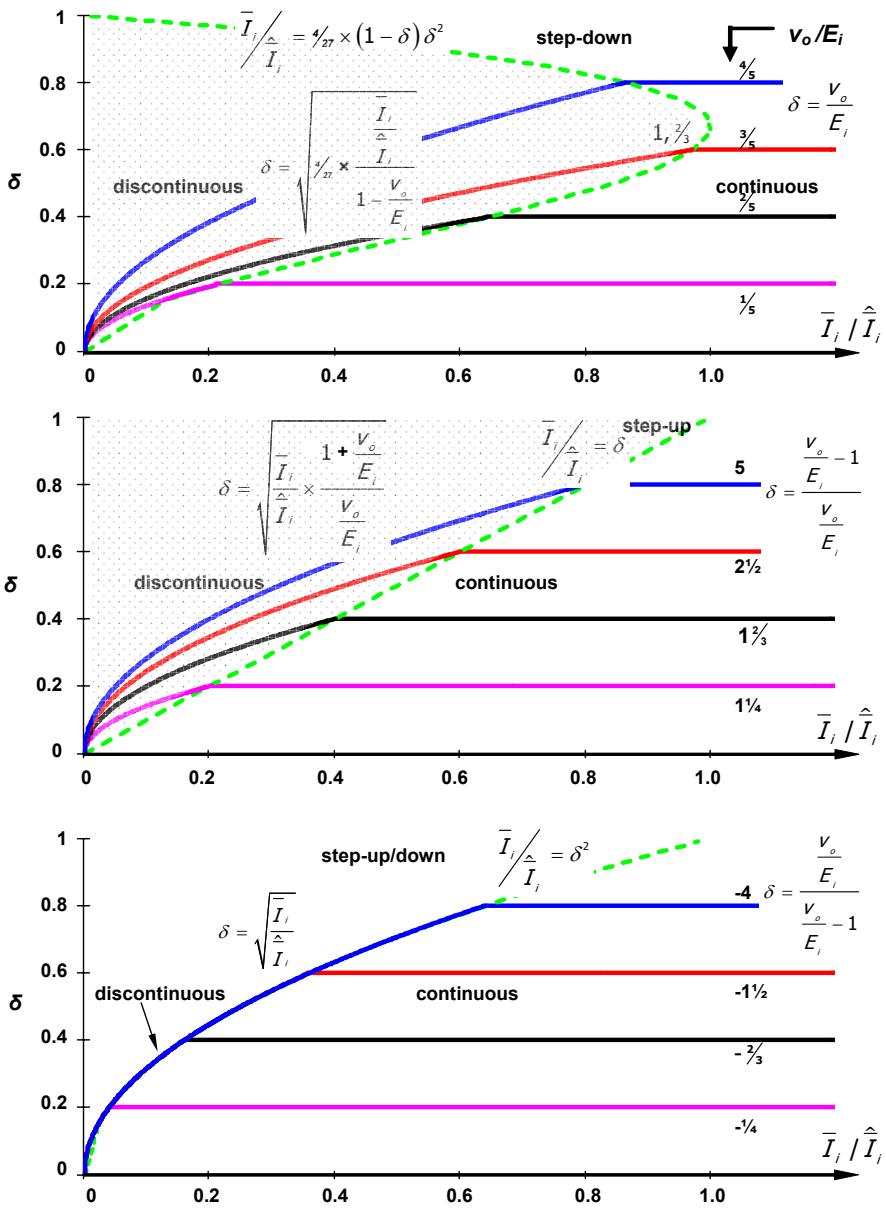


Figure 17.25. *Characteristics for three dc-dc converters with respect to $\bar{I}_i$, when the input voltage E_i is held constant. See table 17.6.*

Table 17.6: Transfer functions with constant input voltage, E_i, with respect to $\bar{I}_i$

E_i constant	converter		
	step-down	step-up	step-up/down
reference equation	(17.4)	(17.48)	(17.78)
continuous inductor current conduction (and change of variable)	$\frac{v_o}{E_i}=\delta$	$\frac{v_o}{E_i}=\frac{1}{1-\delta}$	$\frac{v_o}{E_i}=\frac{-\delta}{1-\delta}$
	$\delta=\frac{v_o}{E_i}$	$\delta=\frac{\frac{v_o}{E_i}-1}{\frac{v_o}{E_i}}$	$\delta=\frac{\frac{v_o}{E_i}}{\frac{v_o}{E_i}-1}$
reference equation	(17.20)	(17.64)	(17.94)
discontinuous inductor current conduction	$\frac{v_o}{E_i}=1-\frac{2L\bar{I}_i}{\delta^2\tau E_i}$	$\frac{v_o}{E_i}=\frac{1}{1-\frac{E_i\tau\delta^2}{2L\bar{I}_i}}$	$\frac{v_o}{E_i}=\frac{v_o\tau\delta^2}{2L\bar{I}_i}$
normalised $\frac{v_o}{E_i}=$	$\frac{v_o}{E_i}=1-\frac{4}{27\delta^2}\times\frac{\bar{I}_i}{\hat{\bar{I}}_i}$ where $\hat{\bar{I}}_i=\frac{4}{27}\times\frac{E_i\tau}{2L}$	$\frac{v_o}{E_i}=\frac{1}{1-\delta^2/\left(\frac{\bar{I}_i}{\hat{\bar{I}}_i}\right)}$ where $\hat{\bar{I}}_i=\frac{E_i\tau}{2L}$	$1=\delta^2/\frac{\bar{I}_i}{\hat{\bar{I}}_i}$ where $\hat{\bar{I}}_i=\frac{E_i\tau}{2L}$
$\bar{I}_i=\hat{\bar{I}}_i=1$pu @	$\delta=⅔$; $\frac{v_o}{E_i}=⅔$	$\delta=1$; $\frac{v_o}{E_i}\to\infty$	$\delta=1$; $\frac{v_o}{E_i}\to-\infty$
change of variable $\frac{\bar{I}_i}{\hat{\bar{I}}_i}=$	$\frac{\bar{I}_i}{\hat{\bar{I}}_i}=\frac{27}{4}\delta^2\left(1-\frac{v_o}{E_i}\right)$	$\frac{\bar{I}_i}{\hat{\bar{I}}_i}=\delta^2\times\frac{\frac{v_o}{E_i}}{\left(\frac{v_o}{E_i}-1\right)}$	$\frac{\bar{I}_i}{\hat{\bar{I}}_i}=\delta^2$
change of variable $\delta=$	$\delta=\sqrt{\frac{4}{27}\times\frac{\bar{I}_i}{\hat{\bar{I}}_i}\times\frac{1}{1-\frac{v_o}{E_i}}}$	$\delta=\sqrt{\frac{\bar{I}_i}{\hat{\bar{I}}_i}\times\frac{\frac{v_o}{E_i}-1}{\frac{v_o}{E_i}}}$	$\delta=\sqrt{\frac{\bar{I}_i}{\hat{\bar{I}}_i}}$
conduction boundary	$\frac{\bar{I}_i}{\hat{\bar{I}}_i}=\frac{27}{4}\times\left(1-\frac{v_o}{E_i}\right)\left(\frac{v_o}{E_i}\right)^2=\frac{27}{4}\delta^2(1-\delta)$	$\frac{\bar{I}_i}{\hat{\bar{I}}_i}=\frac{\left(\frac{v_o}{E_i}-1\right)}{\frac{v_o}{E_i}}=\delta$	$\frac{\bar{I}_i}{\hat{\bar{I}}_i}=\left(\frac{\frac{v_o}{E_i}}{\frac{v_o}{E_i}-1}\right)^2=\delta^2$
conduction boundary	$\frac{\bar{I}_i}{\hat{\bar{I}}_i}=\frac{27}{4}\delta^2(1-\delta)$	$\delta=\frac{\bar{I}_i}{\hat{\bar{I}}_i}$	$\delta=\sqrt{\frac{\bar{I}_i}{\hat{\bar{I}}_i}}$

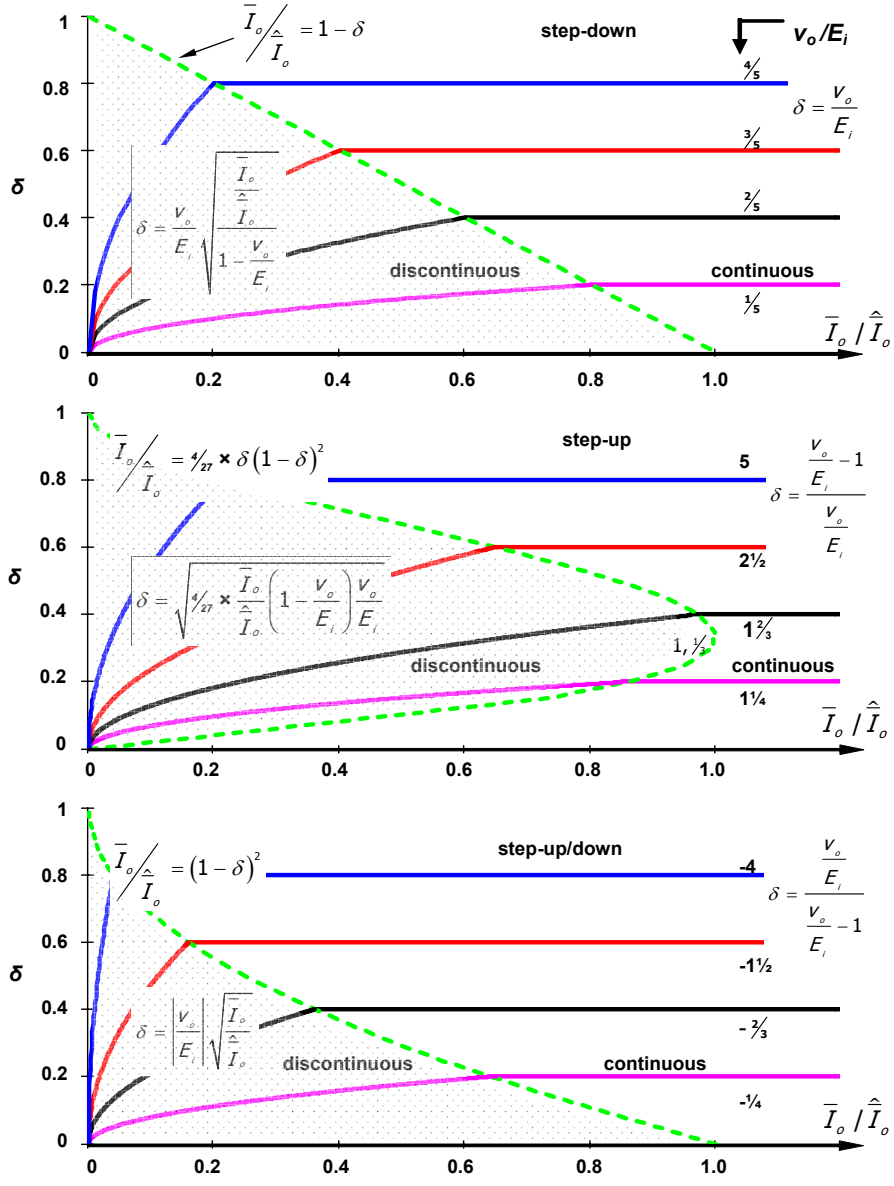


Figure 17.26. *Characteristics for three dc-dc converters with respect to $\bar{I}_o$, when the output voltage v_o is held constant. See table 17.7.*

Table 17.7: Transfer functions with constant output voltage, v_o, with respect to $\bar{I}_o$

v_o constant	converter		
	step-down	step-up	step-up/down
reference equation	(17.4)	(17.48)	(17.78)
continuous inductor current conduction (and change of variable)	$\frac{v_o}{E_i}=\delta$	$\frac{v_o}{E_i}=\frac{1}{1-\delta}$	$\frac{v_o}{E_i}=\frac{-\delta}{1-\delta}$
	$\delta=\frac{v_o}{E_i}$	$\delta=\frac{\frac{v_o}{E_i}-1}{\frac{v_o}{E_i}}$	$\delta=\frac{\frac{v_o}{E_i}}{\frac{v_o}{E_i}-1}$
reference equation	(17.20)	(17.64)	(17.94)
discontinuous inductor current conduction	$\frac{v_o}{E_i}=1-\frac{2L\bar{I}_i}{\delta^2\tau E_i}$	$\frac{v_o}{E_i}=\frac{1}{1-\frac{E_i\tau\,\delta^2}{2L\bar{I}_i}}$	$\frac{v_o}{E_i}=\frac{v_o\tau\delta^2}{2L\bar{I}_i}$
normalised $\frac{v_o}{E_i}=$	$\frac{v_o}{E_i}=1-\frac{1}{4\delta^2}\times\frac{\bar{I}_o}{\hat{I}_o}\times\left(\frac{v_o}{E_i}\right)^2$ where $\hat{I}_o=\frac{v_o\tau}{2L}$	$\frac{v_o}{E_i}=\frac{1}{1-27/4\,\delta^2/\left(\frac{\bar{I}_o}{\hat{I}_o}\times\left(\frac{v_o}{E_i}\right)^2\right)}$ where $\hat{I}_o=\frac{4}{27}\times\frac{v_o\tau}{2L}$	$\frac{v_o}{E_i}=\delta^2/\left(\frac{\bar{I}_o}{\hat{I}_o}\times\frac{v_o}{E_i}\right)$ where $\hat{I}_o=\left\|\frac{v_o\tau}{2L}\right\|$
$\bar{I}_o=\hat{I}_o=1$pu @	δ = 0; $\frac{v_o}{E_i}=0$	δ = ⅓; $\frac{v_o}{E_i}=1½$	δ = 0; $\frac{v_o}{E_i}=0$
change of variable $\frac{\bar{I}_o}{\hat{I}_o}=$	$\frac{\bar{I}_o}{\hat{I}_o}=\delta^2\times\frac{\left(1-\frac{v_o}{E_i}\right)}{\left(\frac{v_o}{E_i}\right)^2}$	$\frac{\bar{I}_o}{\hat{I}_o}=27/4\,\delta^2\times\frac{1}{\left(\frac{v_o}{E_i}-1\right)\frac{v_o}{E_i}}$	$\frac{\bar{I}_o}{\hat{I}_o}=\delta^2\times\frac{1}{\left(\frac{v_o}{E_i}\right)^2}$
change of variable $\delta=$	$\delta=\frac{v_o}{E_i}\sqrt{\frac{\bar{I}_o}{\hat{I}_o}\times\frac{1}{1-\frac{v_o}{E_i}}}$	$\delta=\sqrt{4/27\times\frac{\bar{I}_o}{\hat{I}_o}\times\left(\frac{v_o}{E_i}-1\right)\frac{v_o}{E_i}}$	$\delta=\left\|\frac{v_o}{E_i}\right\|\sqrt{\frac{\bar{I}_o}{\hat{I}_o}}$
conduction boundary	$\frac{\bar{I}_o}{\hat{I}_o}=1-\frac{v_o}{E_i}=1-\delta$	$\frac{\bar{I}_o}{\hat{I}_o}=27/4\times\frac{\left(\frac{v_o}{E_i}-1\right)}{\left(\frac{v_o}{E_i}\right)^3}=27/4\,\delta(1-\delta)^2$	$\frac{\bar{I}_o}{\hat{I}_o}=\frac{1}{\left(1-\frac{v_o}{E_i}\right)^2}=(1-\delta)^2$
conduction boundary	$\delta=1-\frac{\bar{I}_o}{\hat{I}_o}$	$\frac{\bar{I}_o}{\hat{I}_o}=27/4\,\delta(1-\delta)^2$	$\delta=1-\sqrt{\frac{\bar{I}_o}{\hat{I}_o}}$

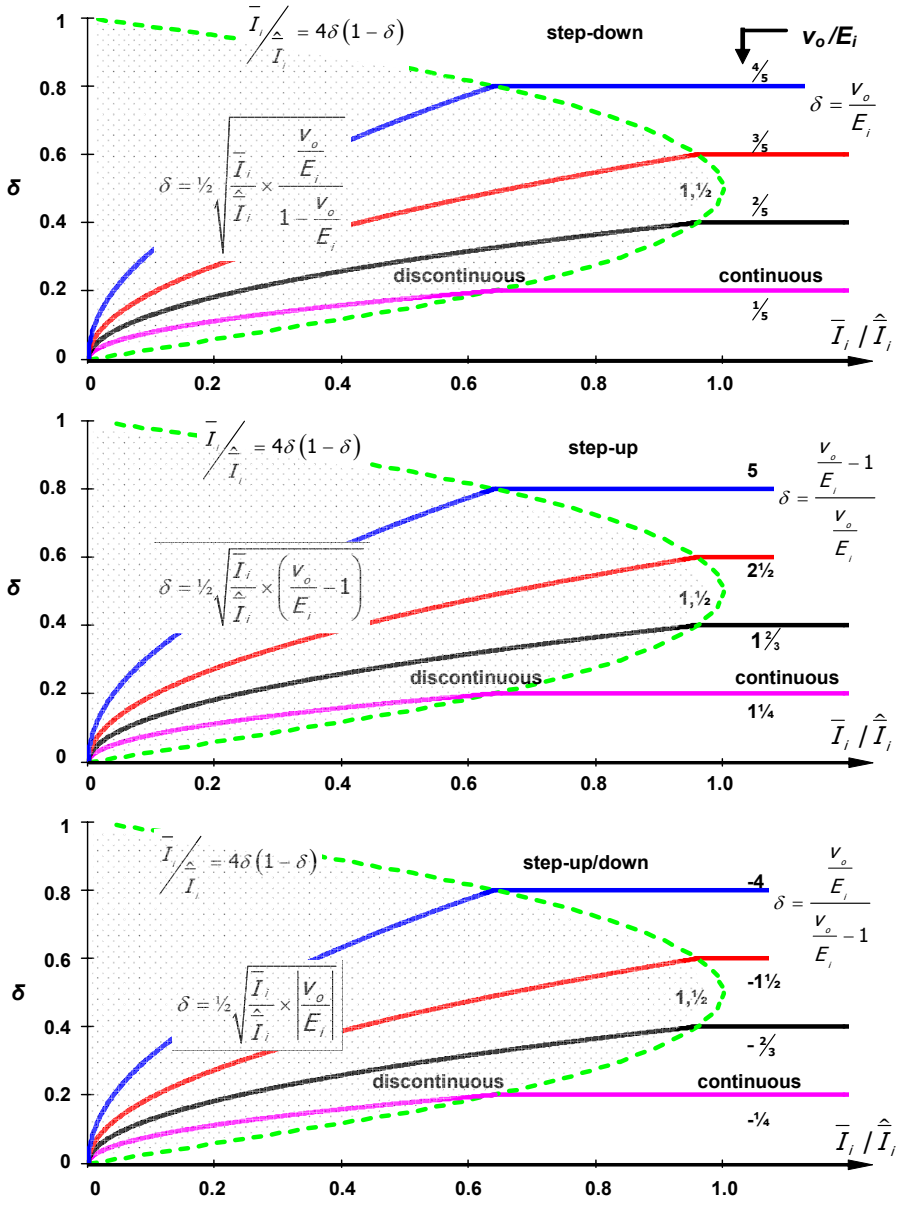


Figure 17.27. *Characteristics for three dc-dc converters with respect to $\bar{I}_i$, when the output voltage v_o is held constant. See table 17.8.*

Table 17.8: Transfer functions with constant input voltage, v_o, with respect to $\bar{I}_i$

v_o constant	converter		
	step-down	step-up	step-up/down
reference equation	(17.4)	(17.48)	(17.78)
continuous inductor current conduction (and change of variable)	$\frac{v_o}{E_i}=\delta$	$\frac{v_o}{E_i}=\frac{1}{1-\delta}$	$\frac{v_o}{E_i}=\frac{-\delta}{1-\delta}$
	$\delta=\frac{v_o}{E_i}$	$\delta=\frac{\frac{v_o}{E_i}-1}{\frac{v_o}{E_i}}$	$\delta=\frac{\frac{v_o}{E_i}}{\frac{v_o}{E_i}-1}$
reference equation	(17.21)	(17.63)	(17.94)
discontinuous inductor current conduction	$\frac{v_o}{E_i}=\frac{1}{1+\frac{2L\bar{I}_i}{\delta^2\tau v_o}}$	$\frac{v_o}{E_i}=1+\frac{\delta^2 v_o\tau}{2L\bar{I}_i}$	$\frac{v_o}{E_i}=-\frac{\delta^2 v_o\tau}{2L\bar{I}_i}$
normalised $\frac{v_o}{E_i}=$ where $\hat{\bar{I}}_i=\left\lvert\frac{v_o\tau}{8L}\right\rvert$	$\frac{v_o}{E_i}=\frac{1}{1+\frac{1}{4\delta^2}\times\frac{\bar{I}_i}{\hat{\bar{I}}_i}}$	$\frac{v_o}{E_i}=1+4\delta^2/\frac{\bar{I}_i}{\hat{\bar{I}}_i}$	$\frac{v_o}{E_i}=-4\delta^2/\frac{\bar{I}_i}{\hat{\bar{I}}_i}$
$\bar{I}_i=\hat{\bar{I}}_i=$ 1pu @	$\delta=½;\ \frac{v_o}{E_i}=½$	$\delta=½;\ \frac{v_o}{E_i}=2$	$\delta=½;\ \frac{v_o}{E_i}=-1$
change of variable $\frac{\bar{I}_i}{\hat{\bar{I}}_i}=$	$\frac{\bar{I}_i}{\hat{\bar{I}}_i}=4\delta^2\times\frac{\left(1-\frac{v_o}{E_i}\right)}{\frac{v_o}{E_i}}$	$\frac{\bar{I}_i}{\hat{\bar{I}}_i}=4\delta^2\times\frac{1}{\frac{v_o}{E_i}-1}$	$\frac{\bar{I}_i}{\hat{\bar{I}}_i}=-4\delta^2\times\frac{1}{\frac{v_o}{E_i}}$
change of variable $\delta=$ all with a boundary $\delta=½+½\sqrt{1-\frac{\bar{I}_o}{\hat{\bar{I}}_o}}$	$\delta=½\sqrt{\frac{\bar{I}_i}{\hat{\bar{I}}_i}\times\frac{\frac{v_o}{E_i}}{1-\frac{v_o}{E_i}}}$	$\delta=½\sqrt{\frac{\bar{I}_i}{\hat{\bar{I}}_i}\times\left(\frac{v_o}{E_i}-1\right)}$	$\delta=½\sqrt{\frac{\bar{I}_i}{\hat{\bar{I}}_i}\times\left\lvert\frac{v_o}{E_i}\right\rvert}$
conduction boundary $\delta=½+½\sqrt{1-\frac{\bar{I}_o}{\hat{\bar{I}}_o}}$	$\frac{\bar{I}_i}{\hat{\bar{I}}_i}=4\frac{v_o}{E_i}\left(1-\frac{v_o}{E_i}\right)$ $=4\delta(1-\delta)$	$\frac{\bar{I}_i}{\hat{\bar{I}}_i}=4\times\frac{\left(\frac{v_o}{E_i}-1\right)}{\left(\frac{v_o}{E_i}\right)^2}$ $=4\delta(1-\delta)$	$\frac{\bar{I}_i}{\hat{\bar{I}}_i}=-4\times\frac{\frac{v_o}{E_i}}{\left(1-\frac{v_o}{E_i}\right)^2}$ $=4\delta(1-\delta)$

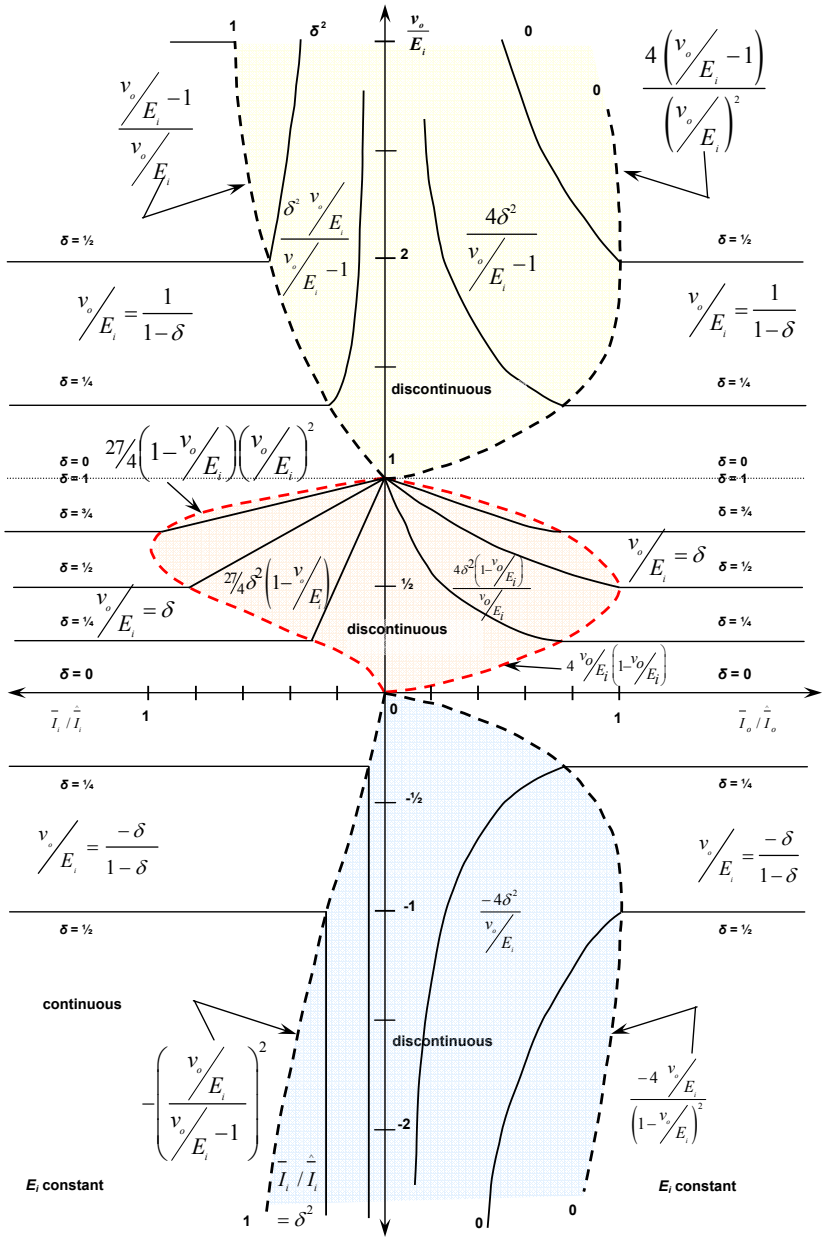


Figure 17.28. *Characteristics for three dc-dc converters, when the input voltage E_i is held constant.*

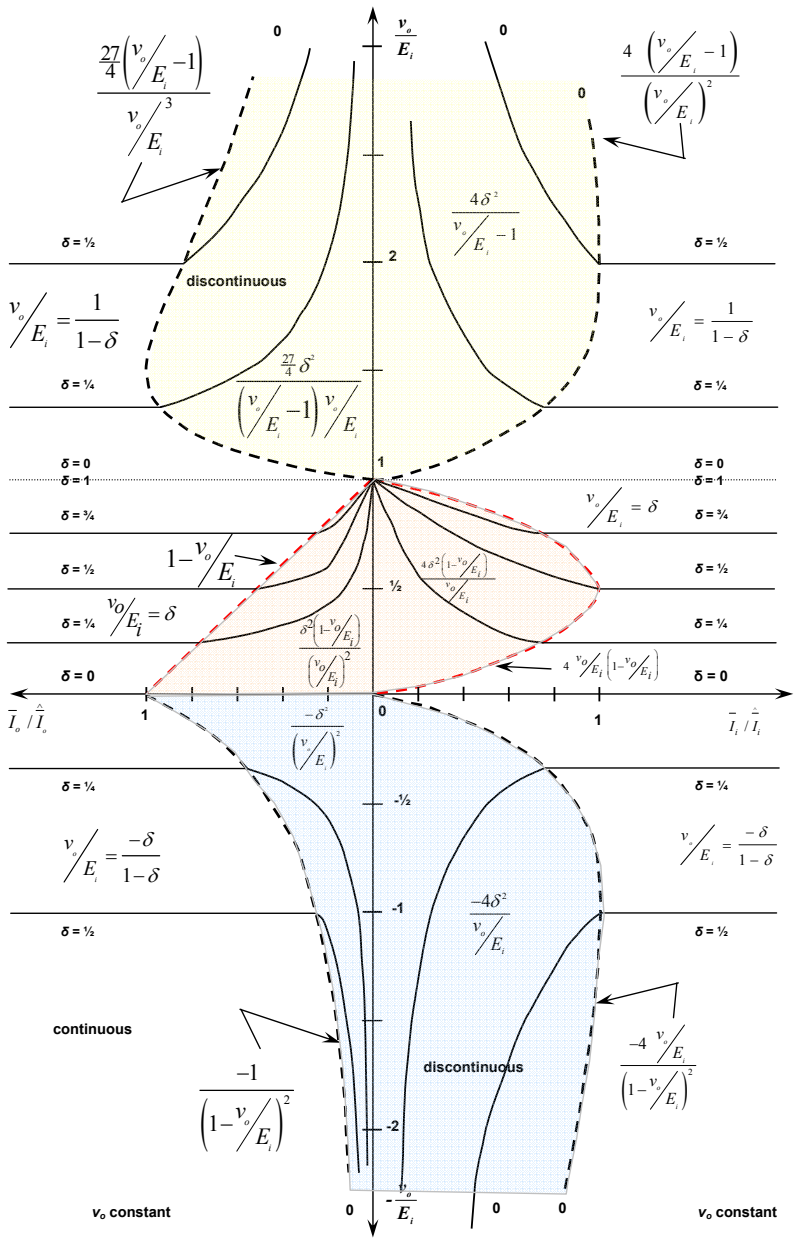


Figure 17.29. *Characteristics for three dc-dc converters, when the output voltage v_o is held constant.*

Table 17.9: Converter parameters for discontinuous and continuous inductor conduction regions and boundaries

$k = \frac{R\tau}{L}$; $0 \le \delta = \frac{t_T}{\tau} \le 1$	Converter	Forward step-down		Flyback step-up		Flyback step-up/down	
		discontinuous	*continuous*	*discontinuous*	*Continuous*	*discontinuous*	*continuous*
$\delta_{critical}(k) = k \ge \frac{2}{\delta(1-\delta)} \times \frac{v_o}{E_i}$		$\delta \le 1-\frac{2}{k}$ $k \ge \frac{2}{1-\delta}$	$\delta \ge 1-\frac{2}{k}$ $k \le \frac{2}{1-\delta}$	$k > 27/2$ $\delta(1-\delta)^2 \le 2/k$	$k \le 27/2$	$\delta \le 1-\sqrt{\frac{2}{k}}$ $k \ge \frac{2}{(1-\delta)^2}$	$\delta \ge 1-\sqrt{\frac{2}{k}}$ $k \le \frac{2}{(1-\delta)^2}$
$\frac{v_o}{E_i}(k,\delta) = \frac{\bar{I}_L}{\bar{I}_o} = \bar{I}_o \times \frac{R}{E_i}$		$\frac{1}{2}k\delta^2\left[-1+\sqrt{1+\frac{8}{k\delta^2}}\right]$	δ	$\frac{1}{2}\left[1+\sqrt{1+2k\delta^2}\right]$	$\frac{1}{1-\delta}$	$-\delta\sqrt{\frac{1}{2}k}$	$\frac{-\delta}{1-\delta}$
$\delta_D = \frac{t_D}{\tau}$		$1-\frac{v_o}{E_i}$ $\delta \times \frac{v_o}{E_i}$	$1-\delta$	$\delta \times \frac{1}{\frac{v_o}{E_i}-1}$	$1-\delta$	$\delta \times \frac{1}{\frac{v_o}{E_i}}$	$1-\delta$
$\delta_x = \frac{t_x}{\tau}$ $= 1-(\delta+\delta_D)$		$1-\delta \times \frac{1}{\frac{v_o}{E_i}}$	0	$\frac{\frac{v_o}{E_i}}{1-\delta \times \frac{v_o}{E_i}-1}$	0	$1+\frac{\left\|\frac{v_o}{E_i}\right\|}{1-\delta \times \frac{\left\|\frac{v_o}{E_i}\right\|}{\frac{v_o}{E_i}}}$	0
$\hat{i}_L \times \frac{R}{E_i}$, $\hat{i}_L \times \frac{R}{E_i}$, $k\delta\left[1-\frac{v_o}{E_i}\right], 0$		$\frac{v_o}{E_i} \pm \frac{1}{2}k\delta(1-\delta)$	$\frac{v_o}{E_i}$	$k\delta, 0$	$\frac{1}{1-\delta} \times \frac{v_o}{E_i} \pm \frac{1}{2}k\delta$	$k\delta, 0$	$\frac{1}{1-\delta} \times \left\|\frac{v_o}{E_i}\right\| \pm \frac{1}{2}k\delta$
$\bar{I}_L \times \frac{R}{E_i}$ $\bar{I}_L = \frac{1}{2}\left(\hat{i}_L + \check{i}_L\right)(\delta+\delta_D)$		$1-\frac{v_o}{E_i}$ $\frac{1}{2}k\delta^2 \times \frac{v_o}{E_i}$	$\frac{v_o}{E_i}$	$\frac{\frac{v_o}{E_i}}{\frac{1}{2}k\delta^2 \times \frac{v_o}{E_i}-1}$	$\frac{1}{1-\delta} \times \frac{v_o}{E_i}$	$\frac{1+\frac{\left\|\frac{v_o}{E_i}\right\|}{\frac{1}{2}k\delta^2 \times \frac{\left\|\frac{v_o}{E_i}\right\|}{\frac{v_o}{E_i}}}}{}$	$\frac{1}{1-\delta} \times \left\|\frac{v_o}{E_i}\right\|$
$I_c = 0$ $i_L(\delta \le t \le \delta+\delta_D) \le I_o$		$t/\tau\big\|_{t<t_T} = \delta / k\delta_D$ $t/\tau\big\|_{t_T<t\le\tau} = \delta_D - 1/k$	$t/\tau\big\|_{t<t_T} = \frac{1}{2}\delta$ $\forall\delta$ $t/\tau\big\|_{t_T<t\le\tau} = \frac{1}{2}(1-\delta)$	$t/\tau\big\|_{t<t_T} = \delta$ $\forall\delta$ $t/\tau = \delta_D\left(1-\frac{v_o/E_i}{k\delta}\right)$ $t_T < t \le \tau$	$t/\tau\big\|_{t<t_T} = \delta$ $\delta \le 1-\sqrt{\frac{2}{k}}$ $k \ge \frac{2}{1-\delta} \times \frac{v_o}{E_i}$ if $k \le 2$ then $\forall\delta$ $t/\tau\big\|_{t_T<t\le\tau} = \frac{1}{2}(1-\delta)+\frac{v_o}{E_i}/k$ $t_T < t \le \tau$	$t/\tau\big\|_{t<t_T} = \delta$ $\forall\delta$ $t/\tau = \delta_D\left(1-\frac{\left\|v_o/E_i\right\|}{k\delta}\right)$ $t_T < t \le \tau$	$t/\tau\big\|_{t<t_T} = \delta$ $\delta \le 1+\frac{1}{k}-\sqrt{\left(1+\frac{1}{k}\right)^2-1}$ $k \ge \frac{2}{1-\delta} \times \frac{\left\|\frac{v_o}{E_i}\right\|}{\frac{v_o}{E_i}}$ $t/\tau\big\|_{t_T<t\le\tau} = \frac{1}{2}(1-\delta)+\frac{\left\|\frac{v_o}{E_i}\right\|}{\frac{v_o}{E_i}}/k$ $t_T < t \le \tau$

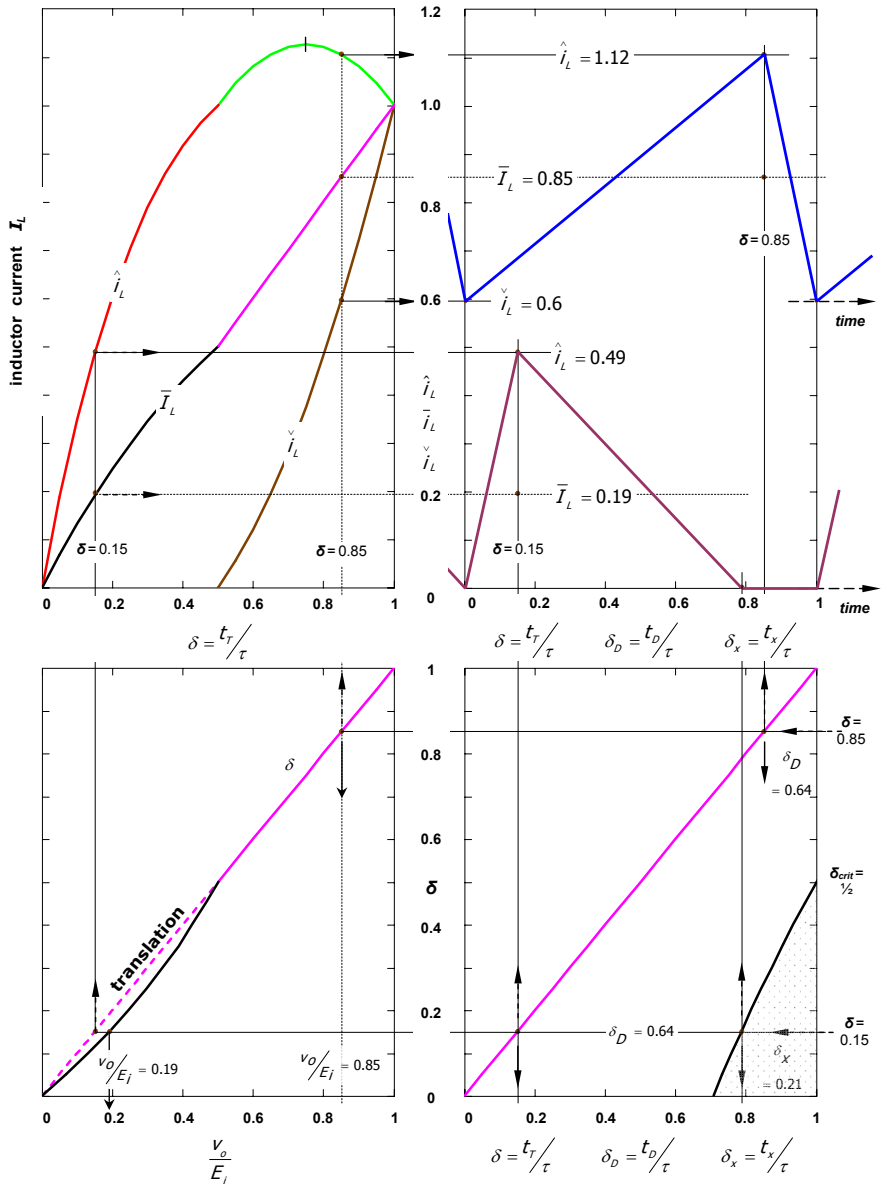


Figure 17.30. *Step-down converter normalised performance monogram for k = 4, giving discontinuous inductor conduction for $\delta_{crit} \le ½$. Inductor time domain current waveforms for $\delta_{cont} = 0.85$ (continuous inductor current) and $\delta_{dis} = 0.15$ (discontinuous inductor current).*

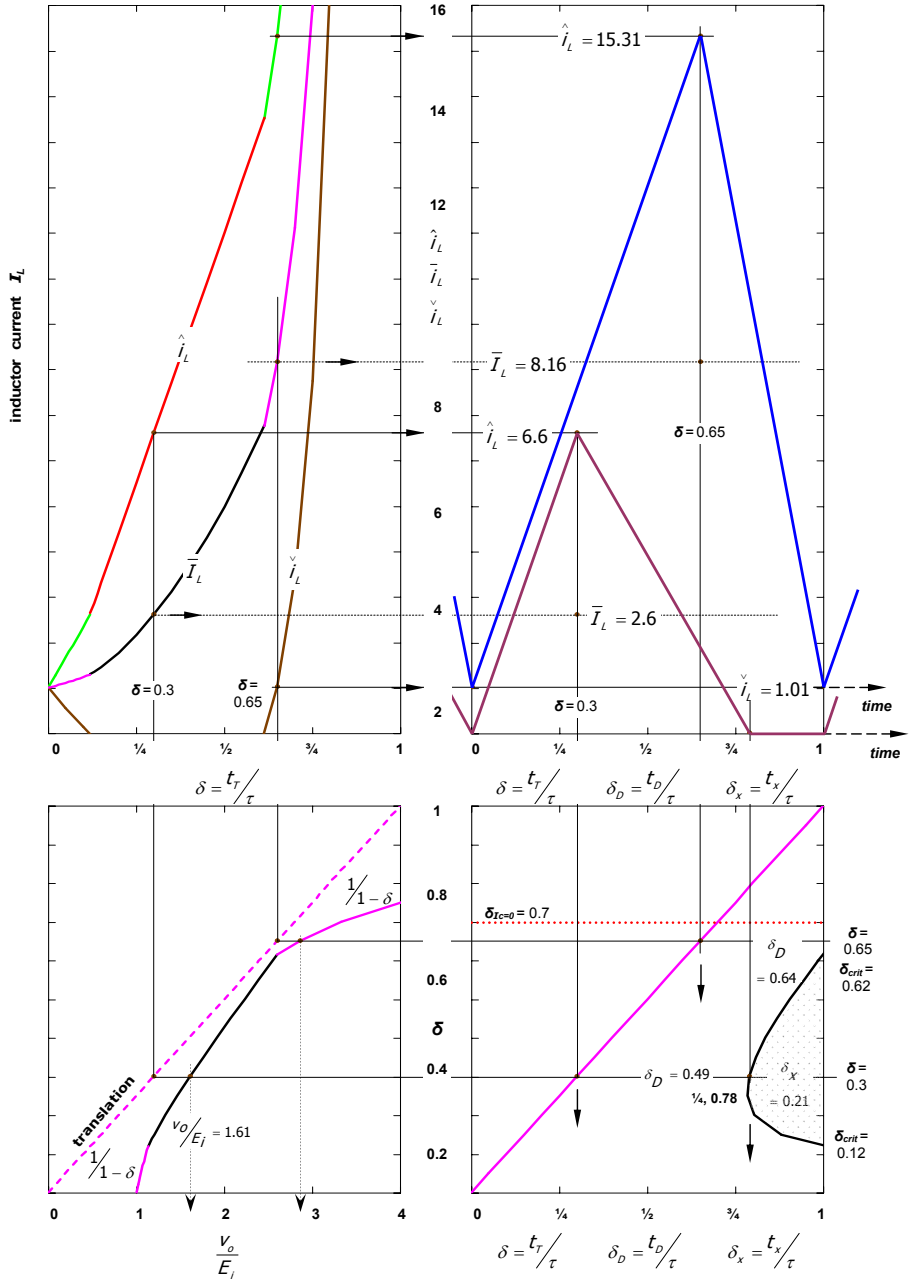


Figure 17.31. *Step-up converter performance monogram for k = 22, giving discontinuous inductor current for $0.12 \le \delta_{crit} \le 0.62$. Inductor time domain current waveforms for $\delta_{cont} = 0.65$ (continuous inductor current) and $\delta_{dis} = 0.3$ (discontinuous inductor current).*
Capacitor discharge in switch-off period when δ ≤ 0.7.

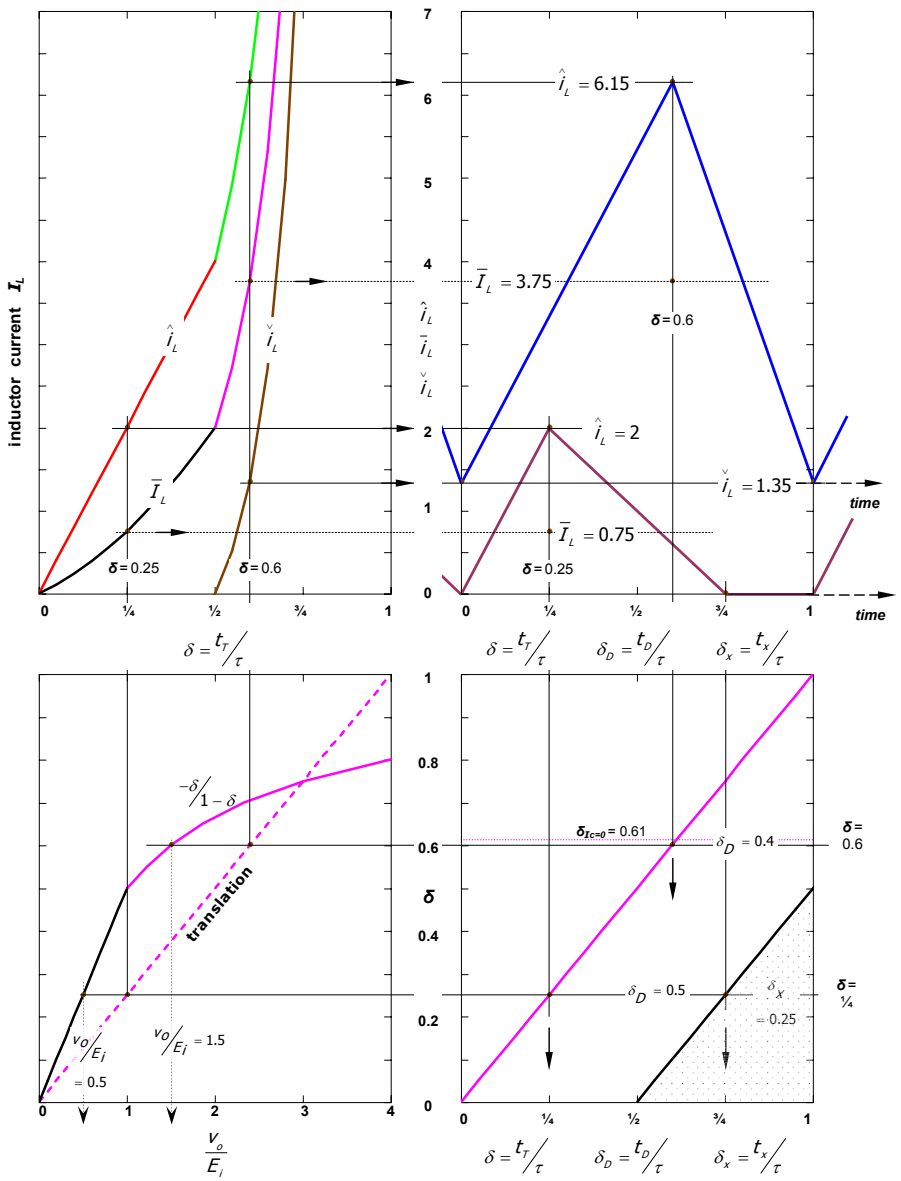


Figure 17.32. *Step-up/down converter normalised performance monogram for k = 8, giving discontinuous inductor current for $\delta_{crit} \le ½$. Inductor time domain current waveforms for $\delta_{cont} = 0.6$ (continuous inductor current) and $\delta_{dis} = ¼$ (discontinuous inductor current). Capacitor discharge in switch-off period when $\delta \le 0.61$.*

Reading list

Fisher, M. J., *Power Electronics*, PWS-Kent Publishing, 1991.

Hart, D.W., *Introduction to Power Electronics*, Prentice-Hall, Inc, 1994.

Hnatek, E. R., *Design of Switch Mode Power Supplies*, Van Nostrand Reinhold, 1981.

Mohan, N., *Power Electronics*, 3rd Edition, Wiley International, 2003.

Thorborg, K., *Power Electronics – in theory and practice*, Chartwell-Bratt, 1993.

http://www.ipes.ethz.ch/

Problems

17.1. An smps is used to provide a 5V rail at 2.5A. If 100 mV p-p output ripple is allowed and the input voltage is 12V with 25 per cent tolerance, design a flyback buck-boost converter which has a maximum switching frequency of 50 kHz.

17.2. Derive the following design equations for a flyback boost converter, which operates in the discontinuous mode.

$$\hat{i}_i = 2 \times \bar{I}_{o\,(\max)} \times \frac{v_o}{E_{i(\min)}} = \text{ constant} \qquad t_D = \frac{1}{f_{(\max)} \dfrac{v_o}{E_{i(\min)}}}$$

$$L = t_{T(\min)} \frac{v_o - E_{i(\min)}}{\hat{i}_i} \qquad f = \frac{1}{\tau} = f_{(\max)} \frac{\bar{I}_o}{\bar{I}_{o\,(\max)}} \times \frac{v_o - E_i}{v_o - E_{i(\min)}}$$

$$\check{C} = \frac{\Delta Q}{\Delta e_o} = \frac{\hat{i}_i\, t_{T(\min)}}{2\Delta e_o} \qquad ESR_{(\max)} = \frac{\Delta e_o}{\hat{i}_i}$$

17.3. Derive design equations for the forward non-isolated converter, operating in the continuous conduction mode.

17.4. Prove that the output rms ripple current for the forward converter in figure 17.2 is given by $\Delta i_o / 2\sqrt{3}$.

17.5. If the smps inductor has series resistance r, show that the voltage transfer function of the boost converter, with continuous inductor current, is given by

$$\frac{v_o}{E_i} = \frac{1}{1-\delta} \times \frac{1}{1 + \dfrac{r}{R(1-\delta)^2}}$$

where *R* is the load resistance. Hence show that the power transfer efficiency is

$$\eta = \frac{1}{1 + \dfrac{r}{R(1-\delta)^2}}$$

17.6. Show that the output voltage of a forward converter is decreased by $\delta v_{sw} + (1-\delta) v_D$ when the switch voltage drop is v_{sw} and the diode forward voltage drop is v_D.

17.7. In the forward converter example 17.1, the load resistance is varied between 1Ω and 16Ω, over which range the inductor current becomes discontinuous. With the aid of table 17.2, plot the output voltage as a function of load resistance over the range 1Ω to 16Ω.

17.8. In the step-up converter example 17.3, the load resistance is varied between 2½Ω and 22½Ω, and the inductor current becomes discontinuous at 22½Ω. With the aid of table 17.2, plot the output voltage as a function of load resistance over the range 2.5Ω to 45Ω.

17.9. In the step-up converter example 17.4, the load resistance is varied between 2½Ω and 37½Ω, and the inductor current becomes discontinuous at 37½Ω. With the aid of table 17.2, plot the output voltage as a function of load resistance over the range 2½Ω to 75Ω.

17.10. The forward converter in example 17.1 dissipates 9.216kW. Specify the necessary inductance change so that the minimum inductor current is 25% of the average inductor current.

17.11. A boost converter has a 12V input voltage and dissipates into a load 960W when the output is 48V. If the inductor ripple current is 50% of the average inductor current, determine the duty cycle and inductance when the switching frequency is 20kHz. If the output voltage ripple is restricted to a maximum of 1%, determine the minimum output capacitance.

17.12. A buck-boost converter has a 12V input voltage and dissipates into a load 960W when the output is -48V. If the inductor ripple current is 50% of the average inductor current, determine the duty cycle and inductance when the switching frequency is 20kHz. If the output voltage ripple is restricted to a maximum of 1%, determine the minimum output capacitance.

17.13. The isolated flyback converter in figure 17.14c has an input voltage of 50V, an output of 25V, an on-state duty cycle ratio of 0.4, and a 20kHz switching frequency. If the load is a 5Ω resistor, determine

i. the transformer turns ratio
ii. the core self inductance such that the ripple current is half its average current.

17.14. The isolated forward converter in figure 17.14a has the following specification: E_i = 96V, $N_1{:}N_2{:}N_3$ = 1 with 4mH self inductance, filter inductance 250μH, load resistance 24Ω, onstate duty cycle = 0.4 and a 40kHz switching frequency. Determine

i. the output voltage and output ripple voltage if C=220μF
ii. average and p-p current in the 250μH output inductor
iii. the peak magnetising current in the model self inductance
iv. peak switching current.

17.15. The push-pull converter in figure 17.6a has the following specification: E_i = 96V, $N_p{:}N_p{:}N_s$ = 1:1:2 with 500μH of output inductance with respect to the primary, 12Ω load resistance, and a 25kHz switching frequency. For an on-state duty cycle of $\frac{1}{3}$ determine

i. the output voltage
ii. the average and p-p output-inductance current
iii. the output voltage ripple across a 470μF output capacitor.

Sketch the switch, diode, source and capacitors currents, using the inductor current as reference.

17.16. Repeat problem 17.15 if the core magnetising inductance (self-inductance) is 2.5mH with respect to the primaries. Having determined the peak magnetising current, add the magnetising inductance current waveform to the other sketched waveforms.

17.17 A forward converter operates at 50kHz with a 60% duty cycle from a 15V dc supply and delivers 27W into a resistive load. Determine the output voltage and inductor rms current. Sketch the capacitor and inductor current and voltage waveforms. What output capacitance will result in 1% output voltage ripple? What inductance will ensure continuous conduction at 3W output?

Blank

18

DC to DC Converters
- Resonant Mode

DC-dc/ac converters have two shortcomings when their switches operate in a hard switched mode.

- During the turn-on and turn-off transients, the switches experience simultaneous high current and voltage, resulting in high power losses, hence high switch stresses. The power loss increases linearly with the switching frequency. To ensure power conversion efficiency, the switching frequency has to be constrained.
- Electromagnetic interference (EMI) is generated by switch *dv/dt* and *di/dt*. The drawbacks have been accentuated by the trend to increase switching frequency in order to reduce the converter size and weight.

The resonant converters minimize these two shortcomings. The switches in resonant converters create a square-wave-like voltage or current waveform with or without a DC component. A resonant *L-C* circuit is incorporated with a resonant frequency close to the switching frequency, which can be varied around the resonant frequency. When the resonant *L-C* circuit frequency is approximately the switching frequency, unwanted harmonics are removed by the circuit. Variation of the switching frequency is a means of controlling the output power and voltage.

L-C resonant converter circuits offer:

- sinusoidal-like wave shapes,
- inherent filter action,
- reduced *dv/dt*, *di/dt* and EMI (radiated and conducted switching noise),
- facilitation of the turn-off process by providing zero current crossing for the switches and output power and voltage control by changing the switching frequency, and
- zero current and/or zero voltage across the switches at switching thus substantially reducing switch losses.

Unlike hard switched converters the switches in soft switched converters, quasi-resonant and some resonant converters are subjected to much lower switching stresses. Note that not all resonant converters offer zero current and/or zero voltage switching, that is, reduced switching power losses. These advantageous features are traded for switches that are subjected to higher forward currents and reverse voltages than encounter in a non-resonant configuration of the same power. Variable operating frequency can be a drawback.

Two main resonant techniques can be used to achieve near zero switching losses

- a resonant load that provides natural voltage or current zero instances for switching
- a resonant circuit across the switch which feeds energy to the load as well as introducing zero current or voltage instances for switching.

DC to ac resonant converters for induction heating type applications are considered in Chapter 16.3.

18.1 Series loaded resonant dc-to-dc converters

The basic converter operating concept involves a H-bridge producing an ac square-wave voltage $V_{H\text{-}B}$. When fed across a series *L-C* filter, a near sinusoidal oscillation current results, provided the square wave fundamental frequency is near the natural resonant frequency of the *L-C* filter. Because of *L-C* filter action, only fundamental current flows and current harmonics produced by the square wave voltage are attenuated due to the gain roll-off of the second order *L-C* filter. The sinusoidal resonant current is rectified to produce a load dc voltage $v_{o/p}$. The output voltage (voltage gain is less than unity) is highly dependant on the frequency relationship between the square-wave drive voltage period and the *L-C* filter resonant frequency.

Figure 18.1a shows the circuit diagram of a series resonant converter, which uses an output full-wave rectifier bridge to converter the resonant ac current oscillation into dc. The converter is based on the series converter in figure 16.2b. The rectified ac output charges the dc output capacitor, across which is the dc load, R_{load}. The non-dc-decoupled resistance, which determines the circuit *Q*, is account for by resistor R_c. The dc capacitor *C* capacitance is assumed large enough so that the output voltage $v_{o/p}$ is maintained constant, without significant ripple voltage. Figures 18.1b and c show how the dc output circuit can be transformed into an ac square wave in series with the *L-C* circuit, and finally this source is transferred to the dc link as a constant dc voltage source $v_{o/p}$ which opposes the dc supply V_s. These transformation steps enable the series *L-C-R* resonant circuit to be analysed with square wave inverter excitation, from a dc source V_s - $v_{o/p}$. This highlights that the output voltage must be less than the dc supply, that is V_s - $v_{o/p} \geq 0$, if current oscillation is to occur. The analysis in chapter 16.3 is valid for this circuit, where V_s is replaced by V_s - $v_{o/p}$. The equations, modified, are here repeated for completeness.

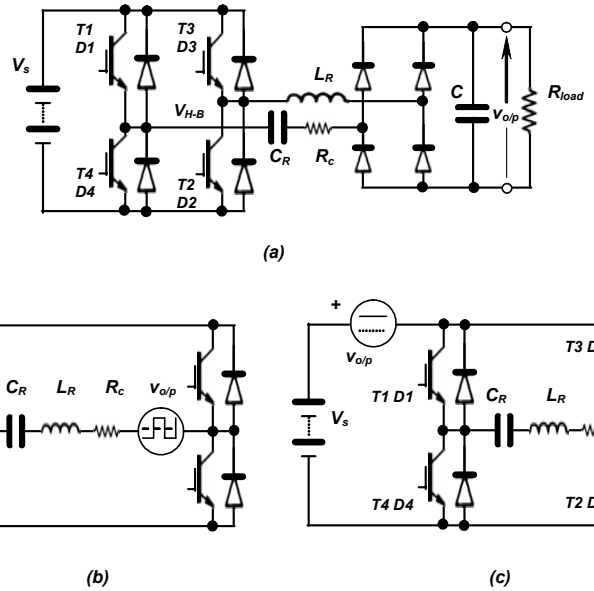


Figure 18.1. *Series resonant converter and its equivalent circuit derivation.*

The series *L-C-R* circuit current for a step input voltage V_s - $v_{o/p}$, with initial capacitor voltage v_o, assuming zero initial inductor current, is given by

$$i(\omega t) = \frac{(V_s - v_{o/p}) - v_o}{\omega L} \times e^{-\alpha t} \times \sin \omega t \qquad (18.1)$$

where

$$\omega^2 = \omega_o^2 (1 - \xi^2) \qquad \omega_o = \frac{1}{\sqrt{L_R C_R}} \qquad \alpha = \frac{R_c}{2L_R} \qquad Z_o = \sqrt{\frac{L_R}{C_R}}$$

ξ is the damping factor. The capacitor voltage is important because it specifies the energy retained in the *L-C-R* circuit at the end of each half cycle.

$$v_c(\omega t) = \left(V_s - v_{o/p}\right) - \left(\left(V_s - v_{o/p}\right) - v_o\right)\frac{\omega_o}{\omega} e^{-\alpha t}\cos\left(\omega t - \phi\right) \tag{18.2}$$

where

$$\tan\phi = \frac{\alpha}{\omega} \quad \text{and} \quad \omega_o^2 = \omega^2 + \alpha^2$$

At the series circuit resonance frequency ω_o, the lowest possible circuit impedance results, $Z = R_c$. The series circuit quality factor or figure of merit, Q_s is defined by

$$Q_s = \frac{\omega_o L}{R_c} = \frac{1}{2\xi} = \frac{Z_o}{R_c} \tag{18.3}$$

Operation is characterised by turning on switches T1 and T2 to provide energy from the dc source during one half of the cycle, then having turned T1 and T2 off, T3 and T4 are turned on for the second resonant half cycle. Energy is again drawn from the dc supply, and when the current reaches zero, T3 and T4 are turned off.

Without bridge freewheel diodes, the switches support high reverse bias voltages, but the switches control the start of each oscillation half cycle. With freewheel diodes the oscillations can continue independent of the switch states. The diodes return energy to the supply, hence reducing the net energy transferred to the load. Correct timing of the switches minimises currents in the freewheel diodes, hence minimises the energy needlessly being returned to the supply. Energy to the load is maximised. The switches can be used to control the effective load power factor. By advancing turn-off to occur before the switch current reaches zero, the load can be made to appear inductive, while delaying switch turn-on produces a capacitive load effect.

The series circuit steady-state current at resonance for the H-bridge with a high circuit Q can be approximated by assuming $\omega_o \approx \omega$, such that:

$$i(\omega t) = \frac{2}{1 - e^{\frac{-\alpha\pi}{\omega}}} \times \frac{\left(V_s - v_{o/p}\right)}{\omega L_R} \times e^{-\alpha t} \times \sin\omega t \tag{18.4}$$

which is valid for the ± $(V_s - v_{o/p})$ voltage loops of cycle operation at resonance. For a high circuit Q this equation is approximately

$$i(\omega t) \approx \frac{4}{\pi} \times Q \times \frac{\left(V_s - v_{o/p}\right)}{\omega_o L_R} \times \sin\omega_o t = \frac{4}{\pi} \times \frac{\left(V_s - v_{o/p}\right)}{R_c} \times \sin\omega_o t \tag{18.5}$$

The maximum current is

$$\hat{I} \approx \frac{4}{\pi} \times \frac{\left(V_s - v_{o/p}\right)}{R_c} \tag{18.6}$$

while the average current with this peak value must equal the load current, that is

$$\bar{I} = \frac{2}{\pi} \times \hat{I} = \frac{8}{\pi^2} \times \frac{\left(V_s - v_{o/p}\right)}{R_c} = \frac{v_{o/p}}{R} \tag{18.7}$$

The output voltage is obtained from equation (18.7) by isolating $v_{o/p}$

$$v_{o/p} = \frac{V_s}{1 + \frac{8}{\pi^2} \times \frac{R_c}{R}} \tag{18.8}$$

In steady-state the capacitor voltage maxima are

$$\begin{aligned} \hat{V}_c &= \left(V_s - v_{o/p}\right)\frac{1 + e^{-\alpha\pi/\omega}}{1 - e^{-\alpha\pi/\omega}} = \left(V_s - v_{o/p}\right) \times \coth\left(\alpha\pi/2\omega\right) = -\check{V}_c \\ &\approx \left(V_s - v_{o/p}\right) \times 2\omega_o/\alpha\pi = \frac{4}{\pi} \times Q \times \left(V_s - v_{o/p}\right) \end{aligned} \tag{18.9}$$

The peak-to-peak capacitor voltage, by symmetry is therefore

$$V_{c_{p-p}} \approx \frac{8}{\pi} \times Q \times \left(V_s - v_{o/p}\right) \tag{18.10}$$

The energy lost in the coil resistance R_c, per half sine cycle (per current pulse) is

$$\begin{aligned} W &= \int_0^{\pi/\omega} i^2 R_c \, dt \approx \int_0^{\pi/\omega_o} \left(4\frac{2}{\pi} \times \frac{\left(V_s - v_{o/p}\right)}{R_c} \times \sin\omega_o t\right)^2 R_c \, dt \\ &= \frac{8}{\pi\,\omega_o R_c}\left(V_s - v_{o/p}\right)^2 \end{aligned} \tag{18.11}$$

The relationship between the output voltage $v_{o/p}$ and H-bridge switching frequency, ω_s, is not a simple linear function. Because of the *L-C* series filter cut-off frequency $\omega_s \approx \omega_o$, only fundamental current flows as a result of the fundamental of the square-wave $V_{H\text{-}B}$, which has a magnitude of $\frac{4}{\pi}V_s$. A series *R-C-L* ac circuit at frequency ω_s can be used to derive a relationship between the output voltage $v_{o/p}$ and ω_s. The effective ac load resistance for a series L-C resonant circuit, from equation (18.7) becomes $R_{eq} = \frac{8}{\pi^2}R_{load} = \frac{8}{\pi^2}(V/I)$ such that Kirchhoff's voltage law for the series circuit, shown in figure 18.2a, is

$$\begin{aligned} \frac{4}{\pi}V_s &= i \times \left(R_{eq} + R_c + j\left(\omega_s L_R - \frac{1}{\omega_s C_R}\right)\right) = i \times \left(R_{eq} + R_c + j\left(X_L - X_C\right)\right) \\ \frac{4}{\pi}v_{o/p} &= i \times R_{eq} \end{aligned} \tag{18.12}$$

The equivalent output voltage is therefore given by

$$v_o(\omega_s) = V_s \times \frac{R_{eq}}{R_{eq} + R_c + j\left(\omega_s L_R - \frac{1}{\omega_s C_R}\right)} \tag{18.13}$$

where the H-bridge switching frequency is $\omega_s = 2\pi f_s$.

Figure 18.2 shows equation (18.13) for different circuit Q. The plot can be used to extract output voltage $v_{o/p}$ and H-bridge switching frequency. The output voltage is scaled to eliminate the coil resistance component R_c from the total resistive voltage.

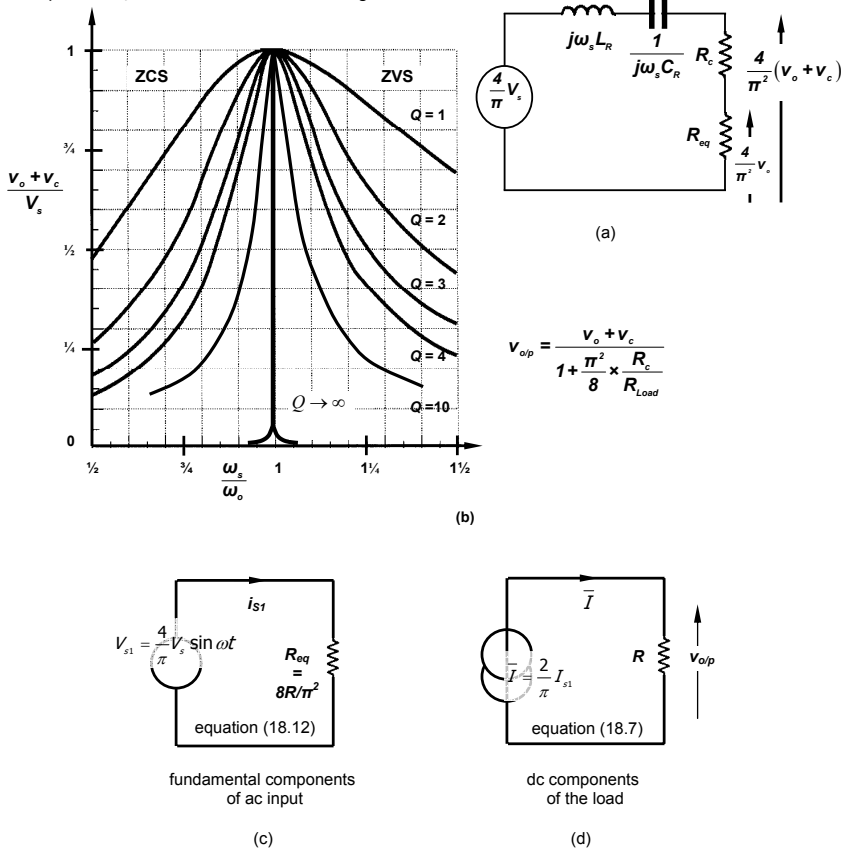


Figure 18.2. *Series resonant converter SRC: (a) equivalent circuit; (b) normalised output voltage curves; (c) input ac equivalent circuit; and (d) load dc equivalent circuit.*

Ignoring coil resistance, the voltage transfer function is

$$\frac{v_o(\omega)}{V_s} = \frac{1}{R_{eq} + j\left(\omega_o L_R - \frac{1}{\omega_o C_R}\right)} = \frac{1}{1 + jQ\left(\frac{\omega}{\omega_o} - \frac{\omega_o}{\omega}\right)}$$

$$\left|\frac{v_o(\omega)}{V_s}\right| = \frac{1}{\sqrt{1 + Q^2\left(\frac{\omega}{\omega_o} - \frac{\omega_o}{\omega}\right)^2}} \le 1; \qquad \psi = -\tan^{-1} Q\left(\frac{\omega}{\omega_o} - \frac{\omega_o}{\omega}\right) \tag{18.14}$$

At $f = f_o$, which is the condition for maximum voltage gain, the input magnitude equals the output magnitude. When $f < f_o$, ψ is positive, and the output voltage, whence output current, leads the input voltage, meaning the load is capacitive. For $f > f_o$, ψ is negative, and the output voltage, whence output current, lags the input voltage, meaning the load is inductive, as indicated in figure 18.3.

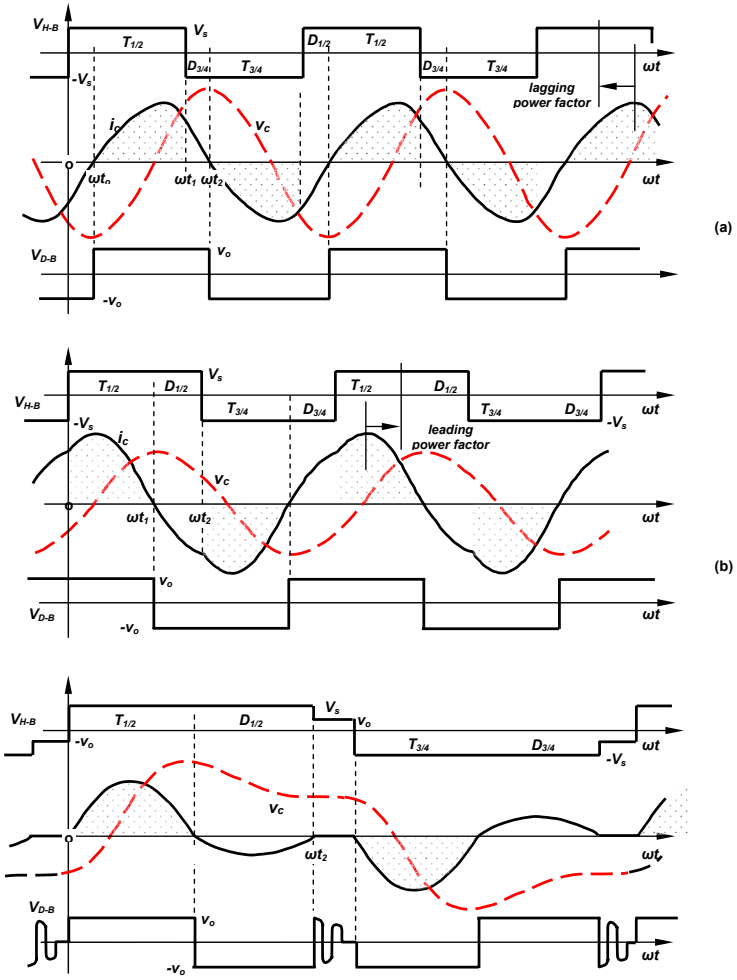


Figure 18.3. *Three modes of series load resonant converter operation: (a) $f_s > f_o$; (b) $½f_o < f_s < f_o$; and (c) $f_s < ½f_o$.*

18.1.1 Modes of operation - series resonant circuit

The basic series converter can be operated in any of three difference modes, depending on the switching frequency f_s in relation to the *L-C* circuit natural resonant frequency f_o. In all cases, the controlled output voltage is less than the input voltage, that is $V_s - v_{o/p} \ge 0$. The switching frequency involves one complete symmetrical square-wave output cycle from the inverter bridge, $V_{H\text{-}B}$. Waveforms for the three operational modes are shown in figure 18.3.

i. **$f_s < ½f_o$** :- discontinuous inductor current (switch conduction $1/2f_o \le t_T \le 1/f_o$)
If the switching frequency is less than half the *L-C* circuit natural resonant frequency, as shown in figure 18.3c, then discontinuous inductor current results. This is because once one complete *L-C* resonant ac cycle occurs and current stops - being unable to reverse, since the switches are turned off when the diodes conduct and the capacitor voltage is less than V_s+v_o. Turn-off occurs at zero current. Subsequent turn-on occurs at zero current but the voltage is determined by the voltage retained by the capacitor. Thyristors are therefore applicable switches in this mode of operation. The freewheel diodes turn on and off with zero current. Since the capacitor current rectified provides the load current average current, the H-bridge switching frequency controls the output voltage. Therefore at low switching frequencies, relative to the resonance frequency, the peak resonant current will be relatively high.

ii. **$½f_o < f_s < f_o$** :- continuous inductor current – leading power factor
If the switching frequency is just less than natural resonant frequency, as shown in figure 18.3b, such that turn-on occurs after half an oscillation cycle but before the completion of an ac oscillation cycle, continuous inductor current results. Switch turn-on occurs with finite inductor current and voltage conditions, with the diodes freewheeling. Diode reverse recovery losses occur and noise is injected into the circuit at voltage recovery snap. Fast recovery diodes are therefore necessary. Switch turn-off occurs at zero voltage and current, when the inductor current passes through zero and the freewheel diodes take up conduction. Thyristors are applicable as switching devices with this mode of control.

iii. **$f_s > f_o$** :- continuous inductor current – lagging power factor
If switch turn-off occurs before the resonance of half a resonant cycle is complete, as shown in figure 18.3a, continuous inductor current flows, hard switching results, and commutable switches must be used. Switch turn-on occurs at zero voltage and current hence no diode recovery snap occurs. This zero electrical condition turn-on allows lossless turn-off snubbers to be employed (a capacitor in parallel with each switch, or in parallel with any one switch in each leg).

For continuous inductor conduction, under light load conditions, resonant energy is continuously transferred to the output stage, which tends to progressively overcharge the output voltage towards the input voltage level, V_s. The charging progressively decreases as the supply voltage V_s is backed off by the increasing output voltage. That is $V_s - v_{o/p}$ shown in figure 18.1c tends to zero such that the effective square wave input is reduced to zero, as will the input energy.

18.1.2 Circuit variations

The number of semiconductors can be reduced by using a split dc rail as in figure 18.4a, at the expense of halving the bridge output voltage swing $V_{H\text{-}B}$ to $\pm ½V_s$. The maximum dc output voltage is $v_o \le ½V_s$. Although the number of semiconductors is halved, the already poor switch utilisation associated with any resonant converter, is further decreased. The switches and diodes support V_s. In the full-bridge case the corresponding switch and diode voltages are also both V_s.
Voltage and impedance matching, for example voltage step-up, can be obtained by using a high-frequency transformer coupled circuit as shown in figures 18.4b and c. When a transformer is used as in figure 18.4c, a centre tapped secondary can reduce the number of high frequency rectifying diodes from four to two, but diode reverse voltage rating is doubled from $V_{D\text{-}B}$ to $2V_{D\text{-}B}$. Secondary copper winding utilisation is halved.
A further modification to any series converter is to used the resonant capacitor to form a split dc rail as shown in figure 18.4c, where each capacitance is $½C_R$. In ac terms the resonant capacitors are in parallel, with one charging while the other discharges, and visa versa, such that their voltage sums to V_s.

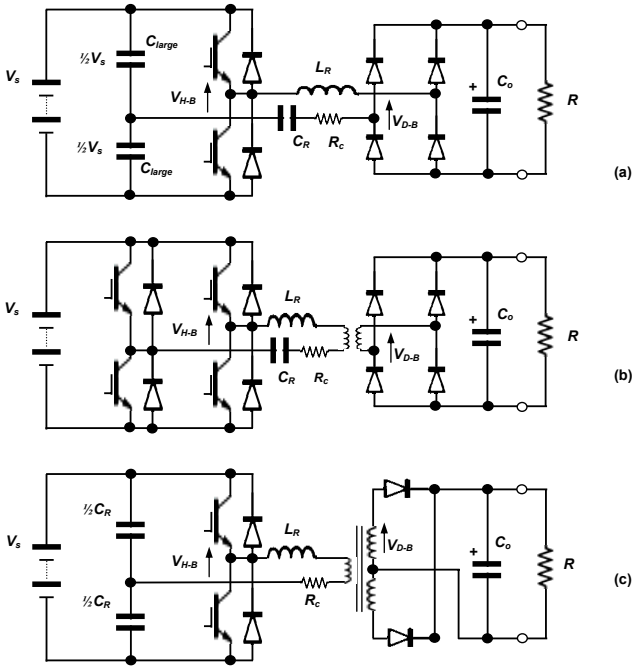


Figure 18.4. *Series load resonant converters variations:*
(a) half bridge, split dc supply rail; (b) transformer couple full bridge; and
(c) split resonant polarised capacitor, with a centre tapped output rectifier stage.

18.2 Parallel loaded resonant dc-to-dc converters

The basic parallel load resonant dc-to-dc converter is shown in figure 18.5a and its equivalent circuit is shown in figure 18.5b. The inductor L_o in the rectified output circuit produces a near constant current. A key feature is that the output voltage $v_{o/p}$ can be greater than the input voltage V_s, that is $0 \le v_{o/p} \ge V_s$.
The capacitor voltage and inductor current equations for the equivalent circuit in figure 18.5b, for a constant current load I_o, and high circuit Q, are

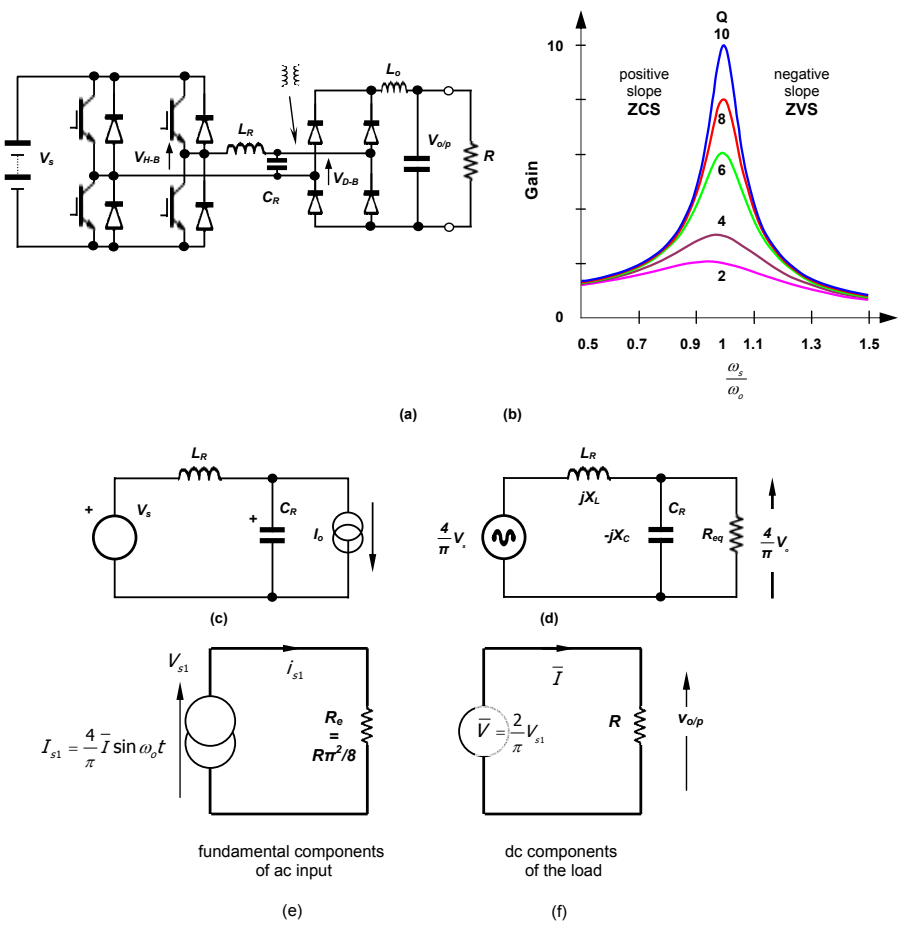
$$\begin{aligned} i_L(t) &= I_o + (i_{Lo} - I_o)\cos\omega_o t + \frac{V_s - v_{Co}}{Z_o}\sin\omega_o t \\ &= I_o + \frac{V_s}{Z_o}\sin\omega_o t \quad \text{for } v_{Co} = 0 \text{ and } i_{Lo} = I_o \end{aligned} \tag{18.15}$$

$$\begin{aligned} v_c(t) &= V_s - (V_s - v_{Co})\cos\omega_o t + Z_o(i_{Lo} - I_o)\sin\omega_o t \\ &= V_s(1 - \cos\omega_o t) \quad \text{for } v_{Co} = 0 \text{ and } i_{Lo} = I_o \end{aligned} \tag{18.16}$$

The relationship between the output voltage and the bridge switching frequency can determined from the equivalent circuit shown in figure 18.5c where the output resistance has been replaced by its equivalent resistance related to the H-bridge output fundament frequency magnitude, $V_{H\text{-}B} = \frac{2}{\pi}V_s$. The voltage across the resonant capacitor C_R is assumed to be sinusoidal.
Kirchhoff analysis of the equivalent circuit in figure 18.5b gives

$$\left|\frac{v_o}{V_s}\right| = \frac{8}{\pi^2} \times \left|\frac{1}{1 - \frac{X_L}{X_C} + j\frac{X_L}{R_{eq}}}\right| = \frac{8}{\pi^2}\left|\frac{1}{\left(1 - \frac{\omega^2}{\omega_o^2}\right) + j\frac{8}{\pi^2}\frac{1}{Q}\frac{\omega}{\omega_o}}\right| = \frac{8}{\pi^2} \times \frac{1}{\sqrt{\left(1 - \frac{X_L}{X_C}\right)^2 + \left(\frac{X_L}{R_{eq}}\right)^2}} \tag{18.17}$$

where the load resistance R is related to the equivalent ac resistance, at the switching frequency ω_s, by $R_{eq} = \frac{\pi^2}{8} \times R$, for the parallel resonant load circuit. Series stray non-load resistance, which reduces the circuit Q ($=R/\omega_o L_R$), has been neglected. The current gain is the voltage gain (0 to ∞) divided by the Q.



fundamental components of ac input

dc components of the load

Figure 18.5. *Parallel resonant dc-to-dc converter PRC:*
(a) circuit; (b) dc characteristics; (c) equivalent ac circuit; (d) equivalent fundamental input voltage circuit; (e) input ac equivalent circuit; and (f) load dc equivalent circuit.

18.2.1 Modes of operation - parallel resonant circuit

Three modes of operation are applicable to the parallel-resonant circuit, dc-to-dc converter, and waveforms are shown in figure 18.6.

i. **$f_s < \frac{1}{2}f_o$** :- discontinuous inductor current (switch conduction $1/2f_o \le t_z \le 1/f_o$)
Initially all switches are off and the load current energy stored in L_o freewheels through the output bridge diodes.
Both the inductor current and capacitor voltage are zero at the beginning of the cycle and at the end of the cycle. Thus switch turn-on and turn-off occur with zero current losses. At H-bridge turn-on the resonant inductor current increases linearly according to $i = V_s t / L_R$ until the output current level I_o is reached at time $t_1 = L_R I_o / V_s$, when the capacitor is free to resonant. The capacitor voltage is given by

$$v_c(t) = V_s(1 - \cos\omega_o t) \tag{18.18}$$

while the inductor current is given by

$$i_L(t) = I_o + \frac{V_s}{Z_o}\sin(\omega_o t) \tag{18.19}$$

The resonant circuit inductor current reverses as the on-switch antiparallel freewheel diodes conduct, at which time the switches may be turned off with zero current and voltage conditions. Any further inductor current reversal is therefore not possible. At the attempted reversal instant, any retained capacitor charge is discharged at a constant rate I_o into the inductor L_o. The capacitor voltage falls linearly to zero at which time the current in L_o freewheels in the output rectifier diodes.

ii. $\frac{1}{2}f_o < f_s < f_o$:- continuous inductor current – leading power factor
When switching below resonance, the switches commutate naturally at turn-off, as shown in figure 18.6b, making thyristors a possibility.
Hard turn-on results, necessitating the use of fast recovery diodes.

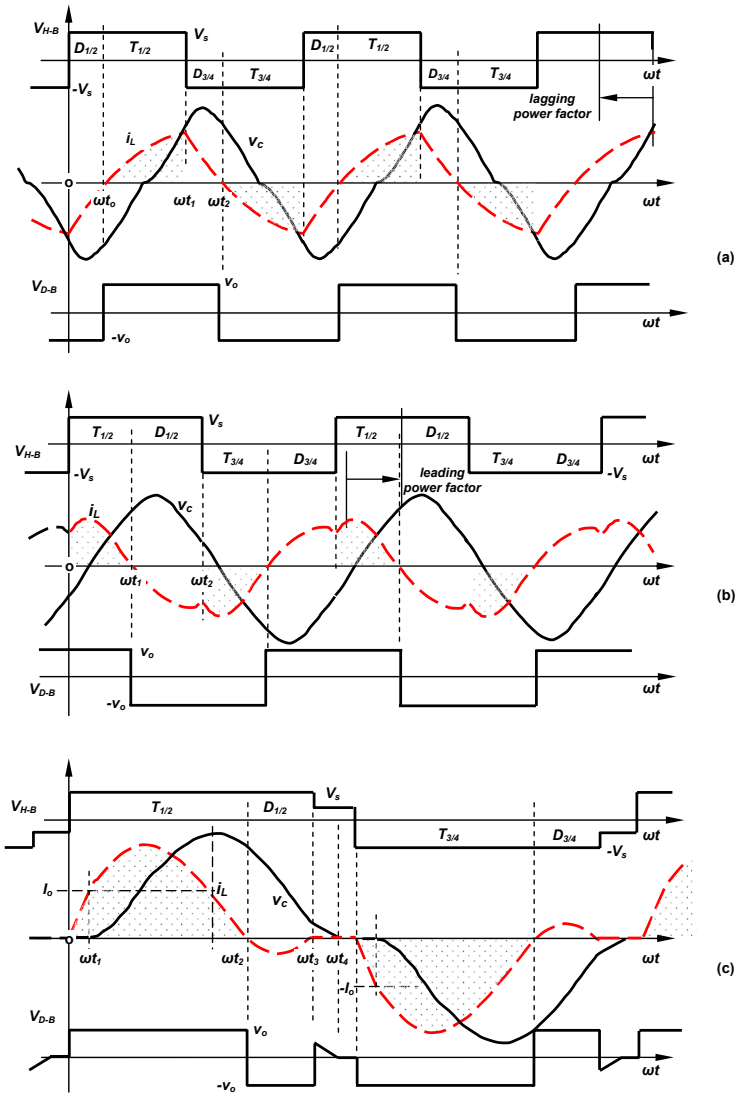


Figure 18.6. *Three modes of parallel load resonant converter operation: (a) $f_s > f_o$; (b) $\frac{1}{2}f_o < f_s < f_o$; and (c) $f_s < \frac{1}{2}f_o$.*

iii. $f_s > f_o$:- continuous inductor current – lagging power factor
When switching at frequencies above the natural resonance frequency, no turn-on losses result since turn-on occurs when a switches antiparallel diode is conducting. Turn-off occurs with inductor current flowing, hence hard turn-off results, with switch current commutated to a freewheel diode. In mitigation, lossless capacitive turn-off snubber can be used (a capacitor in parallel with each switch).

18.2.2 Circuit variations

A parallel resonant circuit approach, with or without transformer coupling, can also be use as indicated in figure 18.5a. The centre tapped secondary approach shown in figure 18.4c is also applicable. A dc capacitor split dc rail can also be used, as shown for the series load resonate circuit in figure 18.4a.
In each case the constant current inductor L_o is in the rectified output circuit.

18.3 Series–parallel load resonant dc to dc converters

Three passive element, parallel plus series, resonant circuits can be employed as shown in figures 18.7a and 18.8a, which attempt to combine to best features of series and parallel resonant tank circuits. The main limitation eliminated is the high no-load circulating current associated with parallel resonance.

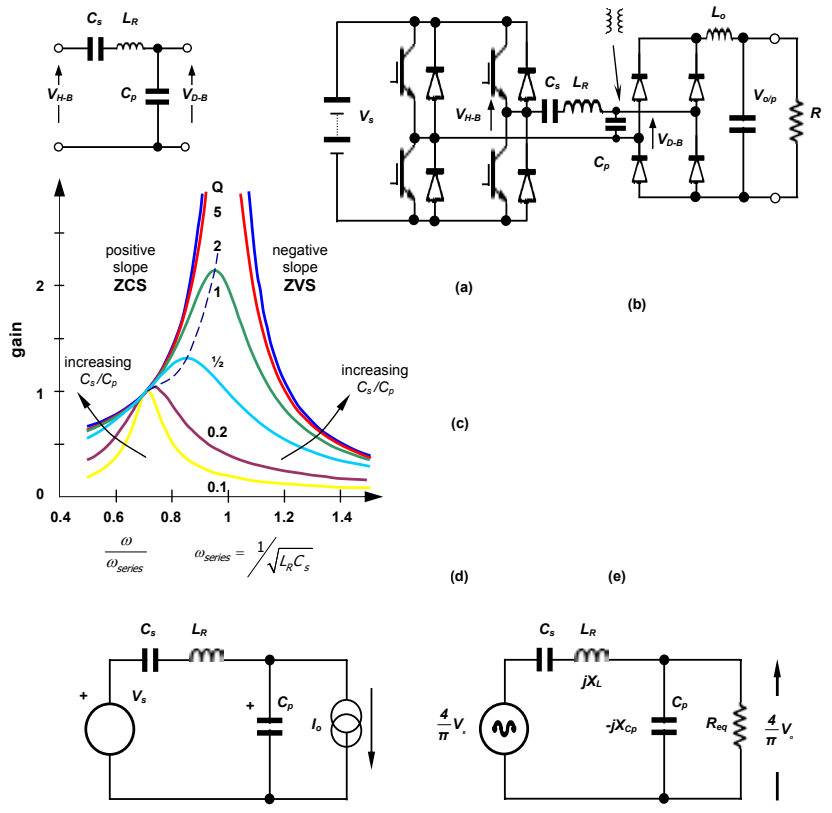


Figure 18.7. *Series-parallel LCC resonant dc-to-dc converter: (a) basic series–parallel resonant circuit; (b) series–parallel load resonant converter circuit; (c) dc characteristics; (d) equivalent ac circuit; and (e) equivalent fundamental input voltage circuit.*

18.3.1 LCC resonant tank circuit

A further load resonant converter variation is a combined series-parallel (third order) resonant stage as shown in figure 18.7a and the implemented circuit is figure 18.7b. This resonant approach combines the properties of the series and the parallel resonant approaches. Analysis is based on the same assumptions as for the parallel resonant cases, namely the H-bridge produces a square-wave voltage $V_{H\text{-}B}$ at a frequency near the filter resonance frequency f_o which results in a near sinusoidal voltage across the parallel resonance capacitor C_p. The voltage across the parallel output capacitor is rectified by the output bridge rectifier, while the bridge output inductor, under steady state conditions, ensures near constant load current I_o.

Figure 18.7d shows the converter equivalent circuit, while figure 18.7e shows the ac equivalent circuit where it is assumed that the bridge square wave frequency, which is its fundamental frequency, is near the resonant frequency. The ac equivalent resistance is the same as for the parallel resonant case, namely

$$R_{eq} = \frac{\pi^2}{8} R \tag{18.20}$$

In this case, the voltage transfer ratio, in terms of the fundamental input component, is given by

$$\frac{v_o}{V_s} = \frac{8}{\pi^2} \times \frac{1}{\sqrt{\left(1 + \frac{C_p}{C_s} - \omega^2 L_R C_p\right)^2 + \left(\frac{8}{\pi^2 R}\right) \times \left(\omega L_R - \frac{1}{\omega C_s}\right)^2}} \tag{18.21}$$

The magnitude of the dc to ac voltage transfer function is

$$\left|\frac{V_o}{V_s}\right| = \frac{1}{\sqrt{(1+\lambda)^2 \left(1 - \left(\frac{\omega}{\omega_o}\right)^2\right)^2 + \frac{1}{Q^2} \times \left(\frac{\omega}{\omega_o} - \frac{\omega_o}{\omega} \frac{\lambda}{1+\lambda}\right)^2}} \tag{18.22}$$

where $\lambda = C_s/C_p$.

At light and no-load, the operating point tends to the resonant peak at the upper resonant frequency

$$\omega_{o2}^2 = \frac{1}{L_R}\left(\frac{1}{C_p} + \frac{1}{C_s}\right) = \frac{1}{L_R} \frac{1}{C_{equiv}} = \frac{C_p + C_s}{L_R C_p C_s} \tag{18.23}$$

The lower resonant frequency is $\omega_{o1}^2 = 1/L_R C_s$

All the circuit variations applicable to the parallel resonant circuit are also applicable to the series – parallel resonant converter. Additional advantage can be gained by the fact that a series resonant capacitor can be used to form the ac split dc rail arrangement shown in figure 18.4c.

18.3.2 LLC resonant tank circuit

The inductive elements in the tank circuit shown in figure 18.8a form the leakage inductance and magnetising inductance when the load is transformer coupled. The circuit is less sensitive to load changes than series resonant circuits and has lower circulating current under light and no-load conditions, than parallel resonant circuits. Two resonant frequencies occur

$$f_{o1} = \frac{1}{2\pi\sqrt{L_R C_R}} \quad \text{and} \quad f_{o2} = \frac{1}{2\pi\sqrt{(L_p + L_R) C_R}} = f_{o1}\sqrt{\frac{\lambda}{\lambda+1}} \tag{18.24}$$

The voltage transfer function is approximately (expression is correct at resonance)

$$\frac{V_{o/p}}{V_s} = \frac{j\omega \frac{L_p}{L_s}}{j\omega\left(\frac{L_p}{L_s} + 1 - \frac{1}{\omega^2}\right) + \frac{L_p}{L_s}\left(1-\omega^2\right)\frac{Z}{R_{dc}}\frac{\pi^2}{8}} = \frac{1}{\left(1 + \lambda - \frac{\lambda}{\omega^2}\right) - j\frac{\left(1-\omega^2\right)}{\omega}\frac{Z}{R_{dc}}\frac{\pi^2}{8}} \tag{18.25}$$

where $Z = \sqrt{\frac{L_s}{C_s}}$, $R_{eq} = \frac{8}{\pi^2} R_{dc}$, $\lambda = \frac{L_R}{L_p}$, and $\omega = \omega_s / \omega_o$ is the normalised switching frequency. Then

$$\left|\frac{V_{o/p}}{V_s}\right| = \frac{1}{\sqrt{\left(1 + \lambda - \frac{\lambda}{\omega^2}\right)^2 + \left(\frac{1}{\omega} - \omega\right)^2 \left(\frac{Z}{R_{dc}}\frac{\pi^2}{8}\right)^2}} = \frac{1}{\sqrt{\left(1 + \lambda - \frac{\lambda}{\omega^2}\right)^2 + Q^2\left(\frac{1}{\omega} - \omega\right)^2}} \tag{18.26}$$

where $Q = \frac{Z_o}{R_{eq}} = \frac{Z_o}{n^2 R_{eq}} = \frac{\pi^2 Z_o}{8 n^2 R_{dc}} = \frac{\pi^2 Z_o P_{out}}{8 n^2 V_{dc}^2}$. A transformer is account for by the turns ratio n:1 term n.

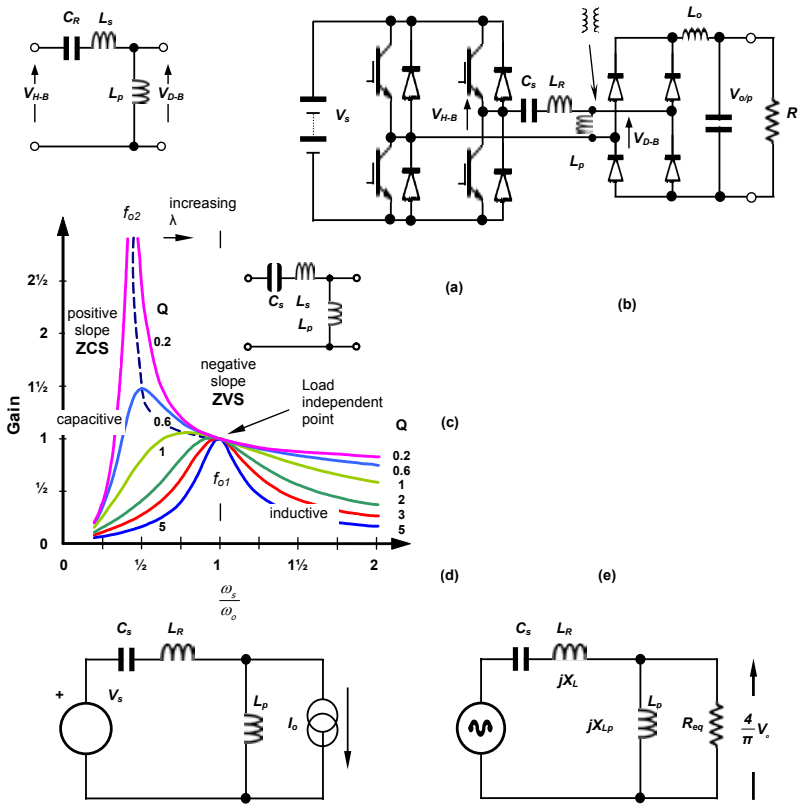


Figure 18.8. *Series-parallel LLC resonant dc-to-dc converter: (a) basic series–parallel resonant circuit; (b) series–parallel load resonant converter circuit; (c) dc characteristics; (d) equivalent ac circuit; and (e) equivalent fundamental input voltage circuit.*

The dc characteristic on the right of f_{o1} for the LLC converter are same characteristic as a SRC.

On the left of f_{o1}, the PRC and SRC characteristics emerge but oppose. At heavy loading, SRC characteristics dominant. At lighter loads, PRC characteristic dominate and the no load output magnitude, as Q tends to zero, is

$$\left|\frac{V_{o/p}}{V_s}\right| = \frac{1}{\left|1 + \lambda - \frac{\lambda}{\omega^2}\right|} \tag{18.27}$$

By operating the LLC converter at a resonant frequency lower than that of a SRC, it retains a ZVS characteristic because PRC characteristics dominant in that frequency range.

Load resonant circuit comparison

The series-parallel approach combines the better switching properties of the series converter and the parallel converter, which are summarised in table 18.1. In common, a positive slope on the voltage transfer function versus frequency characteristics are associated with zero current switching ZCS, while a negative slope is indicative of zero voltage switching ZVS characteristics. Basically, the tank circuit is capacitive below resonance and becomes inductive above the resonance frequency.

Unlike the series converter, both the parallel and series-parallel converters can operate at zero load current, without output voltage over-charging. But advantageously, because of the LCC series input capacitor C_s, the efficiency at low power output levels is like that of the series converter, which is much higher than that of the series–parallel converter. The resonant currents with either parallel approach tend to be independent of the load.

Series resonant converter SRC
The resonant tank is in series with the load. The resonant tank and the load act as a voltage divider. By changing the frequency of input excitation voltage, the impedance of resonant tank will change, which allows to load voltage to be varied. As a voltage divider, the DC gain of SRC is always less than 1. At the resonant frequency, the maximum voltage gain frequency, the impedance of series resonant tank is small; all the input voltage (fundamental) is developed across the load.
Operating region is on the right side of resonant frequency fo is preferred because zero voltage switching (ZVS) is preferred for the LC series converter. When the switching frequency is below the resonant frequency, the converter operates with zero current switching (ZCS). When the DC gain slope is negative, the converter operates under a zero voltage switching condition. When the DC gain slope is positive, the converter switches under a zero current switching condition.
The voltage gain characteristics indicate that at light load, the switching frequency needs to increase, to reduce the gain, in order to maintain output voltage regulation.
The major problems are: light load regulation, high circulating energy and turn off current at high input voltage condition.

Parallel resonant converter PRC
For the parallel resonant converter, the resonant LC tank is again in series. It is called parallel resonant converter because the load is in parallel with the resonant capacitor (or inductor). Technically this converter should be termed a series resonant converter with a connected parallel load.
Compare with a SRC, the operating region is much smaller. At light load, the frequency is minimally changed to maintain output voltage regulation. So light load regulation is not a problem with the PRC.
For the PRC, circulating energy is high even at light load. Since the load is in parallel with the resonant capacitor, even at no load condition, the input sees the impedance of the series resonant tank. This induces a high circulating energy even when the load is zero.
The major problems are: high circulating energy, high turn-off current at high input voltage condition.

Series-parallel resonant converter SPRC
The resonant tank of SPRC can be considered a combination of SRC and PRC converters, producing the best characteristics of each. With load in series with series tank L and C, the circulating energy is smaller than with the PRC. With the parallel capacitor, the SPRC can regulate the output voltage at a no load condition, by operating at a frequency above resonance.
The operating region of the SPRC has a narrower switching frequency range with load change than the SRC. The input current in much smaller than that of the PRC and slightly larger than for the SRC. This means for the SPRC, the circulating energy is less than with the PRC.
A SPRC combines the better characteristics of a SRC and a PRC. Smaller circulating energy and is not too sensitive to load change.

Table 18.1: Switch and diode turn-on/turn-off conditions for resonant switch converters

	series load resonance		parallel load resonance		
	switch diode on off	switch diode off on	switch diode on off	switch diode off on	power factor
$f_s < ½f_o$	zcs	zcs zvs	zcs	zcs zvs	N/A
$f_s < f_o$	hard$^+$	zcs zvs	hard$^+$	zcs zvs	leading
$f_s > f_o$	zcs zvs	hard*	zcs zvs	hard*	lagging

Lossless, * capacitive turn-off or $^+$ inductive turn-on snubbers can be used, as appropriate

18.4 Resonant coupled-load configurations

The concept of voltage and current transformer coupling is introduced. A series LC resonant circuit produces a resonant sinusoidal current and anti-phased sinusoidal voltages across both the circuit L and C components. A current transformer can therefore be used in series with the resonant circuit, while, alternatively a voltage transformer can be used in parallel with either the circuit L or C.

In the case of a parallel transformer resonant circuit, a voltage transformer is used across the resonant capacitor (or inductor) as shown in figure 18.9b, while for a series transformer resonant circuit, a current transformer is used in series with the resonant circuit inductor and capacitor, as shown in figure 18.9a.

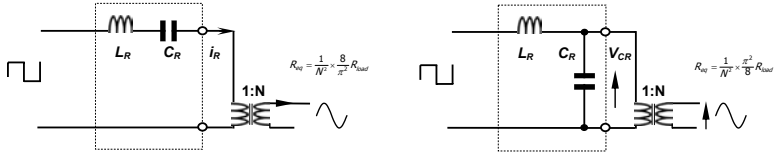


Figure 18.9. *Voltage-source transformer resonant converter circuits: (a) series transformer LC resonant circuit and (b) parallel transformer LC resonant circuit.*

In series, as shown in figure 18.9a, the primary current is a controlled sinusoid, hence by Ampere-turns transformer action, the secondary is a current source sinusoidal waveform. The secondary current is faithfully delivered into any voltage. This mode of transformer action is therefore termed 'current' transformer operation since the transformer primary is driven to obey $I_{prim} = N \times I_{sec}$. However, the secondary circuit must have low series inductance, since the secondary is driven by a current source.

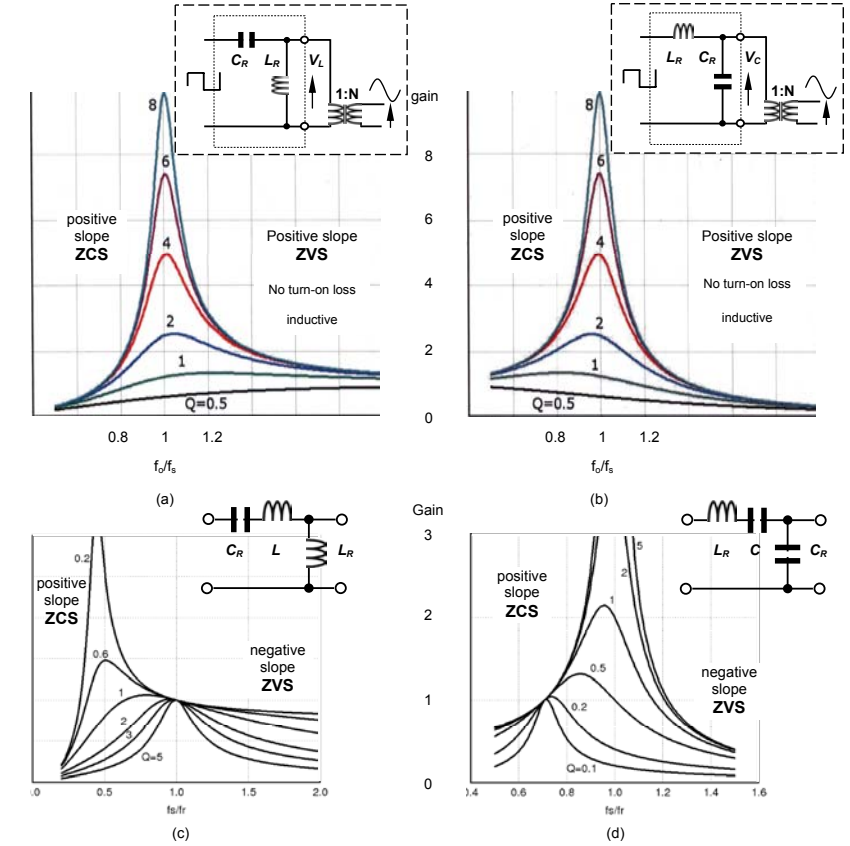


Figure 18.10. *Voltage transformer coupled, resonant converter circuits, LC, LLC, LCC: (a) and (c) parallel resonant inductor L_R transformer coupling circuits and (b) and (d) parallel resonant capacitor C_R transformer coupling circuit.*

A voltage coupling transformer requires a voltage source, so the voltage across the capacitor C_R (or alternatively inductor L_R) of a series resonant circuit can be utilised as shown in figure 18.9b and figure 18.10. The primary voltage is controlled by the capacitor C_R voltage, as is the secondary voltage when complying with $V_{sec} = N \times V_{prim}$. Although the voltage is faithfully reproduced on the secondary, the secondary current, hence primary current, are determined by the secondary load circuit. A transformer used in this mode is termed a 'voltage' transformer. However, in order to prevent uncontrolled secondary currents, the secondary inductor, L_o, in figures 18.5, 18.7 and 18.8, is introduced to give controlled output current source properties.

The advantage of the parallel transformer resonant circuit over the series approach is that the turns ratio of the transformer can be reduced (increased primary turns for a given output specification) because of the higher gain (greater than unity) of parallel-coupled resonant circuits (see figures 18.2b and 18.5b). For example, if the circuit *Q* is four, the transformer turns ratio can be decreased by a similar factor over that needed for a series transformer resonance circuit.
Both the parallel transformer coupled circuits in figure 18.10 can take advantage of inherent voltage gain in order to reduce the necessary transformer turns ratio. Figure 18.10 shows that the voltage transformer can by parallel coupled to the inductor, as in figure 18.10a, or the capacitor, as in figure 18.10b. At resonance, the two circuits behave identically. The off-resonance characteristics are different (opposite), particular at frequencies above resonance, where ZVS and inductive operation occur in both cases. Above resonance, the inductor supports an increasing amount of the H-bridge output ac voltage as the capacitor reactance, hence voltage and output gain diminish to zero in figure 18.10b, but the output is unity in figure 18.10a. The ability to reduce the gain to near zero, with ZVS, is important at converter start-up, and enables the start-up current inrush into any output capacitor to be better controlled. The inductance of the output inductor L_o can thereby be minimised.

Example 18.1: ***Transformer-coupled, series-resonant, dc-to-dc converter***

The series resonant dc step-down voltage converter in figure 18.1a is operated at just above the resonant frequency of the load circuit and is used with a step up transformer, 1:2 (n_T = ½), as shown in figure 18.4a. It produces an output voltage for the armature of a high voltage dc motor that has a voltage requirement that is greater than the 50Hz ac mains rectified, 340V dc, with an *L-C* dc link filter. The resonant circuit parameters are L_R=100μH, C_R=0.47μF, and the coil resistance is R_c = 1Ω.
For a 10Ω armature resistance, R_{load}, calculate

i. the circuit *Q* and ω_o
ii. the output voltage, hence dc armature current and power delivered
iii. the secondary circuit dc filter capacitor voltage and rms current rating
iv. the resonant circuit rms ac current and capacitor rms ac voltage
v. the converter average input current and efficiency
vi. the ac current in the input *L-C* dc rectifier filter decoupling capacitor
vii. the H-bridge square-wave switching frequency ω_s, greater than ω_o.

Solution

i. The resonant circuit *Q* is

$$Q = \sqrt{\frac{L_R}{C_R}} \Big/ R_c = \sqrt{\frac{100\mu H}{0.47\mu F}} \Big/ 1\Omega = 14.6$$

For this high Q, the circuit resonant frequency and damped frequency will be almost the same, that is

$$\omega \approx \omega_o = 1/\sqrt{L_R C_R}$$
$$= 1/\sqrt{100\mu H \times 0.47\mu F} = 146 \text{ krad/s}$$
$$= 2\pi f$$
$$f = 146 \text{ krad/s} / 2\pi = 23.25 \text{ kHz}$$

ii. From equation (18.7), which will be accurate because of a high circuit *Q* of 14.6,

$$\bar{I} = \frac{2}{\pi}\hat{I} = \frac{8}{\pi^2}\frac{(V_s - n_T \times v_{o/p})}{R_c} = \frac{8}{\pi^2} \times \frac{(340\text{ V} - ½ v_{o/p})}{1\Omega}$$
$$= 0.81 \times (340\text{ V} - ½ v_{o/p})$$

Note that the output voltage $v_{o/p}$ across the dc decoupling capacitor has been referred to the primary by n_T, hence halved, due to the turns ratio of 1:2.
The rectified resonant current provides the load current, that is

$$\bar{I} = \frac{1}{n_T} \times \frac{v_{o/p}}{R_{load}} = 2 \times \frac{v_{o/p}}{10\Omega} = \frac{v_{o/p}}{5}$$

Again the secondary current has been referred to the primary. Solving the two average primary current equations gives

$$\bar{I} = 0.81 \times (340 - ½ v_o) = \frac{v_{o/p}}{5}$$
$$v_{o/p} = 456\text{V and } \bar{I} = 91.2\text{A}$$

That is, the load voltage is 456V dc and the load current is 456V/10Ω = 91.2A/2 = 45.6A dc. The power delivered to the load is 456^2/10Ω = 20.8kW.

iii. From part ii, the capacitor dc voltage requirement is at least 456V dc. The secondary rms current is

$$I_{Srms} = n_T \times I_{Prms} = n_T \times \frac{1}{\sqrt{2}} \times \hat{I}_p = ½ \times \frac{1}{\sqrt{2}} \times \frac{\pi}{2} \times \bar{I}_p$$
$$= 0.555 \times \bar{I}_p = 0.555 \times 91.2\text{A}$$
$$= 50.65\text{A rms}$$

The primary rms current is double the secondary rms current, 101.3A rms.
By Kirchhoff's current law, the secondary current (50.65A rms) splits between the load (45.6A dc) and the decoupling capacitor. That is the rms current in the capacitor is

$$I_{Crms} = \sqrt{I_{Srms}^2 - \bar{I}_S^2} = \sqrt{50.65\text{A}^2 - 45.6\text{A}^2} = 22\text{A rms}$$

That is, the secondary dc filter capacitor has a dc voltage requirement of 456V dc and a current requirement of 22A rms at 46.5kHz, which is double the resonant frequency because of the rectification process.

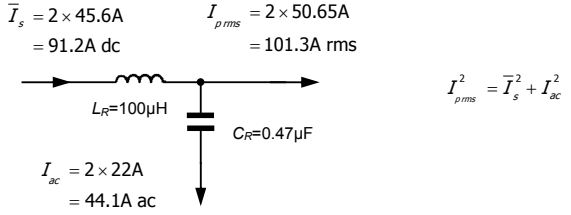
iv. The primary rms current is double the secondary rms current, namely from part iii, $I_{P\,rms}$ = 101.3A rms (essentially 101.3A at 23.25kHz). The 0.47μF resonant capacitor voltage is given by

$$v_{cap} = I_{Prms} X_c = \frac{I_{Prms}}{\omega_o C_R}$$
$$= \frac{101.3\text{A}}{146\text{krad/s} \times 0.47\mu\text{F}} = 1476\text{V rms} \quad \text{(predominately at 23.25kHz))}$$

The resonant circuit capacitor has an rms current rating requirement of 101.3A rms and an rms voltage rating of 1476 V rms.

v. From part ii, the average input current is 91.2 A. The supply input power is therefore 340Vdc×91.2A ave = 31kW. The power dissipated in the resonant circuit resistance R_c =1Ω is given by $I_{Prms}^2 \times R_c = 101.3^2 \times 1\Omega = 10.26\text{kW}$. The coil power plus the load power (from part ii) equals the input power (20.8kW+10.26kW = 31kW). The efficiency is

$$\eta = \frac{\text{output power}}{\text{input power}} \times 100\%$$
$$= \frac{20.8\text{kW}}{31\text{kW}} \times 100 = 67.1\%$$



vi. The average input dc current is 91.2A dc while the resonant bridge rms current is 101.3A rms. By Kirchhoff's current law, the 340V dc rail decoupling capacitor ac current is given by

$$I_{ac} = \sqrt{I_{Prms}^2 - I_{Pave}^2}$$
$$= \sqrt{101.3^2 - 91.2^2} = 44.1\text{A ac}$$

This is the same ac current magnitude as the current in the dc capacitor across the load in the secondary circuit, 22A, when the transformer turns ratio, 2, is taken into account.

vii. The voltage across the load resistance is given by equation (18.13)

$$\frac{v_o}{V_s} = \frac{R_{eq}}{R_{eq} + R_c + j\left(\omega_s L_R - \frac{1}{\omega_s C_R}\right)} = \frac{\frac{8}{\pi^2} \times n_T^2 \times R_{Load}}{\frac{8}{\pi^2} \times n_T^2 \times R_{Load} + R_c + j\left(\omega_s L_R - \frac{1}{\omega_s C_R}\right)}$$

$$\frac{228\text{V}}{340\text{V}} = \left| \frac{\frac{8}{\pi^2} \times ¼ \times 10\Omega}{\frac{8}{\pi^2} \times ¼ \times 10\Omega + 1\Omega + j\left(\omega_s\, 100\mu\text{H} - \frac{1}{\omega_s\, 0.47\mu\text{F}}\right)} \right|$$

which yields $\omega_s \approx \omega_o$ =146krad/s.
Because of the high circuit Q = 14.6, but an infinite Q being assumed to evaluate v_o, and relatively high voltage transfer ratio $v_o / V_s = 0.66$, ω_s is very close to ω_o, as can be deduced from the plots in figure 18.2b. The output voltage control will be very sensitive to changes in the H-bridge switching frequency.

♣

18.5 Resonant-switch, dc-to-dc step-down voltage converters

There are two forms of resonant switch circuit configurations for dc-to-dc converters, namely resonant voltage and resonant current switch commutation. Each type reduces the switching losses to near zero. The technique can be used on any hard switch circuit, for example, resonant switching is applicable to the switch in the buck, boost, buck/boost, Cuk, etc. converters.

- In *resonant current commutation* the switch current is reduced to zero by an *L-C* resonant circuit current greater in magnitude than the load current, such that the switch is turned on and off with zero current.
- In *resonant voltage commutation* the switch voltage is reduced to zero by the capacitor of an *L-C* resonant circuit with a voltage magnitude greater than the output voltage, such that the switch can turn on and off with zero voltage.

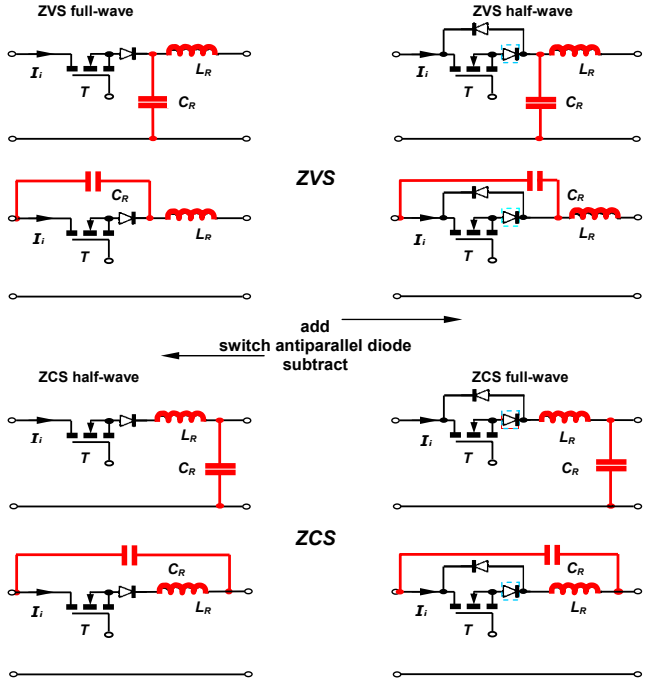


Figure 18.11. *Resonant switches, showing half-wave and full-wave circuits for both ZCS and ZVS.*

Both full-wave or half wave resonance switching is possible, depending on whether half a resonant cycle or a full resonant cycle is allowed by the circuit configuration. Figure 18.11 shows the basic resonant switch building blocks. A common feature, with a dc voltage source, is that the resonant inductor L_R is in series with the switch to be losslessly commutated. Essentially a diode across the switch determines whether full or half wave commutation can occur. In the case of ZCS, the anti-parallel diode allows the resonant current to continue for a full ac resonant cycle, while the same diode prevents a full resonant capacitor voltage cycle since the diode clamps any switch negative voltage to zero.

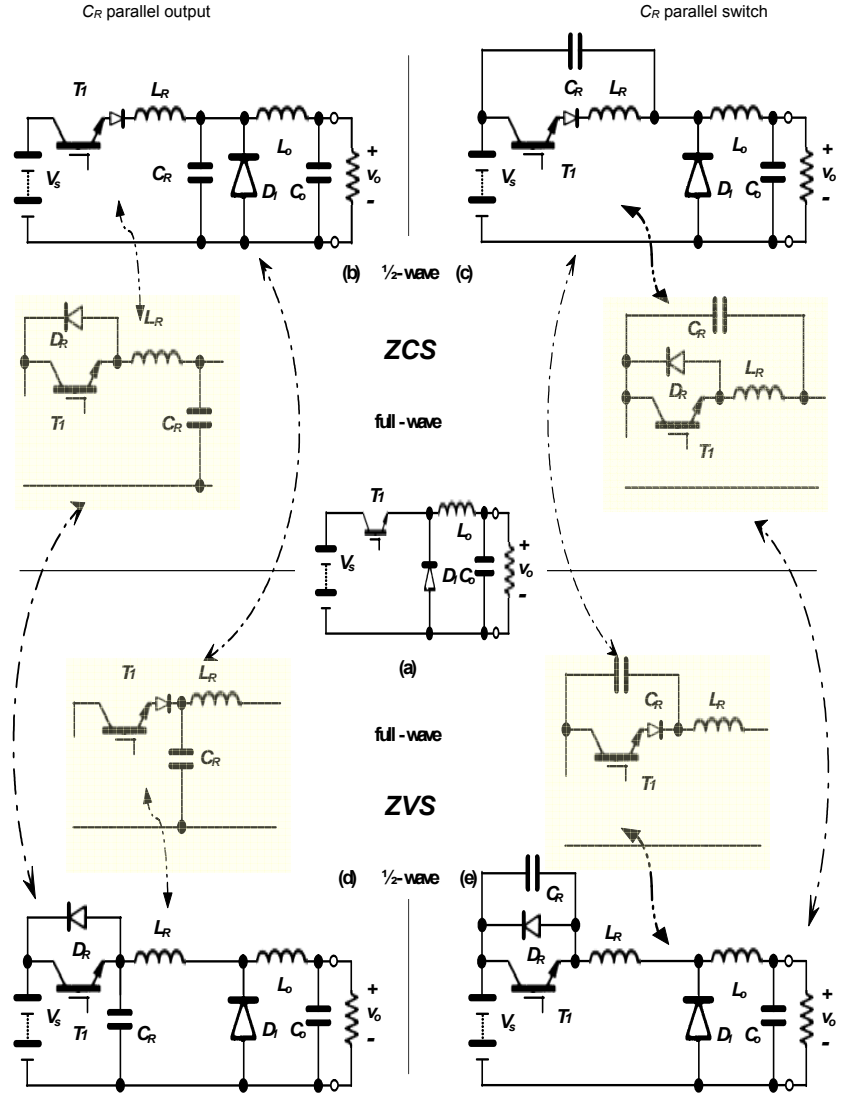


Figure 18.12. *Dc to dc resonant switch step-down converters: (a) conventional switch mode forward step-down converter; (b) and (c) half-wave zero current switching **ZCS** resonant switch converters; and (d) and (e) half-wave zero voltage switching **ZVS** resonant switch converters. Topological translations between half-wave and full-wave versions also shown.*

Although resonant switching is applicable to boost and buck/boost converters, only its application to the buck converter will be specifically analysed. In mitigation, the basic resonant switch possibilities for the buck, boost, and buck/boost converters are shown in matrix form, in the figures 18.20, 18.21, and 18.22 of the Appendix 18.7 at the end of this chapter.

Figure 18.12a shows the basic single switch, forward, step-down voltage switch mode dc-to-dc converter. Resonant switch forward converters are an extension of the standard switch mode forward converter, but the switch is supplement with passive components L_R - C_R to provide resonant operation through the switch, hence facilitating zero current or voltage switching. Common to each circuit is a resonant inductor L_R in series with the switch to be commutated. Parasitic series inductance is therefore not an issue with most resonant switch converters.

An important operational requirement is that the average load current never falls to zero, otherwise the resonant capacitor C_R can never fully discharge in order to fulfil its zero switch voltage/current electrical condition function.

The resonant capacitor C_R, can be either in a parallel or series arrangement as shown in figure 18.12, since small-signal ac-wise, thence resonance wise, the connections are the same. A well-decoupled supply is essential when the capacitor C_R is used in the parallel switch arrangement, as shown in figure 18.12d. A further restriction shown in figure 18.12, is that a diode must be used in series or in anti-parallel with T1 if a switch without reverse blocking capability is used.

The use of a diode D_R in anti-parallel to the switch (shown shaded below figures 18.12 b and c) changes the switching arrangement from half-wave resonant to full-wave resonant operation, where switch reverse voltage block capability is not necessary.

Conversely, removal of the diode D_R in figures 18.12 d and e (shown shaded above figures 18.12 d and e) changes the switching arrangement from half-wave resonant to full-wave resonant operation, where switch reverse voltage block capability is necessary.

Reconnecting the capacitor C_R terminal not associated with V_s, to the other end of inductor L_R in the half-wave circuit in figures 18.12b-e, will create four full-wave resonant switch circuits, with the commutation type, namely ZVS or ZCS, interchanged.

Full and half wave operation is dependent on whether the circuit configuration allows the resonant capacitor to complete a full or half resonant sinusoidal cycle.

Table 18.2: ZCS and ZVS circuit characteristics

Characteristics and properties			ZVS no turn-off loss constant off-time		ZCS no turn-on loss constant on-time	
type	period		I_{cap}	switch	V_{cap}	switch
half-wave	π	½ resonant cycle	current resonates	anti-parallel diode	voltage resonates	reverse blocking
full-wave	2π	1 resonant cycle	current resonates	reverse blocking	voltage resonates	anti-parallel diode
			controlled *di/dt*'s	> twice the dc supply	controlled *dv/dt*'s	> twice the output current

18.5.1 Zero-current, resonant-switch, dc-to-dc converter - ½ wave, C_R parallel with load version

The zero current switching of T1 in figure 18.13 (18.12b) can be analysed in five distinctive stages, as shown in the capacitor voltage and inductor current waveforms in figure 18.13b. The switch is turned on at t_o and turned off after t_4 but before t_5.

Assume the circuit has attained steady state load conditions from one cycle to the next. The cycle commences, before t_o, with both the capacitor voltage and inductor current being zero, and the load current is freewheeling through D1. The output inductor L_o, is large enough such that its current, I_o can be assumed constant. The switch T1 is off.

Time interval I

At t_o the switch is turned on and the series inductor L_R acts as a turn-on snubber for the switch. In the interval t_o to t_1, the supply voltage is impressed across L_R since the switch T1 is on and the diode D1 conducts the output current, thereby clamping the associated inductor terminal to zero volts. Because of the fixed voltage V_s, the current in L_R increases from zero, linearly to I_o in time

$$t_1 = I_o L_R / V_S \tag{18.28}$$

according to

$$i_{L_R}(t) = \frac{V_s}{L_R} t \tag{18.29}$$

During this interval the resonant capacitor voltage is clamped to zero since C_R is in parallel with D1 which is conducting a current decreasing from I_o to zero:

$$v_c(t) = 0 \tag{18.30}$$

Time interval IIA

When the current in L_R reaches I_o at time t_1, the capacitor C_R and L_R are free to resonant. The diode D1 blocks as the voltage across C_R sinusoidally increases. The constant load current component in L_R does not influence its ac performance since a constant inductor current does not produce any inductor voltage. Its voltage is specified by the resonant cycle, provided $I_o < V_s / Z_o$. The capacitor resonantly charges to twice the supply V_s when the inductor current falls back to the load current level I_o, at time t_3.

Time interval IIB

Between times t_3 and t_4 the load current is displaced from L_R by charge from C_R, in a quasi resonance process. The resonant cycle cannot reverse through the switch once the inductor current reaches zero at time t_4, because of the series blocking diode (the switch must have uni-directional conduction characteristics).

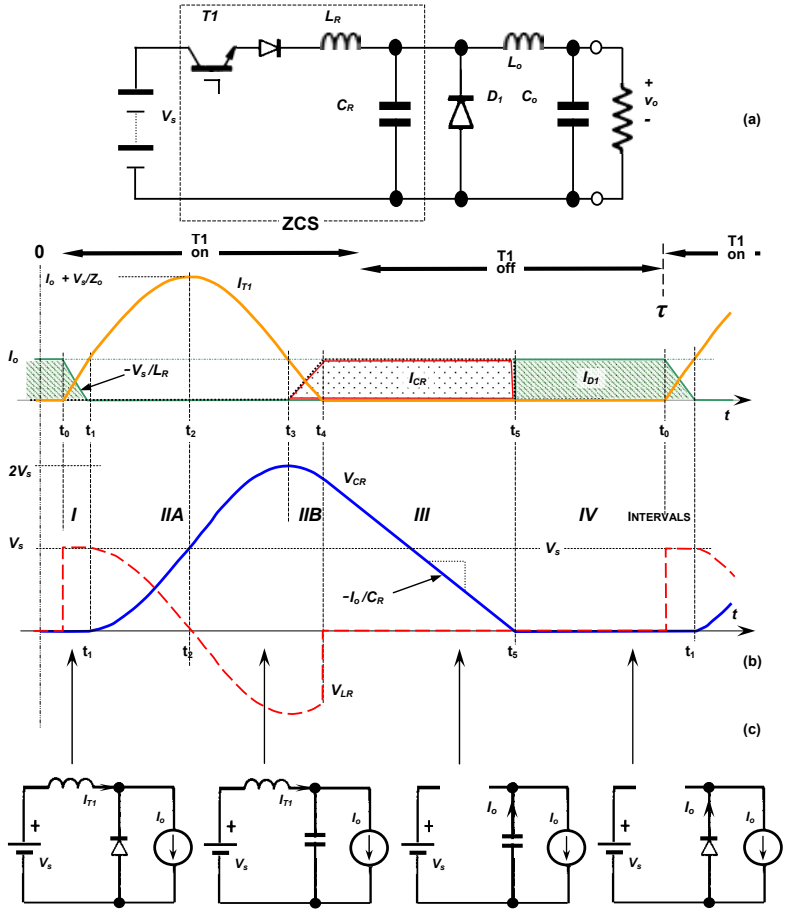


Figure 18.13. *Zero current switching, ZCS, half-wave resonant switch dc to dc converter with the resonant capacitor across the output: (a) circuit; (b) waveforms; and (c) equivalents circuits.*

The capacitor voltage and current for period *IIA* and approximately for period *IIB*, are given by equations 14.60 and 14.61 with the appropriate initial conditions of $i_o = 0$ and $v_o = 0$:

$$v_c(\omega t) = V_s\left(1 - \frac{\omega_o}{\omega} e^{-\alpha t}\cos(\omega t - \phi)\right) \approx V_s\left(1 - e^{-\alpha t}\cos\omega t\right) \quad (18.31)$$

$$i_c(\omega t) = \frac{V_s}{\omega L} \times e^{-\alpha t} \times \sin\omega t \quad (18.32)$$

If the circuit *Q* is high these equations can be approximated by

$$v_{C_R}(t) = V_s(1 - \cos\omega_o t) \quad \text{where} \quad \hat{v}_{C_R} = 2V_s \quad (18.33)$$

$$i_{C_R}(t) = \frac{V_s}{Z_o}\sin\omega_o t \quad (18.34)$$

The inductor current is the constant load current plus the capacitor current:

$$i_{L_R}(t) = I_o + i_{C_R}(t) = I_o + \frac{V_s}{Z_o}\sin\omega_o t \quad \text{where} \quad \hat{i}_{T1} = \hat{i}_{L_R} = I_o + \frac{V_s}{Z_{o\,s}} \quad (18.35)$$

where $Z_o = \sqrt{L_R / C_R}$ and $\omega_o = 1/\sqrt{L_R C_R}$. Equation (18.35) shows that the inductor current only returns to zero if $I_o < V_s / Z_o$, otherwise the switch is commutated with a non-zero current flow.

Setting $i_L = 0$ in equation (18.35) gives the time for period *II* as

$$t_{II} = \left(\pi + \sin^{-1}\left(\frac{I_o Z_o}{V_s}\right)\right)/\omega_o \quad (18.36)$$

after which time the capacitor voltage and inductor current reach

$$V_{C_R\,t4} = V_s\left(1 + \sqrt{1 - \left(\frac{I_o Z_o}{V_s}\right)^2}\right) \quad (18.37)$$

$$I_{L_R\,t4} = 0$$

Time interval III

At time t_4 the input current is zero and the switch T1 can be turned off with zero current, ZCS. The constant load current requirement I_o is provided by the capacitor, which discharges linearly to zero volts at time t_5 according to

$$v_{C_R}(t) = V_{C_R\,t4} - \frac{I_o}{C_R} \times t = V_s\left(1 + \sqrt{1 - \left(\frac{I_o Z_o}{V_s}\right)^2}\right) - \frac{I_o}{C_R} \times t \quad (18.38)$$

where $V_{C_R\,t4}$ is given by equation (18.37).

The inductor current is

$$i_{LR}(t) = 0 \quad (18.39)$$

The time for interval *III* is load current dependant and is given by setting equation (18.38) to zero:

$$t_{III} = \frac{V_{C_R\,t4}\,C_R}{I_o} = \frac{C_R}{I_o} \times V_s\left(1 + \sqrt{1 - \left(\frac{I_o Z_o}{V_s}\right)^2}\right) \quad (18.40)$$

Time interval IV

After t_5, the switch is off, the current freewheels through D1, the capacitor voltage is zero, and the input inductor current is zero. At time t_1 the cycle recommences. The switch off-time, interval *IV*, t_5 to the subsequent t_0, is used to control the rate at which energy is transferred to the load.

Output voltage

The output voltage can be specified by either evaluating the energy from the supply, through the input resonant inductor L_R, or by evaluating the average voltage across the resonant capacitor C_R (or the freewheel diode D_1) which is filtered by the output filter L_o - C_o.

By considering the input inductor energy (volt-second integral) for each period shown in the waveforms in figure 18.13b, the output energy, whence voltage, is given by

$$v_o = \frac{V_s}{\tau}\left(\tfrac{1}{2}t_I + t_{II} + t_{III}\right) = \frac{V_s}{\tau} \times \left(\frac{1}{\omega_o}\left(\pi + \sin^{-1}\left(\frac{I_o Z_o}{V_s}\right)\right) + \tfrac{1}{2}\frac{I_o L_R}{V_S} + V_s\left(1 + \sqrt{1 - \left(\frac{I_o Z_o}{V_s}\right)^2}\right) \times \frac{C_R}{I_o}\right) \quad (18.41)$$

where the time intervals *I*, *II*, and *III* are given by equations (18.28), (18.36), and (18.40) respectively, the switching frequency $f_s = 1/\tau$, and $\tau > t_I + t_{II} + t_{III}$.

The output voltage based on the average capacitor voltage (after resetting time zero references) is

$$v_o = \frac{1}{\tau}\left[\int_0^{t_5 - t_4} V_s(1 - \cos\omega t)\,dt + \int_0^{t_5 - t_4} V_{C_R\,t4}\left(1 - \frac{t}{t_5 - t_4}\right)dt\right] = \frac{1}{\tau} \times \left[\frac{V_s}{\omega_o}\left(\pi + \sin^{-1}\left(\frac{I_o Z_o}{V_s}\right) + \frac{I_o Z_o}{V_s}\right) + \tfrac{1}{2} \times V_s^2 \times \sqrt{1 - \left(\frac{I_o Z_o}{V_s}\right)^2} \times \left(1 + \sqrt{1 - \left(\frac{I_o Z_o}{V_s}\right)^2}\right) \times \frac{C_R}{I_o}\right] \quad (18.42)$$

The output voltage in equation (18.42) reduces to equation (18.41).

The minimum switch commutation period is $t_I + t_{II} + t_{III}$ which limits the upper operating frequency, hence maximum output voltage. The output voltage,which is less than the input voltage, is inversely related to the switching frequency.

The circuit has a number of features:

i. Turn-on and turn-off occur at zero current, hence switching losses are minimal.
ii. Increasing the switch off period (interval *VI*) decreases the average output voltage.
iii. At light load currents the switching frequency may become extreme low.
iv. The capacitor discharge time is $t_{III} \le V_{C_R\,t4} \times C_R / I_o$, thus the output voltage is load current dependant.
v. L_R and C_R are dimensioned such that the capacitor voltage is greater than V_s at time t_4, at maximum load current I_o.
vi. Supply inductance is inconsequential, and decreases the inductance L_R requirement.
vii. Being based on the forward converter, the output voltage is less than the input voltage. The output increases with increased switching frequency.
viii. If a diode in antiparallel to the switch is added as shown below figure 18.12b, reverse inductor current can flow and the output voltage is $v_o \approx V_s \times f_s / f_o$. A full-wave resonant zero current switch circuit is formed.

18.5.1i - Zero-current, full-wave resonant switch converter

By adding a diode in anti-parallel to the switch, as shown in figure 18.14 (and the circuit below figure 18.12b), resonant action can continue beyond $\omega t \ge \pi$.

Assume the circuit has attained steady state load conditions from one cycle to the next. The cycle commences, before t_o, with both the capacitor voltage and inductor current being zero, and the load current is freewheeling through D1. The output inductor L_o, is large enough such that its current, I_o can be assumed constant. The switch T1 is off.

Time interval I

At t_o the switch is turned on and the series inductor L_R acts as a turn-on snubber for the switch. In the interval t_o to t_1, the supply voltage is impressed across L_R since the switch T1 is on and the diode D1 conducts the output current, thereby clamping the associated inductor terminal to V_s. Because of the fixed voltage V_s, the current in L_R increases from zero, linearly to I_o in time

$$t_I = I_o L_R / V_S = \frac{1}{\omega_o}\frac{Z_o I_o}{V_s} \quad (18.43)$$

according to

$$i_{L_R}(t) = \frac{V_s}{L_R}t \quad \text{and also} \quad i_{D1_R}(t) = I_o - i_{L_R}(t) = I_o - \frac{V_s}{L_R}t \quad (18.44)$$

During this interval the resonant capacitor voltage is clamped to -V_s (with respect to input voltage positive terminal) since C_R is in parallel with L_R which is conducting I_o:

$$v_c(t) = -V_s \tag{18.45}$$

Time interval II

When the current in L_R reaches I_o at time t_1, the capacitor C_R and L_R are free to resonant. The diode D1 blocks as the voltage across C_R sinusoidally decreases. The constant load current component in L_R does not influence its ac performance since a constant inductor current does not produce any inductor voltage. Its voltage is specified by the resonant cycle, provided $I_o < V_s / Z_o$. The capacitor resonantly charges to the opposite polarity $+V_s$ when the inductor current falls back to the load current level I_o, at time t_3.

Between times t_3 and t_4 the load current is displaced from L_R by charge from C_R, in a quasi resonance process. The resonant cycle reverses through the switch parallel diode D_R once the inductor current reaches zero at time t_4.

Assuming a high circuit Q, the capacitor voltage and inductor current for period *II*, are given by

$$v_{C_R}(t) = V_s(1 - \cos\omega_o t) \tag{18.46}$$

$$i_{L_R}(t) = I_o + i_{C_R}(t) = I_o + \frac{V}{Z_o}\sin\omega_o t \tag{18.47}$$

where $Z_o = \sqrt{L_R / C_R}$ and $\omega_o = 1/\sqrt{L_R C_R}$. Equation (18.47) shows that the inductor current only returns to zero if $I_o < V_s / Z_o$, otherwise the switch is commutated with non-zero current flow.

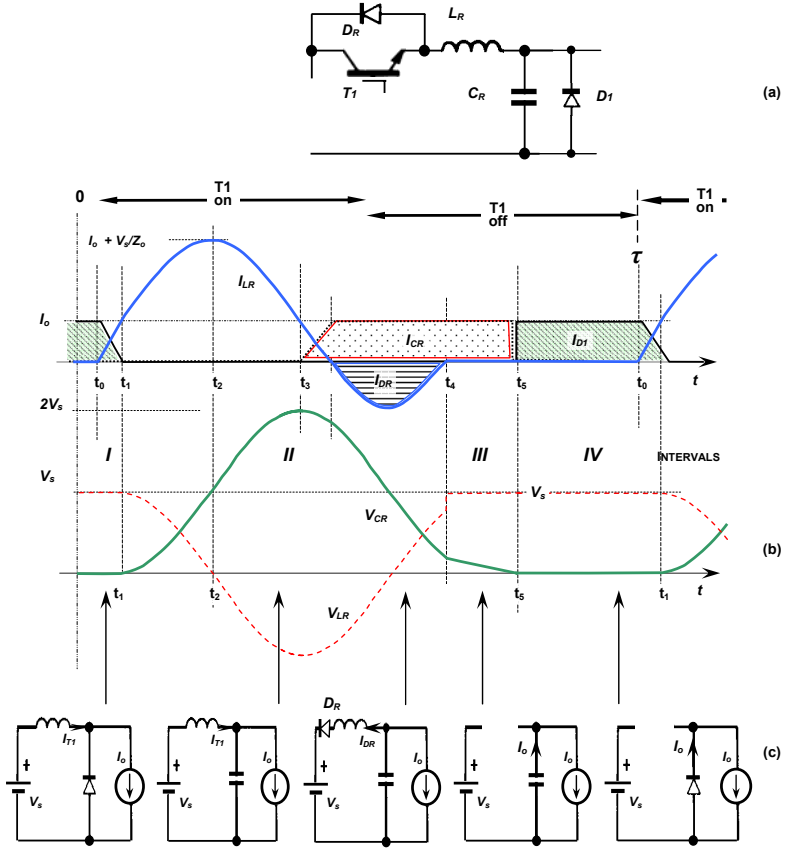


Figure 18.14. *Zero current switching, ZCS, full-wave resonant switch dc to dc converter with the resonant capacitor across the output: (a) circuit; (b) waveforms; and (c) equivalents circuits.*

The peak inductor current hence maximum switch current, from equation (18.47), is

$$\hat{i}_{T1} = \hat{i}_{L_R} = I_o + \frac{V_s}{Z_o} \tag{18.48}$$

By adding a diode in anti-parallel to the switch, resonant action can continue beyond $\omega t \geq \pi$.

The capacitor can resonant to a lower voltage level, hence the capacitor linear discharge period starts from a lower voltage, equation (18.49).

$$V_{C_R t4} = V_s\left(1 + \sqrt{1 - \left(\frac{I_o Z_o}{V_s}\right)^2}\right) \tag{18.49}$$

The lower limit of load current for proper circuit action is therefore decreased with full wave resonant circuits. Equations (18.31) to (18.42) remain valid except the time for interval *II* is extended to the fourth quadrant where $i_L = 0$ and the capacitor voltage at t_4 is decreased. That is

$$t_{II} = \left(2\pi - \sin^{-1}\left(I_o Z_o / V_s\right)\right) / \omega_o \tag{18.50}$$

Time interval III

Before time t_4 the input current is zero and the switch T1 can be turned off with zero current, ZCS. The constant load current requirement I_o is provided by the capacitor, which discharges linearly to zero volts at time t_5 according to

$$v_{C_R}(t) = V_{C_R t4} - \frac{I_o}{C_R} \times t = V_s\left(1 + \sqrt{1 - \left(\frac{I_o Z_o}{V_s}\right)^2}\right) - \frac{I_o}{C_R} \times t \tag{18.51}$$

where $V_{C_R t4}$ is given by equation (18.49).

The inductor current is

$$i_{L_R}(t) = 0 \tag{18.52}$$

The time for interval *III* is load current dependant and is given by setting equation (18.51) to 0:

$$t_{III} = \frac{V_{C_R t4}\, C_R}{I_o} = \frac{C_R}{I_o} \times V_s\left(1 + \sqrt{1 - \left(\frac{I_o Z_o}{V_s}\right)^2}\right) \tag{18.53}$$

Time interval IV

After t_5, the switch is off, the current freewheels through D1, the capacitor voltage is zero, and the input inductor current is zero. At time t_o the cycle recommences. The switch off-time, interval *IV*, t_5 to the subsequent t_o, is used to control the rate at which energy is transferred to the load.

Output voltage

Since switch turn-off is dependent on the resonant cycle, the output voltage does not depend on the duty cycle, but is resonant period depend according to

$$v_o = V_s \frac{2\pi\sqrt{L_R C_R}}{\tau} = V_s \frac{2\pi}{\omega_o \tau} \tag{18.54}$$

18.5.2 Zero-current, resonant-switch, dc-to-dc converter - ½ wave, C_R parallel with switch version

Operation of the ZCS circuit in figure 18.15 (figure 18.12c), where the capacitor C_R is connected in parallel with the switch, is essentially the same as the circuit in figure 18.13. The capacitor connection produces the result that the capacitor voltage has a dc offset of V_s, meaning its voltage swings between $\pm V_s$ rather than zero and twice V_s, as in the circuit in figure 18.13.

The zero current switching of T1 in figure 18.15 is analysed in five distinctive stages, as shown in the capacitor voltage and inductor current waveforms in figure 18.15b. The switch is turned on at t_o and turned off after t_4 but before t_5.

Assume the circuit has attained steady state load conditions from one cycle to the next. The cycle commences, before t_o, with the inductor current being zero, the capacitor charged to V_s with the polarity shown, and the load current freewheeling through D1. The output inductor L_o, is large enough such that its current, I_o can be assumed constant. The switch T1 is off.

Time interval I

At t_o the switch is turned on and the series inductor L_R acts as a turn-on snubber for the switch. In the interval t_o to t_1, the supply voltage is impressed across L_R since the switch T1 is on and the diode D1

conducts the output current, thereby clamping the associated inductor terminal to V_s. Because of the fixed voltage V_s, the current in L_R increases from zero, linearly to I_o in time

$$t_I = I_o L_R / V_S = \frac{1}{\omega_o}\frac{Z_o I_o}{V_s} \tag{18.55}$$

according to

$$i_{L_R}(t) = \frac{V_s}{L_R} t \tag{18.56}$$

During this interval the resonant capacitor voltage is clamped to -V_s (with respect to input voltage positive terminal) since C_R is in parallel with L_R which is conducting I_o:

$$v_c(t) = -V_s \tag{18.57}$$

Time interval II

When the current in L_R reaches I_o at time t_1, the capacitor C_R and L_R are free to resonant. The diode D1 blocks as the voltage across C_R sinusoidally decreases. The constant load current component in L_R does not influence its ac performance since a constant inductor current does not produce any inductor voltage. Its voltage is specified by the resonant cycle, provided $I_o < V_s / Z_o$. The capacitor resonantly charges to the opposite polarity +V_s when the inductor current falls back to the load current level I_o, at time t_3.

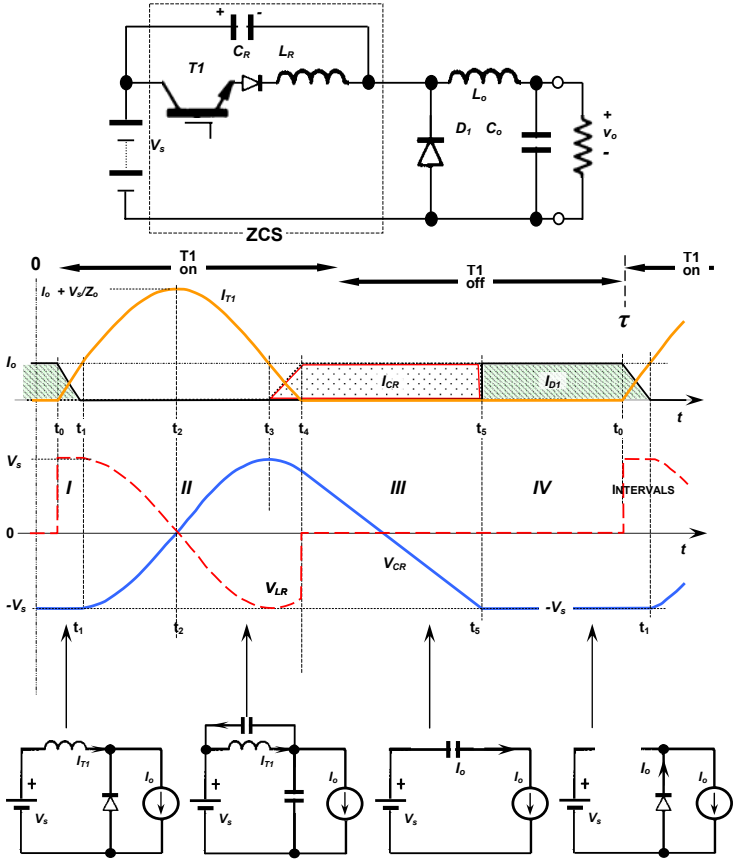


Figure 18.15. *Zero current switching, ZCS, half-wave resonant switch dc to dc converter with resonant capacitor across the switch: (a) circuit; (b) waveforms; and (c) equivalents circuits.*

Between times t_3 and t_4 the load current is displaced from L_R by charge from C_R, in a quasi resonance process. The resonant cycle cannot reverse through the switch once the inductor current reaches zero at time t_4, because of the series blocking diode (the switch must have uni-directional conduction characteristics).

Assuming a high circuit *Q*, the capacitor voltage and inductor current for period *II*, are given by

$$v_{C_R}(t) = -V_s \cos \omega_o t \tag{18.58}$$

$$i_L(t) = I_o + i_{C_R}(t) = I_o + \frac{V}{Z_o}\sin \omega_o t \tag{18.59}$$

where $Z_o = \sqrt{L_R / C_R}$ and $\omega_o = 1/\sqrt{L_R C_R}$. Equation (18.59) shows that the inductor current only returns to zero if $I_o < V_s / Z_o$, otherwise the switch is commutated with non-zero current flow.

Setting $i_L = 0$ in equation (18.59) gives the time for period *II* as

$$t_{II} = \left(\pi + \sin^{-1}\left(I_o Z_o / V_s\right)\right)/\omega_o \tag{18.60}$$

at which time the capacitor voltage and inductor current are

$$\begin{aligned} V_{C_R t4} &= V_s\sqrt{1-\left(\frac{I_o Z_o}{V_s}\right)^2} \\ I_{L_R t4} &= 0 \end{aligned} \tag{18.61}$$

Time interval III

At time t_4 the input current is zero and the switch T1 can be turned off with zero current, ZCS. The constant load current requirement I_o is provided by the capacitor, which discharges linearly to - V_s volts at time t_5 according to

$$v_{C_R}(t) = V_{C_R t4} - \frac{I_o}{C_R}\times t = V_s\sqrt{1-\left(\frac{I_o Z_o}{V_s}\right)^2} - \frac{I_o}{C_R}\times t \tag{18.62}$$

where $V_{C_R t4}$ is given by equation (18.61).

The inductor current is

$$i_{LR}(t) = 0 \tag{18.63}$$

The time for interval *III* is load current dependant and is given by setting equation (18.62) to - V_s:

$$t_{III} = \frac{V_{C_R t4}\, C_R}{I_o} = \frac{C_R}{I_o}\times V_s\left(1+\sqrt{1-\left(\frac{I_o Z_o}{V_s}\right)^2}\right) \tag{18.64}$$

Time interval IV

After t_5, the switch is off, the current freewheels through D1, the capacitor voltage is - V_s, and the input inductor current is zero. At time t_1 the cycle recommences. The switch off-time, interval *IV*, t_5 to the subsequent t_0, is used to control the rate at which energy is transferred to the load.

Output voltage

By considering the input inductor energy (volt-second integral) for each period shown in the waveforms in figure 18.15b, the output energy, whence voltage, is given by

$$\begin{aligned} v_o &= \frac{V_s}{\tau}\left(\tfrac{1}{2}t_I + t_{II} + t_{III}\right) \\ &= \frac{V_s}{\tau}\times\left(\frac{1}{\omega_o}\left(\pi + \sin^{-1}\left(\frac{I_o Z_o}{V_s}\right)\right) + \tfrac{1}{2}\frac{I_o L_R}{V_S} + V_s\left(1+\sqrt{1-\left(\frac{I_o Z_o}{V_s}\right)^2}\right)\times\frac{C_R}{I_o}\right) \end{aligned} \tag{18.65}$$

where the time intervals *I*, *II*, and *III* are given by equations (18.55), (18.60), and (18.64) respectively, the switching frequency $f_s = 1/\tau$, and $\tau > t_I + t_{II} + t_{III}$.

The output voltage based on the average diode voltage (after resetting time zero references) is

$$\begin{aligned} v_o &= \frac{1}{\tau}\left[\int_0^{t_4-t_1} V_s\left(1-\cos\omega t\right)dt + \int_0^{t_5-t_4} V_{C_R t4}\left(1-\frac{t}{t_5-t_4}\right)dt\right] \\ &= \frac{1}{\tau}\times\left[\frac{V_s}{\omega_o}\left(\pi + \sin^{-1}\left(\frac{I_o Z_o}{V_s}\right)\right) + \frac{I_o Z_o}{V_s}\right) + \tfrac{1}{2}\times V_s^2\times\sqrt{1-\left(\frac{I_o Z_o}{V_s}\right)^2}\times\left(1+\sqrt{1-\left(\frac{I_o Z_o}{V_s}\right)^2}\right)\times\frac{C_R}{I_o}\right] \end{aligned} \tag{18.66}$$

The output voltage in equation (18.66) reduces to equation (18.65).

18.5.3 Zero-voltage, resonant-switch, dc-to-dc converter - ½ wave, C_R parallel with switch version

The zero voltage switching of T1 in figure 18.16 (18.12e) can be analysed in four distinctive stages, as shown in the resonant capacitor voltage and inductor current waveforms. The switch is turned off at t_o and turned on after t_4 but before t_5.

The circuit has attained steady state load conditions from one cycle to the next. The cycle commences, before t_o, with the capacitor C_R voltage being zero and the load current I_o being conducted by the switch and the resonant inductor, L_R. The output inductor L_o is large enough such that its current, I_o can be assumed constant. The switch T1 is on.

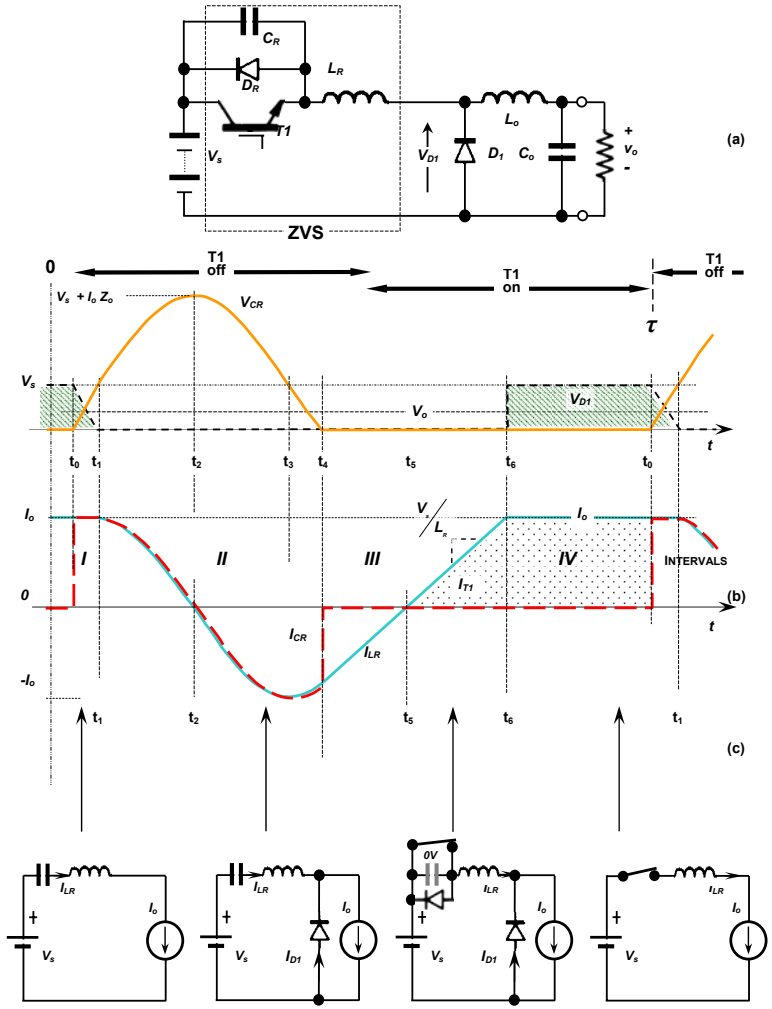


Figure 18.16. *Zero voltage switching, ZVS, half-wave, resonant switch dc to dc converter: (a) circuit; (b) waveforms; and (c) equivalents circuits.*

Time interval I

At time t_o the switch is turned off and the parallel capacitor C_R acts as a turn-off snubber for the switch. In the interval t_o to t_1, the supply current is provided from V_s through C_R and L_R. Because the load current is constant, I_o, due to large L_o, the capacitor charges linearly from 0V until its voltage reaches V_s in time

$$t_1 = \frac{V_s C_R}{I_o} = \frac{V_s}{I_o}\frac{1}{\omega_o Z_o} \tag{18.67}$$

where $Z_o = \sqrt{\frac{L_R}{C_R}}$ and $\omega_o = \sqrt{L_R C_R}$

according to

$$v_c(t) = \frac{I_o}{C_R} \times t = V_s \times \frac{t}{t_1} \tag{18.68}$$

The inductor current is zero, that is

$$i_{L_R}(t) = 0 \tag{18.69}$$

The freewheel diode voltage, which is related to the output voltage, is given by

$$V_{D1} = V_s - v_c = V_s - \frac{I_o}{C_R} \times t = V_s \times \left(1 - \frac{t}{t_1}\right) \tag{18.70}$$

Time interval II

When the voltage across C_R reaches V_s at time t_1, (equation (18.67)), the load freewheel diode conducts, clamping the load voltage to zero volts. The capacitor C_R and L_R are free to resonant, where the initial inductor current is I_o and the initial capacitor voltage is V_s. The energy in the inductor transfers to the capacitor, which increases its voltage from V_s to a maximum, at time t_2, of

$$\hat{v}_C = V_s + I_o Z_o \tag{18.71}$$

The capacitor energy transfers back to the inductor which has resonated from + I_o to – I_o between times t_1 to time t_3. For the capacitor voltage to resonantly return to *zero*, $I_o > V_s / Z_o$. Between t_3 and t_4 the voltage V_s on C_R is resonated through L_R, which conducts – I_o at t_3, as part of the resonance process. Assuming a high circuit *Q*, the resonant capacitor voltage and inductor current during period *II* are given by

$$\begin{aligned} v_{C_R}(t) &= V_s + I_o Z_o \sin \omega_o t \\ i_{L_R}(t) &= I_o \cos \omega_o t \end{aligned} \tag{18.72}$$

and the duration of interval *II* is

$$t_{II} = \left(\pi + \sin^{-1}\frac{V_s}{I_o Z_o}\right) / \omega_o \tag{18.73}$$

At the end of interval *II* the capacitor voltage is zero and the inductor and capacitor currents are

$$i_{C_R}(t_4) = i_{L_R}(t_4) = I_o \cos \omega_o t_{II} = -I_o \sqrt{1 - \left(\frac{V_s}{I_o Z_o}\right)} \tag{18.74}$$

and

$$v_{C_R}(t_{II}) = 0 \tag{18.75}$$

The freewheel diode current at the end of interval *II* is

$$i_{D1}(t_4) = I_o + I_o \sqrt{1 - \left(\frac{V_s}{I_o Z_o}\right)} \tag{18.76}$$

Time interval III

At time t_4 the voltage on C_R attempts to reverse, but is clamped to zero by diode D_R. The inductor energy is returned to the supply V_s via diode D_R and the freewheel diode D_1. The inductor current decreases linearly to zero during the period t_4 to t_5. ***During this period the switch T1 is turned on.***

No turn-on losses occur because the diode D_R in parallel with T1 is conducting during the period the switch is turned on, that is, the switch voltage is zero and the switch T1 can be turned on with zero voltage, ZVS. With the switch on at time t_5 the current in the inductor L_R reverses and builds up, linearly to I_o at time t_6. The current slope is supply V_s dependant, according to $V_s = L_R\, di/dt$, that is

$$i_{L_R}(t) = i_{D_R}(t) = \frac{V_s}{L_R} t + I_o \cos \omega_o t_{II} \tag{18.77}$$

and the time of interval *III* is load current dependant:

$$t_{III} = \frac{I_o L_R}{V_s} \times (1 - \cos \omega_o t_{II}) \qquad (18.78)$$

The freewheel diode current is given by

$$i_{D1}(t) = I_o + i_{D_R}(t) \qquad (18.79)$$

Time interval IV

At t_6, the supply V_s provides all the load current through the switch resonant inductor, and the diode D_1 recovers with a controlled *di/dt* given by V_s / L_R. The freewheel diode D_f supports the supply voltage V_s. The switch conduction interval *IV*, t_6 to the subsequent t_o when the switch is turned off, is used to control the rate at which energy is transferred to the load.

Output voltage

The output voltage, which is always less than the input voltage, can be derived from the diode voltage (shown hatched in figure 18.16b) since this voltage is averaged by the output *L-C* filter.

$$\begin{aligned} v_o &= \frac{1}{\tau}(\text{Volt} \times \text{second area of interval } I + \text{Volt} \times \text{second area of interval } IV) \\ &= \frac{V_s}{\tau}(½t_1 + \tau - t_6) = V_s(1 - f_s(t_6 - ½t_1)) \\ &= 2\pi \frac{V_s}{\omega_o \tau}\left(1 - \left(½ + \frac{I_o Z_o}{2\pi V_s}\right)\right) \end{aligned} \qquad (18.80)$$

The circuit has a number of features:

i. Switch turn-on and turn-off both occur at zero voltage, hence switching losses are minimal.
ii. At light load currents the switching frequency may become extreme high.
iii. The inductor defluxing time is $t_{III} \le I_{L_R III} \times L_R / V_s$, hence the output voltage is load current dependant.
iv. L_R and C_R are dimensioned such that the inductor current is less than zero (being returned to the supply V_s) at time t_5, at maximum load current I_o. Also $I_o > V_s / Z_o$.
v. Being based on the forward converter, the output voltage is less than the input voltage. Increasing the switching frequency decreases the output voltage since $\tau - t_5$ is decreased in equation (18.80).

18.5.3i - Zero-voltage, full-wave resonant switch converter

By removing the supply return diode in the half-wave ZVS converter in figure 18.12e (figure 18.16) a full wave ZVS resonant converter is formed, where the capacitor sinusoidal oscillation can continue past π, t_4, as shown in figure 18.17. Consequently, the inductor current attains a level closer to the load level, I_o, before the capacitor voltage oscillation is complete, thereby shortening the cycle time.

18.5.4 Zero-voltage, resonant-switch, dc-to-dc converter - ½ wave, C_R parallel with load version

Operation of the ZVS circuit in figure 18.12d, where the capacitor C_R is connected in parallel with the load circuit (the freewheel diode D_1), is essentially the same as the circuit in figure 18.17. The capacitor connection produces the result that the capacitor voltage has a dc offset of V_s, meaning its voltage swings between + V_s and -I_o Z_o, rather than zero and V_s - I_o Z_o, as in the ZVS circuit considered in 18.5.3. Specifically the inductor waveforms and expressions are unchanged, as is the output voltage expression (18.80). The expression for the time of each interval is the same and the capacitor voltage waveform equations are negated, with V_s then added.
Any dc supply inductance must be decoupled when using the ZVS circuit in figure 18.12d.

It will be noticed that, at a given load current I_o, a ZCS converter has a predetermined on-time, while a ZVS converter has a predetermined off-time.

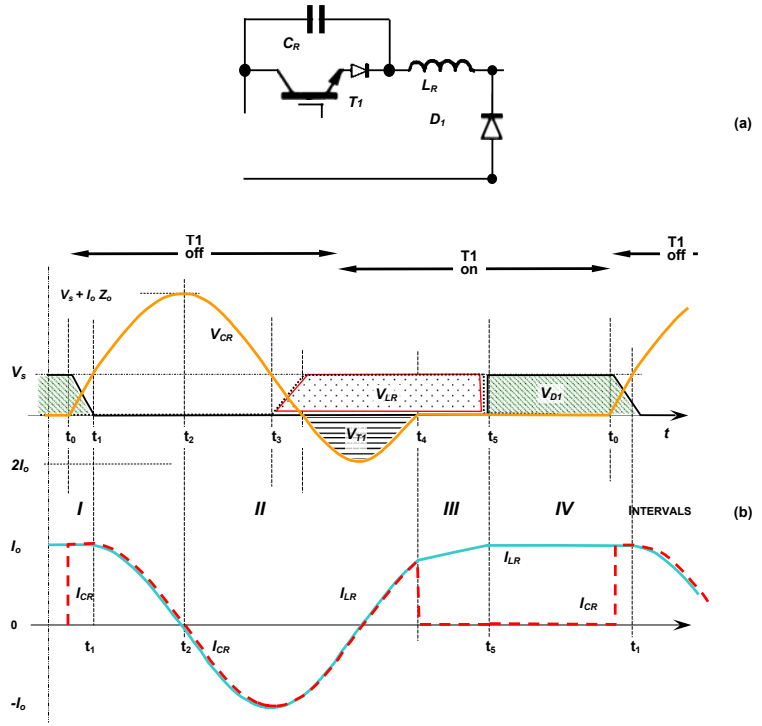


Figure 18.17. *Zero voltage switching, ZVS, full-wave resonant switch dc to dc converter with the resonant capacitor across the switch: (a) circuit and (b) waveforms.*

Example 18.2: ***Zero-current, resonant-switch, step-down dc-to-dc converter - ½ wave***

The ZCS resonant dc step-down voltage converter in figure 18.13a produces an output voltage for the armature of a high voltage dc motor and operates from the voltage produced from the 50Hz ac mains rectified, 340V dc, with an *L-C* dc link filter. The resonant circuit parameters are L_R = 100μH, C_R = 0.47μF, and the high frequency ac resistance of the resonant circuit is R_c = 1Ω.
Calculate

i. the circuit Z_o, Q, and ω_o
ii. the maximum output current to ensure ZCS occurs
iii. the maximum operating frequency, represented by the time between switch turn on and the freewheel diode recommencing conduction, at maximum load current
iv. the average diode voltage (capacitor voltage), hence load voltage at the maximum frequency
v. switching frequency for v_o = 170V dc and R_L = 17Ω, peak input current, and diode maximum reverse voltage.

Solution

i. The characteristic impedance is given by

$$Z_o = \sqrt{\frac{L_R}{C_R}} = \sqrt{\frac{100\mu H}{0.47\mu F}} = 14.6\Omega$$

The resonant circuit Q is

$$Q = \frac{Z_o}{R_c} = \sqrt{\frac{100\mu H}{0.47\mu F}} / 1\Omega = 14.6$$

For this high Q, the circuit resonant frequency and damped frequency will be almost the same, that is

$$\omega \approx \omega_o = 1/\sqrt{L_R C_R}$$
$$= 1/\sqrt{100\mu H \times 0.47\mu F} = 146 \text{ krad/s} \quad = 2\pi f$$
$$f = 146 \text{ krad/s}/2\pi = 23.25 \text{ kHz} \quad \text{or} \quad T = 43\mu s$$

ii. For zero current switching, the load current must not be greater than the peak resonant current, that is

$$I_o < V_s / Z_o = 340V/14.6\Omega = 23.3A$$

iii. The commutation period comprises the four intervals, *I* to *IV*, shown in figure 18.13b.

Interval *I*
The switch turns on and the inductor current rises from 0A to 23.3A in a time given by

$$t_I = L_R \Delta I / V_s$$
$$= 100\mu H \times 23.3A/340V = 6.85\mu s$$

Interval *II*
These two sub-intervals take over half a resonant cycle to complete. Assuming action is purely sinusoidal resonance then from equation (18.36)

$$t_{II} = \left(\pi + \sin^{-1}\left(I_o Z_o / V_s\right)\right)/\omega_o$$
$$= \left(\pi + \sin^{-1}\left(23.3A \times 14.6\Omega / 340V\right)\right)/146\text{krad/s} = 32.27\mu s$$

The capacitor voltage at the end of this period is given by

$$V_{C_R t4} = V_s(1 - \cos\omega_o t_{II})$$
$$= 340V \times (1 - \cos \tfrac{1}{2}\pi) = 340V$$

Interval *III*
The capacitor voltage must discharge from 340V dc to zero volts, providing the 23.3A load current. That is

$$t_{III} = V_{C_R t4} \times C_R / I_o$$
$$= 340V \times 0.47\mu F/23.3A = 6.86\mu s$$

The minimum commutation cycle time is therefore 6.85+32.27+6.86 = 46μs. Thus the maximum operating frequency is 21.75kHz.

iv. The output voltage v_o is the average reverse voltage of freewheel diode D_1, which is in parallel with the resonant capacitor C_R. Integration of the capacitor voltage shown in figure 18.13b gives equation (18.42)

$$v_o = \frac{1}{t_5}\left[\int_0^{t_4 - t_1} V_s(1 - \cos\omega t)\,dt + \int_0^{t_5 - t_4} V_{C_R t4}\left(1 - \frac{t}{t_5 - t_4}\right)dt\right]$$
$$= \frac{1}{46\mu s} \times \left[\int_o^{32.27\mu s} 340V \times (1 - \cos\omega t)\,d\omega t + \int_0^{6.86\mu s} 340V \times \left(1 - \frac{t}{6.86\mu s}\right)dt\right]$$
$$= \frac{1}{46\mu s} \times \left[340V \times \left(\frac{3\pi}{2} + 1\right) \times \frac{43\mu s}{2\pi} \quad + \tfrac{1}{2} \times 340V \times 6.86\mu s\right]$$
$$= \frac{1}{46\mu s} \times \left[13292V\mu s \quad + \quad 1166V\mu s\right] = 314.3Vdc$$

The maximum output voltage is 314V dc. Alternatively, using the input inductor energy based equation (18.41):

$$v_o = \frac{V_s}{\tau}\left(\tfrac{1}{2}t_I + t_{II+III} + t_{IV}\right)$$
$$= \frac{340V}{46\mu s} \times \left(\tfrac{1}{2} \times 6.85\mu s + 32.25\mu s + 6.86\mu s\right) = 314.4V$$

v. When the output current is v_o/R_L=170V/17Ω=10A, the operating frequency is obtained from equation (18.41)

$$v_o = \frac{V_s}{\tau} \times \left(\tfrac{1}{2}\frac{I_o L_R}{V_S} + \frac{\left(\pi + \sin^{-1}\left(\frac{I_o Z_o}{V_s}\right)\right)}{\omega_o} + V_s\left(1 + \sqrt{1 - \left(\frac{I_o Z_o}{V_s}\right)^2}\right) \times \frac{C_R}{I_o}\right)$$

$$170V = \frac{340V}{\tau} \times \left(\tfrac{1}{2} \times \frac{10A \times 100\mu H}{340V} + \frac{\left(\pi + \sin^{-1}\left(\frac{10A \times 14.6\Omega}{340V}\right)\right)}{146\text{krad/s}} + 340V \times \left(1 + \sqrt{1 - \left(\frac{10A \times 14.6\Omega}{340V}\right)^2}\right) \times \frac{0.47\mu F}{10A}\right)$$

That is, τ = 108.2μs, or f_s = 9.25kHz.
The peak input current is the peak inductor current is

$$\hat{I}_{i/p} = \hat{I}_{L_R} = I_o + \frac{V_s}{Z_o} = 10A + \frac{340V}{14.6\Omega} = 33.3A$$

The diode peak reverse voltage is 2×V_s = 640V

♣

Example 18.3: ***Zero-current, resonant-switch, step-down dc-to-dc converter – full-wave***

In example 18.2, a diode is connected in anti-parallel with the switch (figure 18.14), forming a quasi resonant full-wave switch, dc converter. Using the data in example 18.2:

i. Determine the maximum operating frequency with a 10A load current.
ii. Repeat the calculations when an infinite *Q* is not assumed.

Solution

Using the data in example 18.2:

$$\omega_o = 1/\sqrt{L_R C_R} \quad = 1/\sqrt{100\mu H \times 0.47\mu F} = 146 \text{ krad/s}$$
$$Z_o = \sqrt{\frac{L_R}{C_R}} = \sqrt{\frac{100\mu H}{0.47\mu F}} = 14.6\Omega \qquad Q = \frac{Z_o}{R_c} = \sqrt{\frac{100\mu H}{0.47\mu F}}/1\Omega = 14.6$$
$$\alpha = \frac{R}{2L} = \frac{1\Omega}{2 \times 100\mu H} = 5,000$$
$$\omega = \sqrt{\omega_o^2 - \alpha^2} = \sqrt{(146\text{krad/s})^2 - 0.005^2} = 146.1 \text{ krad/s}$$

i. Three intervals are involved.
Interval *I* is given by equation (18.29)

$$t_I = L_R \Delta I / V_s$$
$$= 100\mu H \times 10A/340V = 2.94\mu s$$

The time for interval *II* is given by equation (18.50)

$$t_{II} = \left(2\pi - \sin^{-1}\left(I_o Z_o / V_s\right)\right)/\omega_o$$
$$= \left(2\pi - \sin^{-1}\left(10A \times 14.6\Omega / 340V\right)\right)/146\text{krad/s} = 40.0\mu s$$

The capacitor voltage at the end of interval *II* is

$$V_{C_R t4} = V_s(1 - \cos\omega_o t_{II})$$
$$= 340V \times (1 - \cos(146\text{krad/s} \times 40.0\mu s)) = 32.8V$$

Interval *III*
The capacitor voltage must discharge from 32.8V dc to zero volts, providing the 10A load current. That is

$$t_{IV} = V_{C_R t4} \times C_R / I_o$$
$$= 32.8V \times 0.47\mu F/10A = 1.5\mu s$$

The minimum commutation cycle time is therefore 2.94 + 40 + 1.5 = 44.44μs. Thus the maximum operating frequency is 22.5kHz.

ii. Circuit *Q* does not affect the first interval, which from part i. requires 2.94μs.
When a finite *Q* of 14.6 is used, equations (18.31) and (18.32) are employed for the second interval, the resonant part of the cycle. From equation (18.32)

$$i_c(\omega t) = \frac{V_s}{\omega L} \times e^{-\alpha t} \times \sin \omega t$$

$$-10A = \frac{340\text{V}}{146\text{krad/s} \times 100\mu\text{H}} \times e^{-\frac{1\Omega}{2\times 100\mu\text{H}} t} \times \sin(146\text{krad/s} \times t)$$

which yields t = 39.27μs. The capacitor voltage at this time is given by equation (18.31), that is

$$v_c(\omega \times t) = V_s\left(1 - e^{-\alpha t} \cos \omega t\right)$$

$$v_c(146\text{krad/s} \times 39.27\mu\text{s}) = 340\text{V} \times \left(1 - e^{-5{,}000 \times 39.27\mu\text{s}} \times \cos(146\text{krad/s} \times 39.27\mu\text{s})\right)$$

$$= 101.8\text{V}$$

The time for interval *III* is given by equation (18.40), that is

$$t_{III} = \frac{V_{C_R t4}\, C_R}{I_o} = \frac{101.8\text{V} \times 0.47\mu\text{F}}{10\text{A}} = 4.78\mu\text{s}$$

The minimum commutation cycle time is therefore 2.94 + 39.27 + 4.78 = 47.0μs. Thus the maximum operating frequency is 21.3kHz, which is required for maximum voltage output at 10A.

The main effect of a finite Q is to result in a higher voltage being retained on the capacitor to be discharged into the load at a constant rate, during interval *III*. Never-the-less this voltage is much less than that retained in the half-wave resonant switch case.

♣

Example 18.4: ***Zero-voltage, resonant-switch, step-down dc-to-dc converter - ½ wave***

The zero voltage resonant switch converter in figure 18.16 operates under the following conditions:

V_s = 192V I_o = 25A
L_R = 10μH C_R = 0.1μF

Determine

i. the minimum output current
ii. the switching frequency f_s for v_o = 48V
iii. switch average current and
iv. the peak switch/diode/capacitor voltage.

Solution

$$\omega_o = \frac{1}{\sqrt{L_R C_R}} = \frac{1}{\sqrt{10\mu\text{H} \times 0.1\mu\text{F}}} = 1 \times 10^6 \text{ rad/s} \quad \text{that is } f_o = 159.2\text{kHz}$$

$$Z_o = \sqrt{\frac{L_R}{C_R}} = \sqrt{\frac{10\mu\text{H}}{0.1\mu\text{F}}} = 10\Omega$$

i. For proper resonant action the maximum average output current must satisfy, $I_o > V_s / Z_o$, that is

$$\check{I}_o = \frac{V_s}{Z_o} = \frac{192\text{V}}{10\Omega} = 19.2\text{A}$$

Since the load current, 25A is larger than the minimum current requirement, 19.2A, the switch voltage will be reduced zero giving ZVS turn-off.

ii. The period of interval *I* is given by equation (18.67), that is

$$t_I = \frac{V_s\, C_R}{I_o} = \frac{192\text{V} \times 0.1\mu\text{F}}{25\text{A}} = 0.768\mu\text{s}$$

The period of interval *II* is given by equation (18.73), that is

$$t_{II} = t_3 - t_1 = \left(\pi + \sin^{-1}\frac{V_s}{I_o Z_o}\right) / \omega_o = \left(\pi + \sin^{-1}\frac{192\text{V}}{25\text{A} \times 10\Omega}\right) / 10^6 \text{ rad/s} = 4.017\mu\text{s}$$

The period for the constant current period *III* is given by equation (18.78)

$$t_{III} = \frac{I_o L_R}{V_s} \times (1 - \cos \omega_o t_{I1}) = \frac{25\text{A} \times 10\mu\text{H}}{192\text{V}} \times \left(1 - \cos\left(10^6 \times 4.017\mu\text{s}\right)\right) = 2.136\mu\text{s}$$

After re-arranging equation (18.80), the switching frequency is given by

$$f_s = \frac{\left(1 - \frac{v_o}{V_s}\right)}{t_s - \frac{1}{2}t_1} = \frac{\left(1 - \frac{48\text{V}}{192\text{V}}\right)}{(0.768\mu\text{s} + 4.017\mu\text{s} + 2.136\mu\text{s} - \frac{1}{2} \times 0.768\mu\text{s})} = 114.7\text{kHz}$$

iii. The switch current is shown by hatched dots in figure 18.16. The average value is dominated by interval *IV*, with a small contribution in interval *II* between t_5 and t_6.

$$\bar{I}_T = \frac{I_o}{\tau} \times \left(\frac{\frac{1}{2} \times t_{III}}{1 + \left|\cos \omega_o t_{II}\right|} + (\tau - t_6)\right)$$

$$= 25\text{A} \times 114.7\text{kHz}\left(\frac{\frac{1}{2} \times 2.136\mu\text{s}}{1 + \left|\cos\left(10^6 \times 4.017\mu\text{s}\right)\right|} + \left(\frac{1}{114.7\text{kHz}} - (0.768\mu\text{s} + 4.017\mu\text{s} + 2.136\mu\text{s})\right)\right)$$

$$= 7.0\text{A}$$

iv. The peak switch/diode/capacitor voltage is given by equation (18.71), namely

$$\hat{v}_C = V_s + I_o Z_o$$

$$= 192\text{V} + 25\text{A} \times 10\Omega = 442\text{V}$$

♣

18.6 Resonant-switch, dc-to-dc step-up voltage converters

18.6.1 ZCS resonant-switch, dc-to-dc step-up voltage converters

The zero current resonant ZCS (and ZVS) principle can be applied to the step-up converter (and buck-boost converter), as shown in figure 18.18b. The resonant *L-C* circuit around the switch does not affect the primary boosting function, but only facilitates resonant switching of switch *T*. But now the output voltage is determined by the switch off-time.
When the switch *T* is off, the input inductor *L* provides near constant current to the output circuit through diode *D*. The inductor current I_i comprises the load current I_o and the output capacitor current I_c. The resonant capacitor C_R is charged to the output voltage v_o, as is the output capacitor *C*.

Period 1: t_{P1}
When the switch is turned on at t_o, the input current I_i is progressively diverted to the resonant inductor L_R as its current builds up linearly according to

$$i_{LR}(t) = \frac{v_o}{L_R} t \tag{18.81}$$

The current to the output circuit, I_D, through diode *D*, decreases linearly according to

$$i_D(t) = I_i - i_{LR}(t) = I_i - \frac{V_o}{L_R} t \tag{18.82}$$

At time t_1 the resonant inductor consumes all the input current, when

$$t_{P1} = \frac{L_R I_I}{v_o} \tag{18.83}$$

Period 2: t_{P2}
The resonant capacitor can now resonate through L_R and the switch *T*. The inductor resonant current is superimposed on the constant input current, the constant current not producing any voltage across the inductor since *di/dt* is zero for a constant current.
The inductor, hence switch, current is

$$i_{LR}(t) = I_i + \frac{V_o}{Z} \sin \omega t \tag{18.84}$$

where $Z = \sqrt{\frac{L_R}{C_R}}$ and $\omega = \frac{1}{\sqrt{L_R C_R}}$

while the resonant capacitor current is

$$i_{CR}(t) = -\frac{V_o}{Z} \sin \omega t \tag{18.85}$$

The resonant capacitor and inductor are in parallel hence

$$v_{LR}(t) = v_{CR}(t) = L_R \frac{di_{LR}}{t} = v_o \sin \omega t \tag{18.86}$$

The maximum switch capacitor and inductor currents occur at t_2, namely $\omega t = \frac{1}{2}\pi$, when

$$\hat{I}_{CR} = -\frac{V_o}{Z}$$
$$\hat{I}_T = \hat{I}_{LR} = I_i + \frac{V_o}{Z} \tag{18.87}$$

The resonant capacitor current is zero when the inductor current falls back to the input current level I_i, that is, at time t_3 when $\omega t = \pi$.

The oscillation continues according to equation (18.84) and the resonant inductor current falls to zero at t_3, namely time

$$t_{LR=0} = \frac{1}{\omega}\sqrt{\pi + \sin^{-1}\frac{ZI_i}{v_o}} \tag{18.88}$$

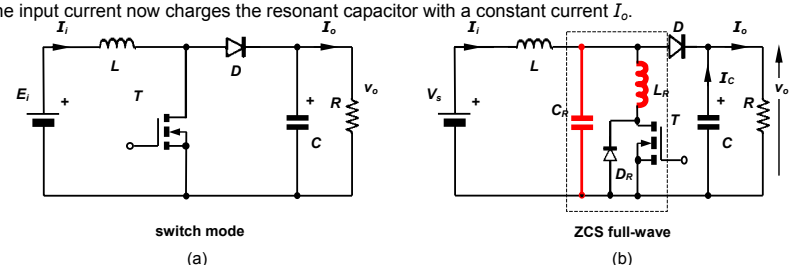
when the resonant circuit voltage from equation (18.86) is

$$v_{LR}(t_{LR=0}) = v_{CR}(t_{LR=0}) = -v_o\sqrt{1-\left(\frac{ZI_i}{v_o}\right)^2} \tag{18.89}$$

and the resonant capacitor current is

$$i_{CR}(t_{LR=0}) = I_i \tag{18.90}$$

The input current now charges the resonant capacitor with a constant current I_o.



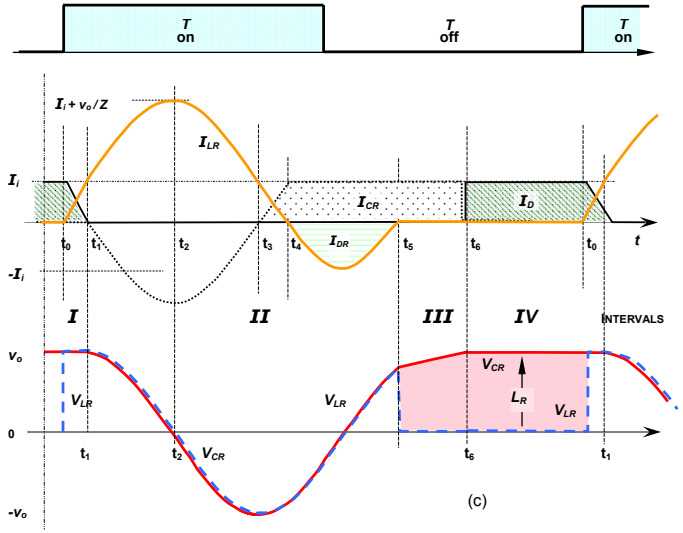


Figure 18.18. *Zero current switching, ZCS, full-wave resonant switch dc to dc step-up voltage converter: (a) conventional smps circuit; (b) ZCS resonant circuit; and (c) waveforms.*

During period 2, all the load current I_o is provided by the output capacitor *C*, and the output diode *D* is reverse biased.

Due to the resonant capacitor retaining a negative voltage at time t_4, the resonant oscillation current reverses for a negative half resonant cycle through the switch antiparallel diode D_R. During this period when the antiparallel diode D_R conducts, the switch can be turned off under a zero current condition.

The inductor current returns to zero at time t_5

$$t_{P2} = \frac{1}{\omega}\left[2\pi - \sin^{-1}\frac{ZI_i}{v_o}\right] \tag{18.91}$$

The capacitor voltage is given by equation (18.86) at the time t_{P2}, namely

$$v_{LR}(t_{P3}) = v_{CR}(t_{P3}) = v_o \sin\omega t_{P3} = v_o\sqrt{1-\left(\frac{ZI_i}{v_o}\right)^2} \tag{18.92}$$

Period 3: t_{P3}
The constant input current I_i charges the resonant capacitor C_R linearly to the output voltage level v_o. At this voltage the output capacitor C ceases to provide load current I_o since diode *D* conducts and the input current provides the load current I_o and replenishes to output capacitor *C* with the remainder of the input current, I_i - I_o. The charging time of the resonant capacitor C_R is load current magnitude I_o dependent.

$$v_{CR}(t) = v_{CR}(t_{P3}) + \frac{I_i}{C_R}t \tag{18.93}$$

18.6.2 ZVS resonant-switch, dc-to-dc step-up voltage converters

The alternative boost resonant ZVS converter in figure 18.19a uses a constant current input as in figure 14.35 in chapter 14.3.4, but the output is half wave rectified by the diode D_{rect}.

Initially, before t_o, the switch is on and the load requirement I_o is being provided by the output capacitor *C*. The large input inductance ensures a constant input current I_i, which is conducted by the switch *T*. The resonant circuit capacitor voltage is zero, as is the initial resonant inductor current.
The switch *T* is turned off at t_o and the resonant circuit waveforms as in figure 18.19 parts b and c occur.

Period 1: t_{P1}
The resonant capacitor charges with the constant input current which is diverted by the turn-off of switch *T*. The capacitor and parallel connected switch voltages increases according to

$$V_T(t) = V_{CR}(t) = \frac{I_i}{C_R}t \tag{18.94}$$

The capacitor and switch voltage rise linearly until equal to the output voltage v_o, when the output rectifying diode becomes forward biased at time t_1. The time for this first period is

$$t_{P1} = \frac{V_o}{I_i}C_R \tag{18.95}$$

Period 2: t_{P2}
The output rectifying diode D_{rect} is able to conducts and allows *L-C* resonance between L_R and C_R where the inductor is clamped to the output voltage v_o and both L_R and C_R are fed from the constant current source I_i. These two dc conditions do not prevent an ac resonant oscillation from occurring. The voltage across the capacitor increasing from v_o at time t_1 according to

$$v_{CR}(t) = v_o + I_i Z \sin\omega t \tag{18.96}$$

where $Z = \sqrt{\frac{L_R}{C_R}}$ and $\omega = \frac{1}{\sqrt{L_R C_R}}$

The inductor voltage hence current are

$$\begin{aligned} v_{LR}(t) &= v_o - v_{CR}(t) = -I_i Z \sin\omega t \\ i_{LR}(t) &= I_i(1-\cos\omega t) \end{aligned} \tag{18.97}$$

This inductor current replenishes to output capacitor *C* whilst providing a portion of the load current I_o.
The capacitor current resonantly decreases from I_i according to

$$i_{CR}(t) = I_i - i_{LR} = I_i\cos\omega t \tag{18.98}$$

This period continues until the capacitor voltage given by equation (18.96) reaches zero at time t_4. This zero voltage condition is necessary if the switch is to turn-on with zero voltage and from equation (18.96) a zero voltage condition occurs provided $I_i Z > v_o$. The capacitor voltage reaches zero and attempts to reverse at time t_4. The duration of the resonant period is

$$t_{P2} = \frac{\pi + \sin^{-1}\left(\frac{v_o}{I_i Z}\right)}{\omega} \tag{18.99}$$

The inductor current (whence capacitor current) at the instant t_4 is

$$i_{LR}(t_2) = I_i\left(1+\sqrt{1-\left(\frac{V_o}{I_i Z}\right)^2}\right)$$

$$i_{CR}(t_2) = i_{LR}(t_2) - I_i = I_i\sqrt{1-\left(\frac{V_o}{I_i Z}\right)^2} \qquad (18.100)$$

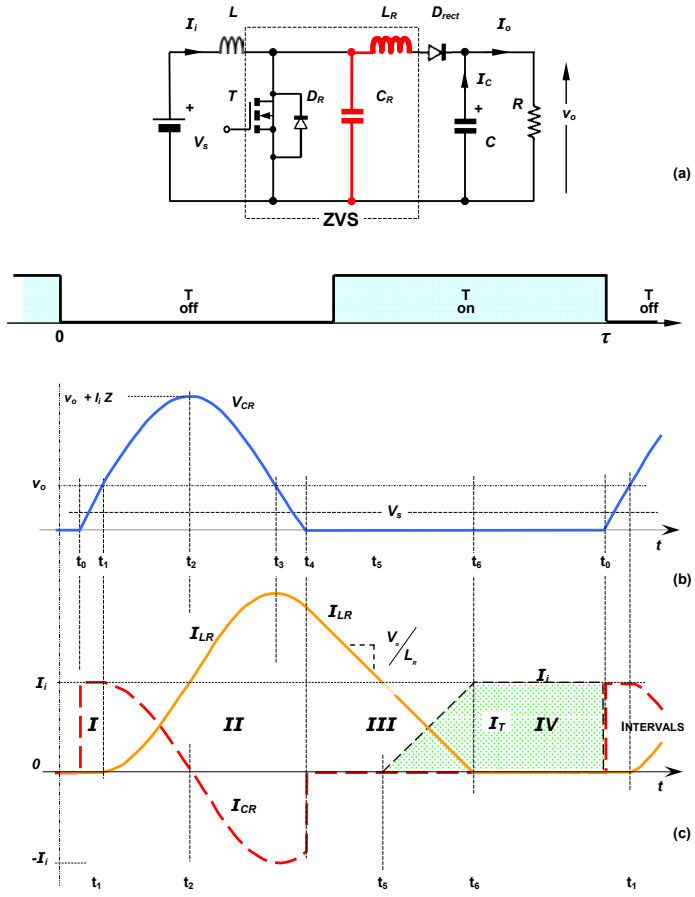


Figure 18.19. *Zero current switching, ZVS, full-wave resonant switch dc to dc step-up voltage converter: (a) ZCS resonant circuit; (b) resonant capacitor voltage; and (c) current waveforms.*

Period 3: t_{P3}

At time t_4 the diode D_R conducts, preventing the resonant capacitor from charging negatively. The resonant inductor releases its energy into the load circuit at a constant voltage v_o, according to

$$i_{LR}(t) = I_i\left(1+\sqrt{1-\left(\frac{V_o}{I_i Z}\right)^2}\right) - \frac{v_o}{L_R}t \qquad (18.101)$$

The diode D_R current decreases linearly to zero at time t_5 according to

$$i_{DR}(t) = I_i - i_{LR}(t) = -I_i\sqrt{1-\left(\frac{V_o}{I_i Z}\right)^2} + \frac{v_o}{L_R}t \qquad (18.102)$$

And reaches zero at time t_5 after a period

$$t_{P3a} = \frac{1}{\omega}\frac{I_i Z}{v_o}\sqrt{1-\left(\frac{V_o}{I_i Z}\right)^2} = L_R\frac{I_i}{v_o}\sqrt{1-\left(\frac{V_o}{I_i Z}\right)^2} \qquad (18.103)$$

At time t_5, the resonant inductor current is the input current I_i and the switch is turned on between t_4 and t_5 in order to achieved zero voltage turn-on ZVS. The inductor current continues to fall at the same rate to zero as the switch current linearly increases to I_i

$$i_T(t) = I_i - I_{LR}(t) = \frac{V_o}{L_R}t \qquad (18.104)$$

The inductor current reaches zero time t_6 that is input current I_i (hence load current I_o) dependent. The time for the inductor current to fall from the input current level I_i at time t_5 to zero at time t_6 is:

$$t_{P3b} = L_R\frac{I_i}{v_o} \qquad (18.105)$$

The time for the third period (t_3 to t_6) is

$$t_{P3} = T_{P3a} + t_{P3b} = L_R\frac{I_i}{v_o}\left(1+\sqrt{1-\left(\frac{V_o}{I_i Z}\right)^2}\right) \qquad (18.106)$$

At time t_6 the switch conducts the input current I_i. and can be turned off so as to control the output voltage v_o.

Summary and comparison of ZCS and ZVS Converters

The main characteristics of ZCS and ZVS are:

- Switch turn-on and turn-off occur at zero current or zero voltage, thus reducing switching losses.
- Rapid switch current and voltage changes are avoided in ZCS and ZVS, respectively. The *di/dt* and *dv/dt* values are small, hence EMI is reduced.
- In the ZCS, the peak current $I_o + V_{dc}/Z_o$ conducted by the switch is more than twice as high as the maximum of the load current I_o.
- In the ZVS, the switch must withstand a forward voltage $V_{dc} + Z_o I_o$, while $Z_o I_o$ must exceed V_{dc}.
- The switching frequency varies the output voltage.
- Switch parasitic capacitances are discharged during turn-on in ZCS, which can produce significant switching loss at high switching frequencies. This does not occur with ZVS.

Table 18.3: Characteristics of resonant tank circuits

characteristic	series	parallel	series/parallel
Resonant frequency ω_o	constant	Load dependent	Load dependent
Open circuit output	OK	Large current near resonance ω_o	Large current near resonance ω_o
Short circuit output	High current near resonance ω_o	Protected by L at all ω	High current near resonance ω_o
Output voltage frequency sensitivity	High at no load and light loads	Good light load regulation but low efficiency	OK but extra resonant component
Equivalent load, R_{eq}	$\frac{8}{\pi^2}R_{load}$	$\frac{\pi^2}{8}R_{load}$	$\frac{\pi^2}{8}R_{load}$

18.7 Appendix: Matrices of resonant switch buck, boost, and buck/boost converters

A series switch diode may not be necessary when an inverse parallel diode is used, as with full-wave ZCS and half-wave ZVS circuits. In the following circuits, the series diode (preferably in the drain circuit) is used to block the MOSFET internal parasitic diode, which may have poor recovery characteristics.

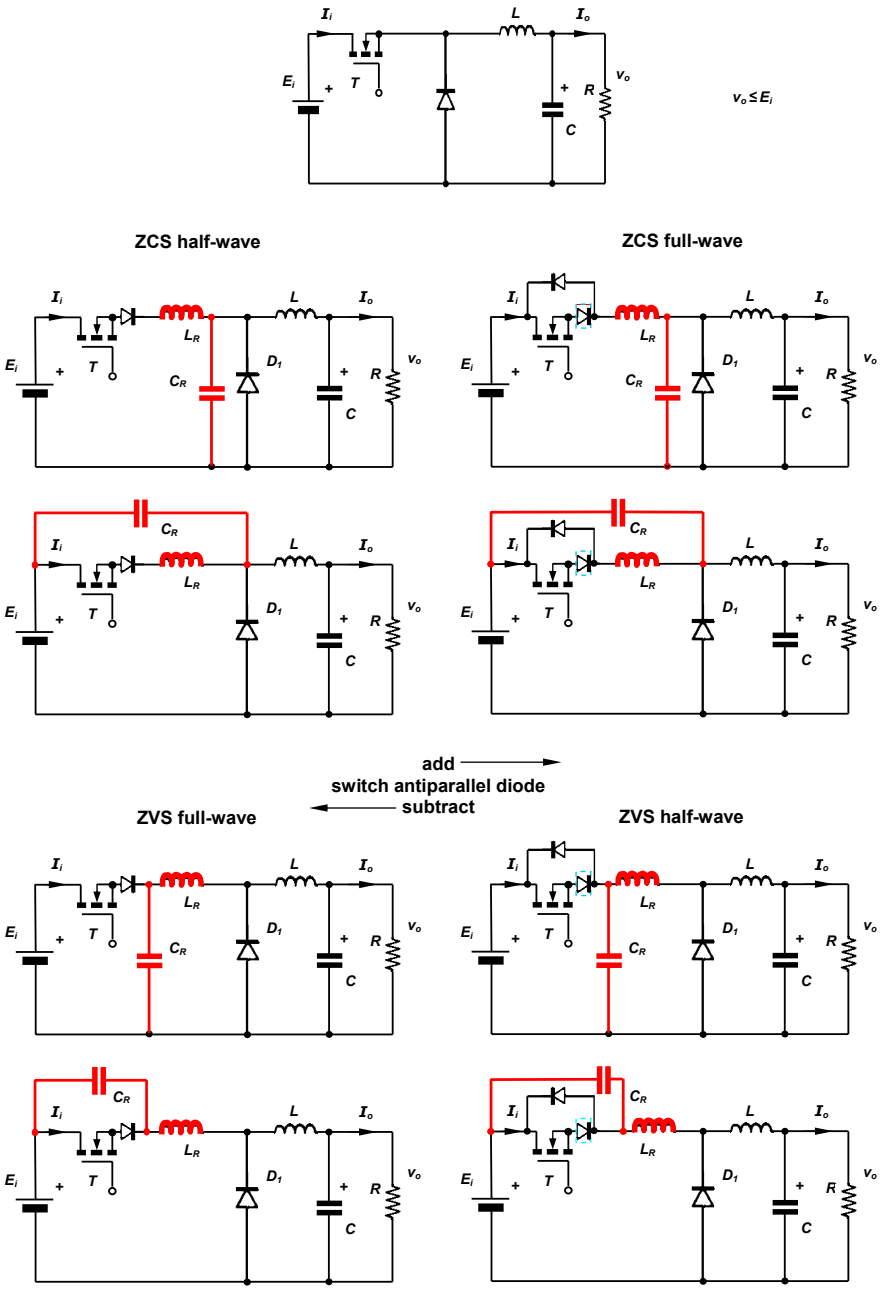


Figure 18.20. *Forward (buck) voltage converter resonant switch circuits.*

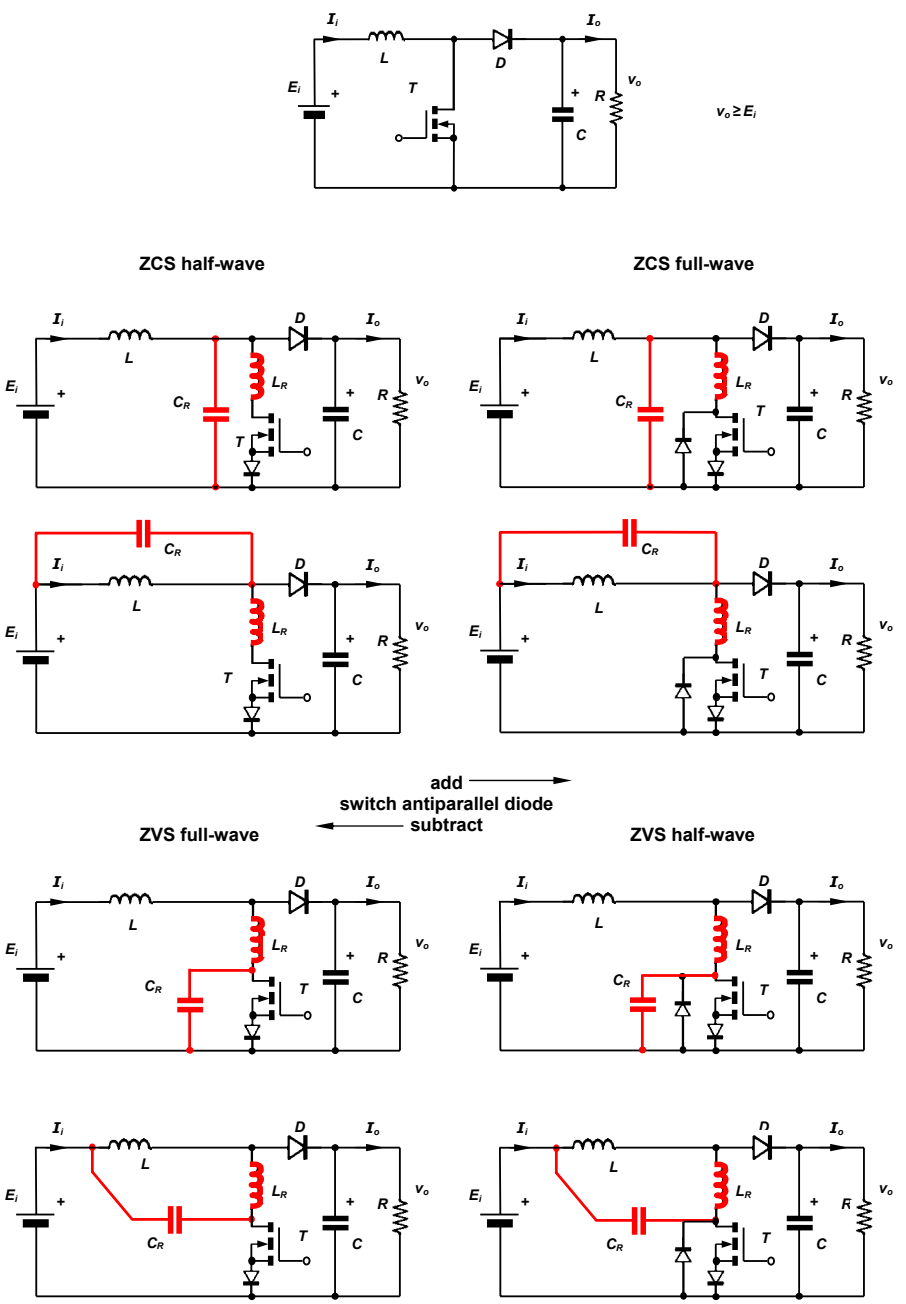


Figure 18.21. *Setup (boost) voltage converter resonant switch circuits.*

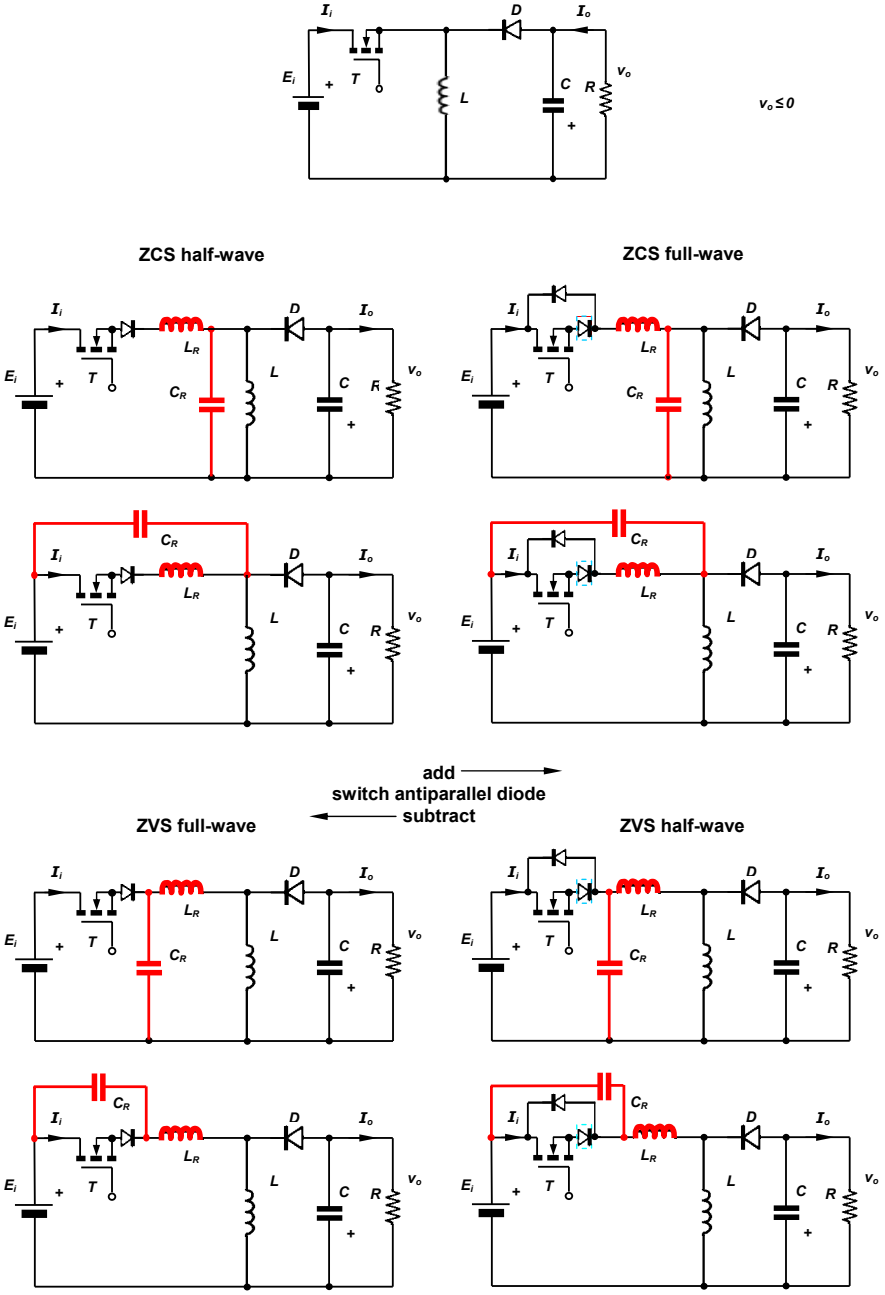


Figure 18.22. *Setup/down (buck/boost) voltage converter resonant switch circuits.*

Reading list

Hart, D.W., *Introduction to Power Electronics,* Prentice-Hall, Inc, 1997.

Mohan, N., et al., *Power Electronics,* 3rd Edition, Wiley International, 2003.

Thorborg, K., *Power Electronics – in theory and practice*, Chartwell-Bratt, 1993.

Problems

Series resonant dc to dc converter

18.1. The series resonant dc converter in figure 18.1a operates from a 340V dc supply at 100kHz with a 17Ω load. If the series *L*-*C* resonant components are 100μH and 47nF, determine the output voltage assuming high resonant circuit Q.

18.2 If the operating frequency in problem 18.1 is decreased to 50kHz, determine suitable *L*-*C* values if the output voltage to be halved.

Parallel resonant dc to dc converter

18.3. The series resonant dc converter in figure 18.5a operates from a 340V dc supply at 100kHz with a 17Ω load. If the parallel *L*-*C* resonant components are 10μH and 470nF, determine the output voltage assuming high resonant circuit Q.

18.4 If the operating frequency in problem 18.3 is decreased to 50kHz, determine suitable *L*-*C* values if the output voltage to be halved.

Zero-current resonant switch converter

18.5 The zero current resonant switch converter in figure 18.13a operates with a 20V dc input supply and resonant *L*-*C* values of 5μH and 10nF, and a 5A output load requirement. Determine
i. the output voltage if the switching frequency is 100kHz
ii. the switching frequency if the output voltage is 10V
In each case determine the maximum capacitor voltage and maximum inductor current.

Zero-voltage resonant switch converter

18.6 The zero current resonant switch converter in figure 18.16a operates with a 20V dc input supply and resonant *L*-*C* values of 10μH and 100nF, and a 5A output load requirement. Determine
i. the output voltage if the switching frequency is 100kHz
ii. the switching frequency if the output voltage is 10V
In each case determine the maximum capacitor voltage and maximum inductor current.

CHAPTER 19

HV Direct-Current Transmission

Originally electrical power generation, transmission, and distribution systems were direct current. The advent of the three-phase induction motor and the ability of transformers to converter one ac voltage to another ac voltage level (at the same frequency), saw the unassailable rise to dominance of ac electrical power systems. But for long distance electrical power transmission, of just a few hundred kilometres of typically about 200 to 350km, a dc transmission system is a viable possibility. For underwater or underground electrical power transmission, ac may not be viable at just 50km due to high capacitive charging currents because of the close proximity of the cables (particularly subsea cables). The highest functional dc voltage for dc transmission, HVDC, is ±600kV over 785km and 805km transmission lines in Brazil. Each of the two bipolar dc transmission systems carry 3.15GW. Also involved are three, three-phase 765kV ac lines which are 1GVAr variable capacitor series compensated (FACTS) at two intermediate substations.

19.1 HVDC Electrical Power Transmission

Electrical power is generated in ac form and is also usually distributed and consumed in an ac form. Its long distance transmission between these two stages may be an ac or a dc transmission system. In a high-voltage dc (HVDC) system the generated 50/60Hz ac is controlled-rectified to dc, transmitted, then at the receiving end, converted from dc back to 50/60Hz ac. A HVDC system is two ac systems connected by a dc transmission system, where the ac systems can be totally independent.

The dc-link of a HVDC transmission system is either

- A controlled voltage dc-link or
- A controlled current dc-link.

A HVDC controlled *current link* has the following characteristics.

The link in highly inductive, achieved with series inductance at each end.

The converter/inverter technology is operated in a controlled current mode, thus the converter/inverter devices require reverse voltage blocking ability. Symmetrical blocking thyristor devices are applicable, but such devices are restricted to line commutation and phased control.

A HVDC controlled *voltage link* has the following characteristics.

The link in highly capacitive, achieved with parallel connected capacitance at each end.

The converter/inverter technology is operated in a controlled voltage mode, thus the converter/inverter devices can be uni-directional voltage blocking IGBT technology. Since devices are gate commutatable, a switching frequency of kHz's is possible, thus PWM techniques can be employed for harmonic minimisation.

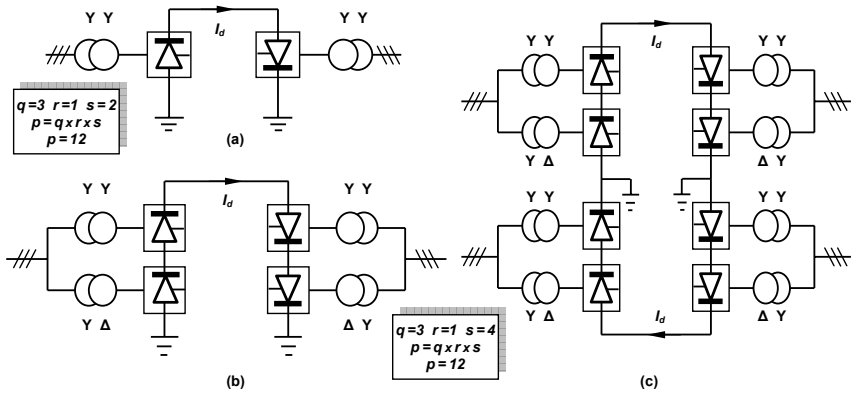


Figure 19.1. *HVDC transmission systems: (a) 6 pulse monopole; (b) 12 pulse monopole; and (c) 12 pulse bipolar, converter bridge configurations.*

19.2 HVDC Configurations

There are a number of different configurations for transmitting dc power, depending on the number of cables employed. Each uses a three-phase fully-controlled thyristor converter (rectifier) coupled through a dc link to another identical three-phase fully-controlled thyristor converter (inverter). Both converters have the same modular structure except the converter connections to the dc link are interchanged for one converter, hence power flow is fully reversible. Since the valves can only conduct current in one direction, power reversal is achieved by changing the polarity of the dc link terminal voltages through control of the converter thyristor firing delay angles. The rectification mode (positive dc link voltage) is achieved with thyristor firing angles of $0 < \alpha < \frac{1}{2}\pi$ while inversion (negative dc link voltage) is achieved with firing angles of $\frac{1}{2}\pi < \alpha < \pi$. Because one converter terminal connection is reversed, a rectifying voltage ($0 < \alpha < \frac{1}{2}\pi$) is opposed by an inverting voltage ($\frac{1}{2}\pi < \alpha < \pi$) – subtractive not additive voltages.

19.2i - Monopole and earth return

The monopole configurations shown in figure 19.1 parts a and b (6 pulse and 12 pulse respectively) use just one transmission cable and earth is used as the negative return. Occasionally, a metal earth return may be used, but importantly any return is at ground potential thus does not need the full transmission voltage insulation. The converter output terminals are reversed relative to one another, as indicated by the direction of the thyristors in the symbol blocks.

Issues involved in using a ground return are

- Electrochemical corrosion of buried metals objects, like pipelines
- Electrode chemical reaction under the sea
- Magnetic field disturbances when the go and return paths become unbalanced

The monopole system is limited in power handling capability, typically 1.5GW above the ground, and 600MW below the ground or under the sea.

19.2ii - Bipolar

In the bipolar arrangement two high voltage conductors, at opposite potentials with respect to ground, are used as shown in figure 19.1c. Any pole imbalance uses an earth return, if a low-voltage metal ground return is not used. The bipolar configuration has a number of advantages over the monopole arrangement.

- Normally no earth-current flows which minimises earth losses and any earth related environmental effects, including minimal corrosion of underground system metal components
- If a fault develops on one pole, the other pole can continue to operate in a monopole arrangement, using the earth as the return path
- For a given power rating, each conductor has half the cross-sectional area of the monopole line, thus reducing the extra cost of using a second conductor
- The same dc transmission line towers can carry two lines with a small additional capital cost

A homopolar hvdc link is formed if the two high voltage conductors have the same polarity, with, undesirably, a high ground or metal return current.

The bipolar system is capable of higher transmission powers than the monopole configuration. Bipolar systems can carry over 3GW at voltages of over ±500kV, over distances well in excess of 500km.

19.2iii - Tripole
Two of the three conductors of an ac system are used in a bipolar configuration, with the third conductor used as a parallel monopole with bidirectional power flow capability. The bidirectional capability of the third conductor allows each of the two bipolar conductors to carry higher than rated current when each in turn is relieved periodically by the monopole system, such that all three conductors do not exceed rated I^2R losses. In this way each of the three conductors experience the same thermal losses. This is achieved if the bipolar currents are cycled every few minutes between 0.366pu $½\left(\sqrt{3}-1\right)$ and 1.366 pu $½\left(\sqrt{3}+1\right)$, with ±1 pu being appropriately alternated in the monopole. As a result, 80% more power can be transmitted compared with the ac equivalent, using the same conductors, towers, etc. Unlike the ac equivalent, the dc system can be fully loaded without system instability or need for reactive power compensation.

19.2iv - Back-to-back
Two different asynchronous ac systems in close proximity, possibly operating at different frequencies, can be interconnected by either a monopole or bipolar system. Since no dc transmission cables are necessary, because of the close proximity of the two systems, the system type and voltage level are not restrictive.

19.2v - Multi-terminal
More than two converters are connected to the same dc link, where simultaneously, at least one converter operates in the rectification mode and at least one other converter operates in an inversion mode. Mechanical switching of the converter terminals is necessary. The control system is more elaborate than for normal point-to-point transmission and power reversal is affected by current reversal.

19.3 Typical HVDC transmission system

A fully modular hardware structure is used.
Each 8.8kV symmetrically blocking thyristor is configured in a module with its gate electronics and RC snubber as shown in figure 19.2. Many thyristors are connected in series to form a valve, with internal static voltage sharing resistors and a saturable reactor turn-on snubber as shown in figure 19.2a.
Six valves are needed to form the 6 pulse valve group converter bridge in figure 19.3a, while 12 valves are used to form the twelve pulse value group converter in figure 19.3b. Each group of four valves in a single vertical stack form quadrivalves as shown in figure 19.3b. Each quadrivalve may contain hundreds of series connected thyristors to give the necessary hundreds of kV pole voltage rating.

Typical six and twelve pulse valve group configurations are shown in figure 19.3, which form the unipolar converters in figure 19.1 parts a and b respectively. A more detailed circuit diagram of the 12-pulse converter and its MOV voltage protection and dc and ac harmonic filtering circuitry, in a substation installation, is shown in figure 19.4. The dc inductors L_{dc} in each pole assist in filtering harmonic currents and smooth the dc side current thereby reducing the current level for the onset of discontinuous current flow. Because the inductors control the dc link side *di/dt*, converter commutation is more robust. No dc filters may be necessary with the back-to-back HVDC converter configuration.

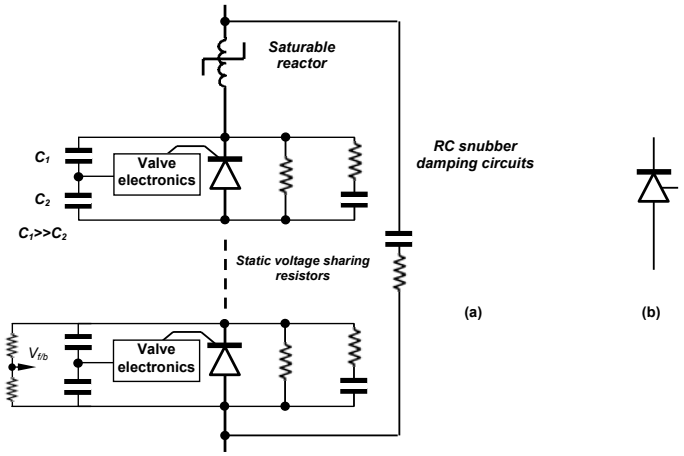


Figure 19.2. *Thyristor valve: (a) modules components assembled into a valve and (b) valve symbol.*

19.4 Twelve-pulse ac line frequency converters

The six-pulse line-frequency fully-controlled thyristor converter was discussed in chapter 12.6. Harmonic filters are required on both the ac and dc side of the converter as shown in figure 19.4. The ac side harmonics occur at 6*n*±1 the fundamental, while the dc side harmonics are generated at 6*n*. In order to reduce the filtering requirements, and increase the effectiveness of the filtering, on both the ac and dc sides, most high power HVDC systems use 12-pulse, phase-shited transformer/converter arrangements. The ac side harmonics now occur at 12*n*±1 and the dc components are generated at 12*n*.
Twelve-pulse converter operation is achieved by using the series bridge connection in conjunction with Δ-Y and Y-Y compound connected transformers as shown in figure 19.3b. (A delta connection is usually employed on the lower voltage side of the transformer.) Figure 19.7a shows the arrangement in more detail, with the necessary transformer turns ratio to ensure each converter bridge produces the same output voltage at the same thyristor firing delay angle. Voltage matching between the ac line and required dc link voltage is achieved with the transformer turns ratio *N*, shown in figure 19.7a. The series thyristors in each bridge provide paths which allow both converter currents to be equal.

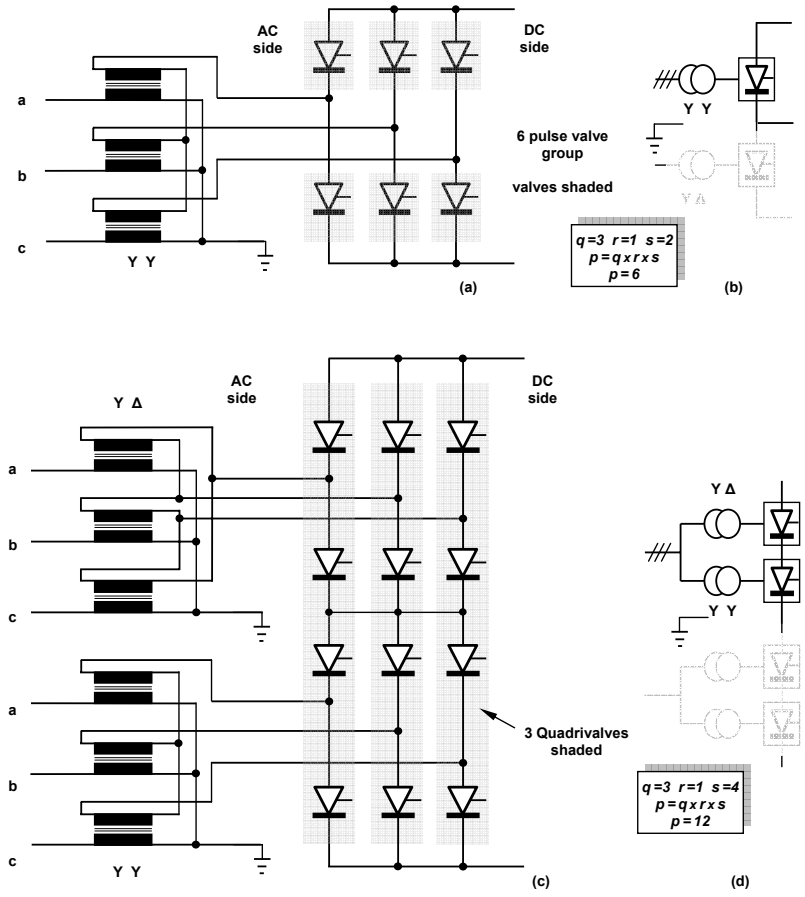


Figure 19.3. *Monopole converter bridges:*
Six-pulse valve group (a) converter bridge schematic and (b) six-pulse valve group converter symbol; Twelve-pulse valve group converter configuration with star-star and star-delta connected converter transformers: (c) converter schematic and (d) twelve-pulse valve group converter symbol.

As a result of the transformer configuration, the corresponding upper and lower transformer voltages are displaced by 30°, where $Va_{Ys\text{-}nY}$ leads $Va_{\Delta s\text{-}n\Delta}$ by $\frac{1}{6}\pi$ radians. The dc-link current I_d is assumed constant because of the large smoothing inductor L_{dc} (linear and typically ½H). If source impedance is neglected, then the various circuit current waveforms are constituted from rectangular current blocks as shown in figure 19.7b. Each converter operates with the same firing delay angle α, with respect to the voltage references shown in figure 19.7b. Because the transformer primaries are in parallel, the input current is the sum of the appropriate two transformer phase currents, namely $i_a = i_{Ya} + i_{\Delta a}$ for phase a. The Fourier series for each transformer primary phase current is obtained from analysis of the appropriate six-pulse converter current, for each converter

$$i_{Ya} = \frac{2\sqrt{3}\,I_d}{N\pi}\left(\cos\theta - \frac{1}{5}\cos 5\theta + \frac{1}{7}\cos 7\theta - \frac{1}{11}\cos 11\theta + \frac{1}{13}\cos 13\theta .....\right) \tag{19.1}$$

and

$$i_{\Delta a} = \frac{2\sqrt{3}\,I_d}{N\pi}\left(\cos\theta + \frac{1}{5}\cos 5\theta - \frac{1}{7}\cos 7\theta - \frac{1}{11}\cos 11\theta + \frac{1}{13}\cos 13\theta .....\right) \tag{19.2}$$

Because of the symmetry of a three-phase system, no triplens exist in each input current. The total line current drawn is

$$i_a = i_{Ya} + i_{\Delta a} = \frac{4\sqrt{3}\,I_d}{N\pi}\left(\cos\theta - \frac{1}{11}\cos 11\theta + \frac{1}{13}\cos 13\theta - \frac{1}{23}\cos 23\theta .....\right) \tag{19.3}$$

The 12-pulse transformer/converter arrangement cancels harmonic components $6\times(2n\text{-}1)\pm1$.

The line current i_a rms value is $\left(1+\frac{1}{\sqrt{3}}\right)I_d/N$ and the rms fundamental is $2\sqrt{6}\,I_d/N\pi$.

The ac line current harmonics occur at $12n\pm1$. The valve side ac line current, shown as $N\times I_{Ysa}$ (or $N\times I_{\Delta sa}$) in figure 19.7b has an rms value of $I_d\sqrt{2/3}$, and once rectified, the valve unipolar current has an rms value of $I_d/\sqrt{3}$.

The more general equation are for the general case of asymmetrical converter firing when $\alpha_1 \neq \alpha_2$. Assuming a turns ratio between the phase windings of the star connected primary and the star connected secondary is 2:1, and the turns ratio between the primary and the delta connected secondary is 2:√3, then

$$i_a = \tfrac{1}{2} i_{Ya} + \tfrac{\sqrt{3}}{2} i_{\Delta a} \tag{19.4}$$

Due to waveform symmetry, no dc component or even harmonics exist.

The Fourier coefficients a_{nY} and b_{nY} of i_{Ya} are

$$a_{nY} = \frac{1}{\pi}\int_0^{2\pi} i_a(t)\cos n\omega t\; d\omega t = \frac{1}{\pi}\left[\int_{\frac{\pi}{6}+\alpha_1}^{\frac{5\pi}{6}+\alpha_1} I_d\cos n\omega t\; d\omega t - \int_{\frac{7\pi}{6}+\alpha_1}^{\frac{11\pi}{6}+\alpha_1} I_d\cos n\omega t\; d\omega t\right] \tag{19.5}$$

$$= -\frac{4I_d}{n\pi}\sin\frac{n\pi}{3}\sin n\alpha_1 \qquad n = 1, 3, 5, \ldots$$

$$b_{nY} = \frac{1}{\pi}\int_0^{2\pi} i_a(t)\sin n\omega t\; d\omega t = \frac{1}{\pi}\left[\int_{\frac{\pi}{6}+\alpha_1}^{\frac{5\pi}{6}+\alpha_1} I_d\sin n\omega t\; d\omega t - \int_{\frac{7\pi}{6}+\alpha_1}^{\frac{11\pi}{6}+\alpha_1} I_d\sin n\omega t\; d\omega t\right] \tag{19.6}$$

$$= \frac{4I_d}{n\pi}\sin\frac{n\pi}{3}\cos n\alpha_1 \qquad n = 1, 3, 5, \ldots$$

The Fourier coefficients $a_{n\Delta}$ and $b_{n\Delta}$ of the delta winding current $i_{\Delta a}$ are

$$a_{n\Delta} = \frac{1}{\pi}\int_0^{2\pi} i_a(t)\cos n\omega t\; d\omega t$$

$$= \frac{2}{\pi}\left[\int_{0+\alpha_2}^{\frac{\pi}{3}+\alpha_2}\frac{I_d}{3}\cos n\omega t\; d\omega t + \int_{\frac{\pi}{3}+\alpha_2}^{\frac{2\pi}{3}+\alpha_2}\frac{2I_d}{3}\cos n\omega t\; d\omega t + \int_{\frac{2\pi}{3}+\alpha_2}^{\pi+\alpha_2}\frac{I_d}{3}\cos n\omega t\; d\omega t\right] \tag{19.7}$$

$$= \frac{2I_d}{3n\pi}\left[-\sin n\alpha_2 - \sin n\left(\frac{\pi}{3}+\alpha_2\right) + \sin n\left(\frac{2\pi}{3}+\alpha_2\right) + \sin n(\pi+\alpha_2)\right]$$

$$b_{n\Delta} = \frac{1}{\pi}\int_0^{2\pi} i_a(t)\sin n\omega t\; d\omega t$$

$$= \frac{2}{\pi}\left[\int_{0+\alpha_2}^{\frac{\pi}{3}+\alpha_2}\frac{I_d}{3}\sin n\omega t\; d\omega t + \int_{\frac{\pi}{3}+\alpha_2}^{\frac{2\pi}{3}+\alpha_2}\frac{2I_d}{3}\sin n\omega t\; d\omega t + \int_{\frac{2\pi}{3}+\alpha_2}^{\pi+\alpha_2}\frac{I_d}{3}\sin n\omega t\; d\omega t\right] \tag{19.8}$$

$$= \frac{2I_d}{3n\pi}\left[\cos n\alpha_2 + \cos n\left(\frac{\pi}{3}+\alpha_2\right) - \cos n\left(\frac{2\pi}{3}+\alpha_2\right) - \cos n(\pi+\alpha_2)\right]$$

$$n = 1, 3, 5, \ldots$$

The Fourier coefficients for the line current are then defined by

$$a_n = \tfrac{1}{2}a_{nY} + \tfrac{1}{2}a_{n\Delta}$$
$$b_n = \tfrac{1}{2}b_{nY} + \tfrac{1}{2}b_{n\Delta} \tag{19.9}$$

From which the input line current is defined by

$$i_a(t) = \sum_{n=1,2,3,\ldots}^{\infty}\sqrt{2}I_{an}\sin(n\omega t + \varphi_n) \tag{19.10}$$

Output voltage

The converter outputs are series connected, hence the output voltage is additive for each pole, namely $V_{dr} = V_{dr1} + V_{dr2}$. The converter output voltage, with the constraint that both converters have a trigger delay of α, is

$$V_{dr} = V_{dr1} + V_{dr2} = \frac{6\sqrt{2}}{\pi N}V_{LL}\cos\alpha = 2.70\frac{V_{LL}}{N}\cos\alpha \tag{19.11}$$

The peak output voltage occurs midway between the peak voltage from each converter, and for $\alpha = 0$

$$\hat{V}_{dr} = 2\sqrt{2}\,\frac{V_{LL}}{\pi N}\cos 15° = 1.932\sqrt{2}\,\frac{V_{LL}}{\pi N} \tag{19.12}$$

Each converter delivers six current blocks of magnitude I_d, comprised of two ⅔π current blocks π radians apart in each converter arm. Since each converter output is the same but shifted by $\frac{1}{6}\pi$ radians, provided the two converters have equal delay angles, the resultant 12 current block per cycle results in the dc side voltage harmonics in V_{dr} being of the order $12n$.

If the two delay angles are controlled individually, the two outputs add but with harmonic components given by two six pulse converter, displaced by $\frac{1}{6}\pi$, where

$$V_{dr1} = \frac{3V_{max}}{\pi}\left(\cos\alpha_1 + \sum_{n=1}^{\infty}\sqrt{\frac{1}{(6n-1)^2} + \frac{1}{(6n+1)^2} - \frac{2\cos 2\alpha_1}{(6n-1)(6n+1)}}\sin\left(6n\omega t + \lambda_{6n1}\right)\right)$$
$$V_{dr1} = \frac{3V_{max}}{\pi}\left(\cos\alpha_2 + \sum_{n=1}^{\infty}\sqrt{\frac{1}{(6n-1)^2} + \frac{1}{(6n+1)^2} - \frac{2\cos 2\alpha_2}{(6n-1)(6n+1)}}\sin\left(6\left(n\omega t - \tfrac{1}{6}\pi\right) + \lambda_{6n2}\right)\right) \tag{19.13}$$

where

$$\lambda_{6n1} = -n\pi + \tan^{-1}\left[\frac{\frac{\cos(6n+1)\alpha_1}{(6n+1)} - \frac{\cos(6n-1)\alpha_1}{(6n-1)}}{\frac{\sin(6n+1)\alpha_1}{(6n+1)} - \frac{\sin(6n-1)\alpha_1}{(6n-1)}}\right]$$

$$\lambda_{6n2} = -n\pi + \tan^{-1}\left[\frac{\frac{\cos(6n+1)\alpha_2}{(6n+1)} - \frac{\cos(6n-1)\alpha_2}{(6n-1)}}{\frac{\sin(6n+1)\alpha_2}{(6n+1)} - \frac{\sin(6n-1)\alpha_2}{(6n-1)}}\right]$$

When the two delays angles differ, $\alpha_1 \neq \alpha_2$, the output voltage harmonics occur at order $6n$.

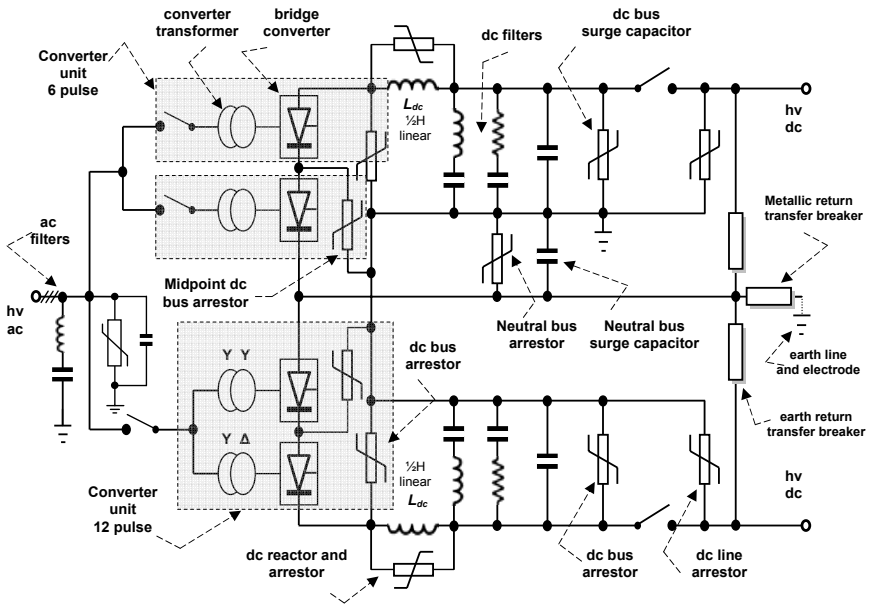


Figure 19.4. *Thyristor HVDC substation.*

19.4.1 Rectifier mode

Figure 19.7b shows that the angle between the input ac voltage and its fundamental current I_{a1} is determine by and equals, the phase delay angle α_r. The phasor diagram for rectification is shown in figure 19.8a. For a constant link current I_d, the fundamental ac input power factor is cos α_r while the input reactive power is given by

$$\begin{aligned} Q_r &= \sqrt{3}\, V_{LL} I_{a1} \sin\alpha_r = P_r \tan\alpha_r \\ &= \sqrt{3}\, V_{LL} \times \frac{2\sqrt{6}}{N\pi} I_d \times \sin\alpha_r = 2.7 \times V_{LL} \times \frac{I_d}{N} \times \sin\alpha_r \end{aligned} \tag{19.14}$$

The rms of the fundamental line current I_{a1} is $2\sqrt{6} I_d / N\pi$.

The real power transfer, which is the rectifier output power, is given by

$$\begin{aligned} P_r &= V_{dr} I_d = \sqrt{3}\, V_{LL} I_{a1} \cos\alpha_r = Q_r / \tan\alpha_r \\ &= \sqrt{3}\, V_{LL} \times \frac{2\sqrt{6}}{N\pi} I_d \times \cos\alpha_r = 2.7 \times V_{LL} \times \frac{I_d}{N} \times \cos\alpha_r \end{aligned} \tag{19.15}$$

To maximize the power flow and minimize the reactive power, the delay angle α_r should be small. From equation (19.15), to minimise the link I^2R loss, both I_d and the delay angle α_r should be small. That is, from equation (19.11), the rectifier output voltage should be maximised.

A low converter firing angle minimise the reactive power, reduce snubber losses, and reduces the harmonic content.

19.4.2 Inverter mode

The same basic rectifier mode equations hold in the inversion mode except that $\alpha_i > ½\pi$. Operational waveforms and the phasor diagram for this mode are shown in figure 19.8b. The reactive power is

$$\begin{aligned} Q_i &= \sqrt{3}\, V_{LL} I_{a1} \sin\alpha_i = \sqrt{3}\, V_{LL} \times \frac{2\sqrt{6}}{N\pi} I_d \times \sin\alpha_i = 2.7 \times V_{LL} \times \frac{I_d}{N} \times \sin\alpha_i \\ &= P_i \tan\alpha_i \end{aligned} \tag{19.16}$$

and the real power transfer, which is inverted into the ac system is given by

$$\begin{aligned} P_i &= V_{di} I_d = \sqrt{3}\, V_{LL} I_{a1} \cos\alpha_i = \sqrt{3}\, V_{LL} \times \frac{2\sqrt{6}}{N\pi} I_d \times \cos\alpha_i = 2.7 \times V_{LL} \times \frac{I_d}{N} \times \cos\alpha_i \\ &= Q_i / \tan\alpha_i \end{aligned} \tag{19.17}$$

To maximize the inverted power flow and minimize the reactive power, the delay angle α_i should be large (→ π). From equation (19.17), to minimise the link I^2R loss, I_d should be minimised (maximise dc link voltage) and the delay angle α_i should be large (→ π). Thus the inversion voltage should be as large as possible, avoiding commutation failure. Then the maximum α_i decreases as current increases since thyristor commutation time increases with current (and temperature).

P-Q operation of the twelve pulse series connected converter

The theory of controlled active and reactive power operation of a series connected 12-pulse converter assumes an ideal input transformer, negligible source impedance, and a constant current from the converter dc output bus. The output voltage contribution due to each constituent 6-pulse converter is given by

$$V_{dri} = \frac{3 V_{max}}{\pi} \cos\alpha_i \tag{19.18}$$

where V_{max} is the peak line voltage of the voltage applied to each converter, α_i is the firing angle of the constituent converter, and *i* represents either converter 1 or 2.

By multiplying (19.18) by the dc load current I_{dc}, the active power is given by

$$Q_i = P_{max} \sin\alpha_i \tag{19.19}$$

where

$$P_{max} = \frac{3}{\pi} \bar{I}_o V_{max}$$

which represents the maximum power delivered from each converter at $\alpha_i = 0$. Similarly, the reactive power absorbed by the converter is given by

$$Q_i = P_{max} \sin\alpha_i \tag{19.20}$$

The active power P_d and reactive power Q_d, drawn by the 12-pulse converter are the sum of the contribution made by each converter, namely

$$\begin{aligned} P_d &= P_{max}(\cos\alpha_1 + \cos\alpha_2) \\ Q_d &= P_{max}(\sin\alpha_1 + \sin\alpha_2) \end{aligned} \tag{19.21}$$

From these two equations, the 12-pulse converter draws active power of $2P_{max}$ and zero reactive power at $\alpha_1 = \alpha_2 = 0$. Similarly, at $\alpha_1 = \alpha_2 = ½\pi$, the converter draws reactive power of $2P_{max}$ and zero active power. Therefore, $2P_{max}$ is the base power, 1 pu, and the equations in (19.21) become

$$\begin{aligned} P_{pu} &= ½(\cos\alpha_1 + \cos\alpha_2) \\ Q_{pu} &= ½(\sin\alpha_1 + \sin\alpha_2) \end{aligned} \tag{19.22}$$

The variation of P_{pu} and Q_{pu} with simultaneous variation of α_1 and α_2 is shown in the three-dimensional plots figure 19.5.

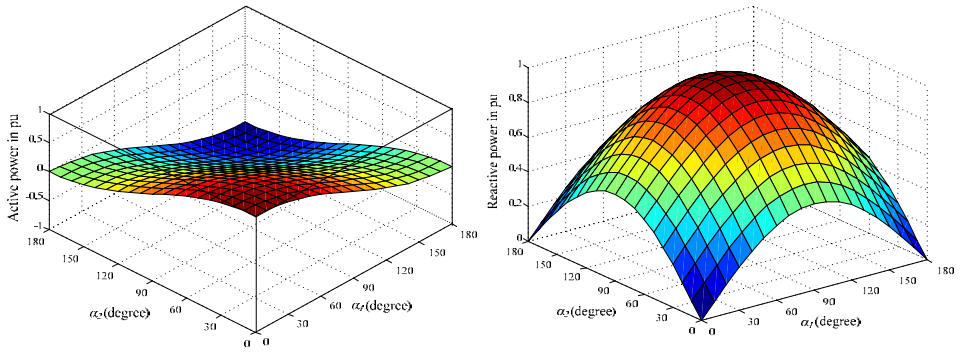
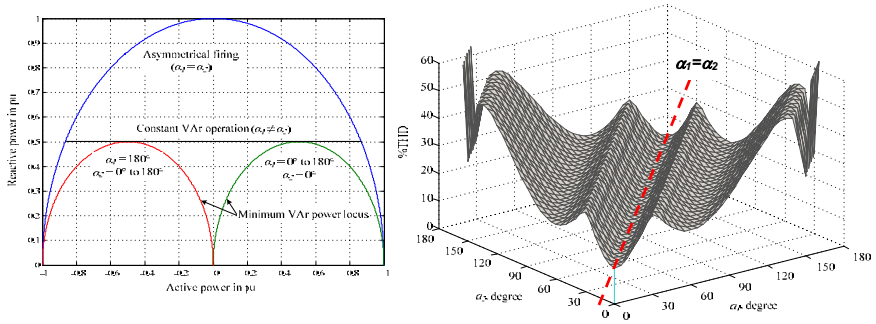


Figure 19.5. *Effects of varying of α_1 and α_2 on: (a) active power and (b) reactive power.*

The 12-pulse converter may operate with symmetrical firing of each 6-pulse converter, where $\alpha_1 = \alpha_2$ or with asymmetrical firing where $\alpha_1 \neq \alpha_2$ (or a combination of both). Different power loci can be realized by different combinations of firing angles (α_1 *and* α_2). Three power loci are shown in part 'a' of figure 19.6. The symmetrical firing *P*-*Q* locus is represented by the outer semicircle with centre '0' and radius of 1 pu, and is obtained by the symmetrical firing of α_1 and α_2 from 0 to π using the parts of equation (19.22). All points in the *P*-*Q* plane within the outer semicircle can be achieved by unique combinations of α_1 and α_2. The figure also shows power loci for asymmetrical firing which are represented by the circumference of the two inner semicircles with radii of 0.5 pu. These are obtained by varying α_1 from 0 to π while α_2 is held at '0' in the case of rectification and by varying α_2 from 0 to π while α_1 is held at π in the case of inversion. Part 'a' also shows the power locus for constant VAr operation indicated by the line parallel to the P_{pu} axis. With asymmetrical firing the maximum reactive power is decreased to one-half the case for symmetrical firing. Part 'b' of figure 19.6 shows the variation of input current THD with varied delay angles and reflects the adverse effect of asymmetrical firing on the THD.



(a) (b)

Figure 19.6: *Twelve-pulse fully-controlled converter: (a) power loci and (b) input current THD.*

The symmetrical firing power locus is associated with the best THD and the worst supply current power factor. On the other hand, asymmetrical firing decreases the reactive power flowing in the system, which improves the input power factor but deteriorates the input current THD, which limits the power factor improvement. As a result, asymmetrical firing of the 12-pulse converter offers the possibility of a smaller reactive power compensator for power factor compensation, at the expense increased input current harmonics.

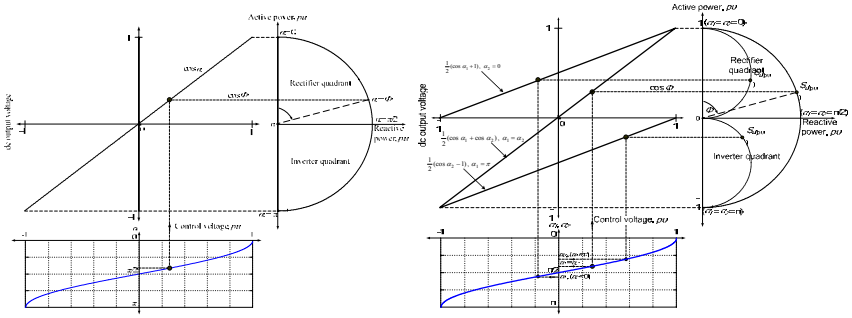


Figure 19.7X: *Power locus of 6-pulse and 12-pulse converters and per unit output voltage.*

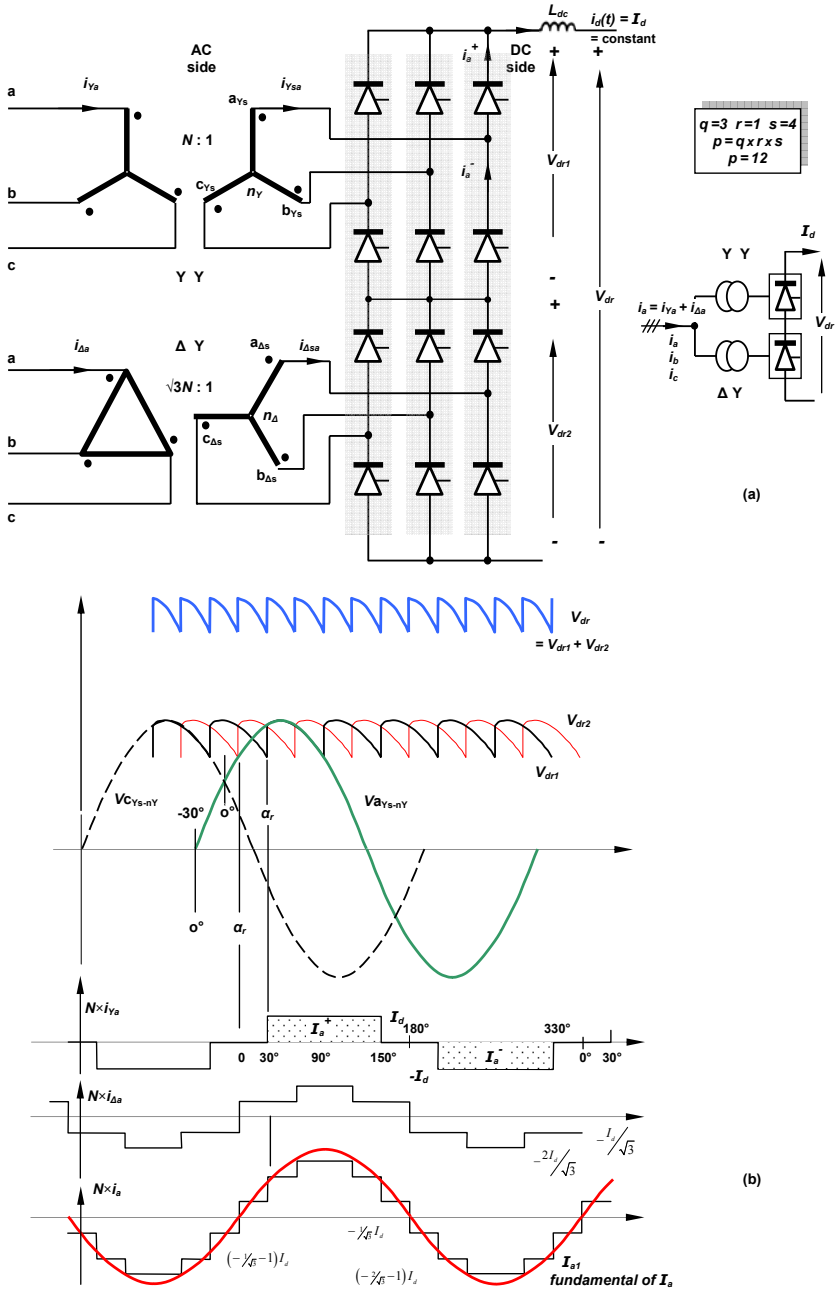


Figure 19.7. *Twelve-pulse valve group converter configuration with star-star and star-delta connected converter transformers: (a) converter schematic and (b) twelve-pulse valve group waveforms.*

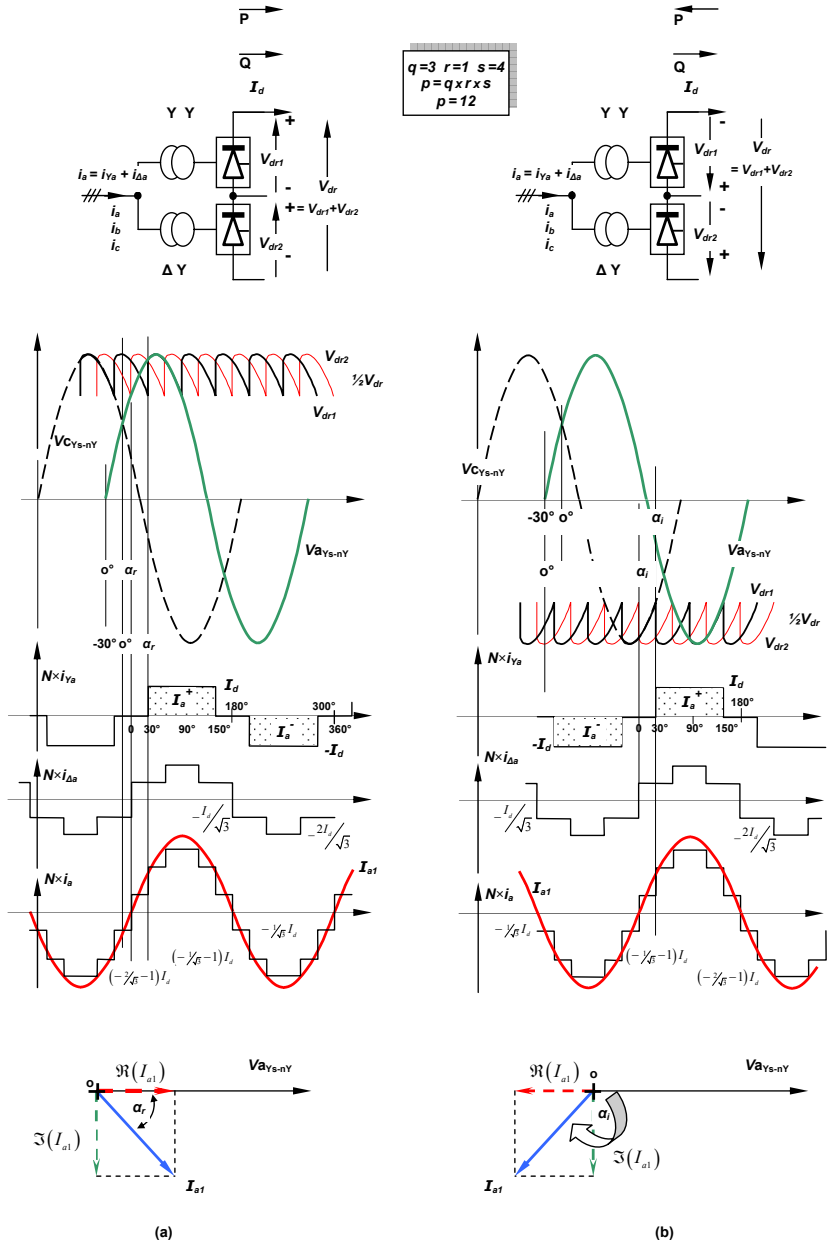


Figure 19.8. *Twelve-pulse valve group converter configuration with star-star and star-delta connected converter transformers operating in: (a) a rectifying mode and (b) an inverting mode.*

19.5 Twelve-pulse ac line frequency converter operation control

Rectification and inversion modes of converter operation and the line (or natural) commutation process of the three-phase fully-controlled thyristor converter have been considered in chapter 12.4 for rectification, with overlap in 12.5, and 12.6 for inversion with overlap.
One converter operates as a rectifier (power flow from ac to dc) and the other dc link converter operates as an inverter (power flow from dc to ac). Either terminal converter can operate as an inverter or rectifier, since the delay angle determines the mode (voltage) of operation. The power flow between the two ac systems connected to the HVDC link is controlled by controlling the delay angle of each converter. Current only flows in one direction in the dc link, from the rectifier to the inverter. A simple system model is shown figure 19.9 where the link dc resistance is represented by R_{dc} and source reactance, hence overlap effects have been neglected. Let the transformer turns ratio factor N equal 1. The dc current I_d is

$$I_d = \frac{V_{dr} + V_{di}}{R_{dc}} \tag{19.23}$$

where

$$V_{dr} = \frac{3}{\pi}\sqrt{2}\,V_{LL}\cos\alpha_r \quad \text{for} \quad 0 \le \alpha_r \le \tfrac{1}{2}\pi$$
$$V_{di} = \frac{3}{\pi}\sqrt{2}\,V_{LL}\cos\alpha_i \quad \text{for} \quad \tfrac{1}{2}\pi \le \alpha_i \le \pi \tag{19.24}$$

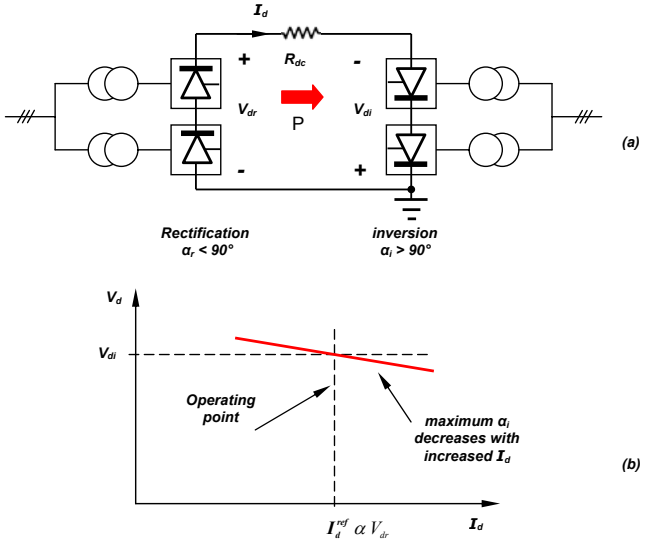


Figure 19.9. *Basic HVDC transmission system: (a) circuit diagram and (b) load line characteristics.*

The rectifier output power is

$$P_r = V_{dr} I_d \tag{19.25}$$

while the power supplied to the inverter is

$$P_i = V_{di} I_d \tag{19.26}$$

where

$$P_r = V_{dr} I_d = V_{di} I_d + I_d^2 R_{dc} = P_i + I_d^2 R_{dc} \tag{19.27}$$

If transformer per phase leakage inductance L_s, referred to the converter side, is accounted for, then the resultant overlap at commutation reduces the output voltage for each six-pulse converter.

$$V_{dr} = 2\times\left(\frac{3}{\pi}\sqrt{2}\,V_{LL}\cos\alpha_r - \frac{3\omega L_s}{\pi} I_d\right) = \frac{6}{\pi}\times\left(\sqrt{2}\,V_{LL}\cos\alpha_r - \omega L_s I_d\right) \tag{19.28}$$

Both the rectifier and inverter dc output voltages can be compensated for leakage reactance commutation overlap. Remember the overlap voltage component is not a loss element in the resistor sense. It represents a 'lossless' loss of voltage which increases with current.
If the link voltage is controlled by the inverter and the dc current controlled by the rectifier, then the load line characteristic in figure 19.9b results. The inverter voltage is kept slightly below the rectifier voltage. As the load current increases, the inverter terminal voltage is reduced. This is because the time to safely commutate the inverter thyristors increases with current, hence α_i must decrease, as shown by the droop in the output characteristics in figure 19.9b. The more detailed practical approach to HVDC power transport control is considered in section 19.5.1.

19.5.1 Control and protection

HVDC transmission systems must operate under tightly controlled conditions. Dc-link current and the two terminal voltages are precisely controlled to affect the desired power transfer. Accurate system quantities measurement are required, which include at each converter bridge, the dc-link current, the dc-side voltages, and the delay angle α for each converter/inverter.
Two terminal dc transmission systems have a preferred control mode during normal operation.

Inverter
Under steady-state conditions, the inverter controls the dc voltage by one of two methods.

- The inverter maintains a constant delay angle $\alpha_i > 90°$, or extinction angle γ, where $\gamma = \pi - \alpha_i$, This constant angle maintenance causes the dc voltage V_d to droop with increasing dc current I_d, as shown in the minimum constant extinction angle $\breve{\gamma}$ characteristic in figure 19.10b and labelled A-B-C-D in figure 19.10c. The weaker the ac system at the inverter, the steeper the droop characteristic.
- Alternatively, the inverter may operate in a dc voltage controlling mode which is the constant V_d characteristic shown dashed in figure 19.10b and labelled B-H-E in figure 19.10c. To achieve this, the extinction angle γ must increase beyond its minimum value characterised in figure 19.10b as $\breve{\gamma}$.

Rectifier
If the inverter is operated in either a minimum constant γ or constant V_d mode, then the rectifier is used to control the dc link current I_d. This is achievable provided the delay angle α_r is not at its minimum limit. The steady-state constant current characteristic of the rectifier is shown in figure 19.10a as the vertical section of the characteristic S-T and is labelled S-C-H-T in figure 19.10c. The rectifier delay angle is increased toward $\alpha_r = 90°$ if the link current attempts to increase beyond the reference level I_d^{ref}. During an attempted short circuit fault, the rectifier delay angle reaches 90° which sets the rectifier output voltage to zero, as shown by equation (19.24), and controls the fault current to I_d^{ref} as shown by trajectory S-T in figure 19.10a.

The operating point of the HVDC system is where the rectifier characteristic intersects the inverter characteristics, either at point C or point H on figure 19.10c, depending on which of the two inverter control methods is being employed. The operating point is tuned over a period of tens of seconds by adjusting the line-side tap changers of the converter transformers.
The inverter controls the dc-link voltage as follows.

- The inverter establishes the desired dc voltage V_d by tap changer adjustment, if it is operated in a constant minimum γ mode.
- If the inverter is operated in a constant V_d mode, the tap changer is adjusted to produce constant V_d with an extinction angle slightly larger than the minimum $\breve{\gamma}$.

The ac-side tap changers on the rectifier end transformers are adjusted so that the delay angle α is small but with a 5° working range, whilst maintain the constant current I_d^{ref}. If the inverter is constant dc voltage operated at the operating point H, and if the dc current I_d^{ref} is increased so that the operating point H moves towards A, the inverter control mode reverts to constant extinction angle γ control when operating in the droop region A-B. The voltage V_d droops to less than the desired value, so the inverter end transformer tap changer progressively boosts the dc-side voltage until dc voltage control recommences, in region B-C-D.

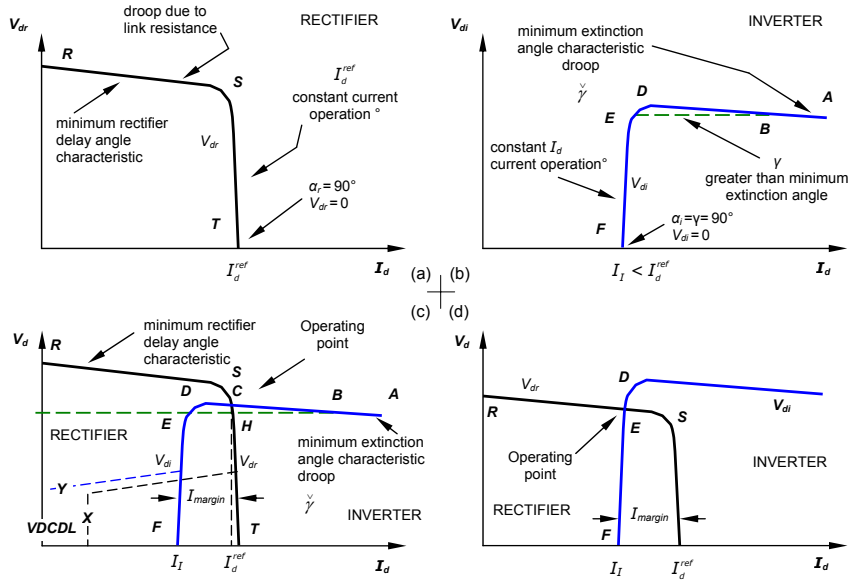


Figure 19.10. *Steady-state V_d - I_d characteristics for a two terminal HVDC system.*

Not all HVDC transmission systems use constant dc voltage control, which is the horizontal characteristic B-H-E in figure 19.10c. Instead, the tap changer in conjunction with the constant extinction angle γ control characteristic A-B-C-D in figure 19.10c, provides dc voltage control.

Current margin
The rectifier and inverter controllers both receive the dc current demand I_d^{ref} but the inverter current demand is decreased by an amount termed the current margin I_{margin}, as shown in Figure 19.10c. This current margin is usually a constant magnitude of about 10% of rated current. The inverter current controller endeavours to control the dc link current to $I_I = I_d^{ref} - I_{margin}$ but the rectifier current controller dominates and maintains the dc current at I_d^{ref}. In steady-state the rectifier controller overrides the inverter controller which is not able maintain a dc current $I_d^{ref} - I_{margin}$. The inverter current controller only becomes active when the rectifier current controller has reduced its delay angle α_r to the minimum limit. This rectifier minimum delay angle limit is characterised by R-S in figure 19.10c.

Control characteristic performance
Variations in the ac voltages change the control operating point of the system as follows.
i. When the rectifier side ac voltage decreases and/or the inverter side ac voltage increases, then the transformer tap adjustment mechanisms on both the rectifier and inverter transformers should attempt to remedy this ac voltage regulation problem. If the disturbance is large enough, the new stable operating point is shown in figure 19.10d. The R-S characteristic falls below points D or E, the operating point will shift from point H to somewhere on the vertical characteristic D-E-F where it is intersected by the lowered minimum α_r R-S characteristic as shown in figure 19.10d. The inverter converts to current control, controlling the dc current I_d to the value $I_I = I_d^{ref} - I_{margin}$, approximately 10% of rated current and the rectifier effectively controls the dc voltage provided it is operating at its minimum delay angle characteristic R-S. The dc power flow is relatively unaffected and safely returns to the normal operating condition shown in figure 19.10c, once the ac disturbances have subsided.
ii. If the rectifier ac voltage increases, the reference current, which is set by the rectifier, is unaffected. Since the link voltage is set by the inverter, any increase in rectifier ac voltage does not affect the power flow. This can be seen in figure 19.10c where increasing the rectifier characteristic R-S does not affect the intersection of the operating point C, whence power flow is unaffected by an increase in the rectifier ac-side voltage.

iii. If the inverter side ac voltage is decreased, the link voltage decreases proportionally, but the current is unaffected because it is set by the rectifier. The power flow is therefore decreased in line with the inverter ac voltage decrease. This can be seen in figure 19.10c where the inverter side ac voltage decrease will lower the inverter characteristic A-B-C-D. The operating point C, the operating voltage, decreases but the current is unaffected. Thus the power transferred is decreased in-line with the inverter ac-side voltage.

vi. Worse case conditions are a dc-link short circuit. As seen in figure 19.10a, the rectifier maximum current is I_d^{ref}, while figure 19.10b shows that the maximum inverter current is I_I. The maximum fault current is therefore limited to the operational current margin I_{margin}. As seen in figures 19.10a and 19.10b, the control angle of both converters moves to 90°, which produces 0V converter output voltages, as shown by equation (19.24). Hence the power associated with any short circuit fault is low, unlike short circuit faults in ac transmission systems.

Voltage dependant current demand limit (VDCDL)

If the ac voltages sag significantly because of weak ac systems, it may not be possible to maintain full load current. In such a case the dc-link voltage is decreased, and the controller characteristics are dictated by the trajectories X and Y in figure 19.10c. A controller which reduces the maximum current demand in such conditions, is termed a voltage dependant current demand limiter, or VDCDL. The current is not reduced to zero so that recovery response is faster once the dc-link voltage has sufficiently recovered.

Power flow reversal

The controllers can be designed such that the transition from the rectifier controlling current to the inverter controlling current is automatic and smooth. That is, seamless automatic power flow reversal is achieved by interchanging the inverter and rectifier functions, as seen in figure 19.11. This is realised by appropriate control of the delay angles, hence terminal polarities, but the dc link current direction does not reverse. Such a bi-directional power flow requirement may be necessary when two ac systems are required to bi-directionally interchange power. The rectifier delay angle is progressively increased while the inverter delay angle is decreased, such that the rectifier and inverter voltage difference is control to be virtually constant. This is achieved if $\alpha_r + \alpha_i \approx 180°$ is maintained.

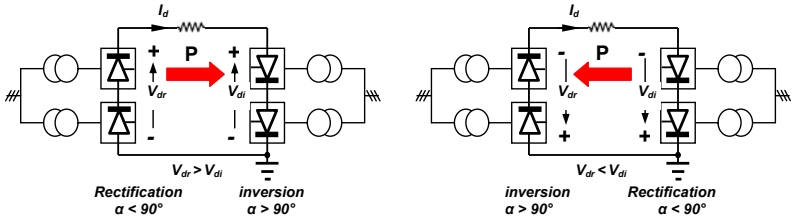


Figure 19.11. *Power reversal in HVDC systems by voltage polarity reversal, not current reversal.*

19.5.2 HVDC Control objectives

The fundamental objectives of a CSI-based HVDC control system are as follows:

- to control basic system quantities such as dc line current, dc voltage, and transmitted power accurately and with sufficient speed of response;
- to maintain adequate commutation margin in inverter operation so that the valves can recover their forward blocking capability after conduction before their voltage polarity reverses;
- to control higher-level quantities such as frequency in isolated mode or provide power oscillation damping to help stabilize the ac network;
- to compensate for loss of a pole, a generator, or an ac transmission circuit by rapid readjustment of power;
- to ensure stable operation with reliable commutation in the presence of system disturbances;
- to minimize system losses and converter reactive power consumption; and
- to ensure proper operation with fast and stable recovery from ac system faults and disturbances.

19.6 Filtering and power factor correction

As shown in figure 19.4, both ac and dc side filtering is used to reduce radiated EMC on the dc link and conducted EMC on the ac side which causes power losses and interference.

On the dc-side, the large link inductance at each converter (typically between 0.2H and 0.5H) is supplemented with *LC* filters, tuned to eliminate selected 12*n* current harmonics in a 12-pulse system. A filter is notch tuned to eliminate (shunt) one specific harmonic, usually the most dominate 12th. At 50/60Hz, these filters are capacitive thus provide reactive power absorbed by the converters.
Harmonics occur at 12*n*±1 on the ac-side. Again, tuned *LC* filters eliminate (shunt) specific low order harmonics and a general high pass shunt filter is used for components above the 11th and 13th. Generally, higher pulse order (>12) transformer/converter arrangements are not attractive in HVDC because of the difficulties in producing high-voltage transformers (auto-transformers tend to be used).
Additional to the VAr compensation provided by the ac harmonic filters, pure capacitance may also be used. In order to avoid overcompensation voltage regulation problems which can occur at low power transmission levels, the extra capacitance tends to be switched in-circuit as needed.

The main transformers may be provided with ac-side voltage taps to adjust the secondary voltage, as considered in section 19.5.1. The taps are switched automatically by motorised tap-changing drives, which only operate when large voltage changes occur for prolonged periods of time. If the transformers have only Y-Y winding configurations, they may also have a low-voltage delta tertiary winding for VAr compensation, provide ancillary supplies, and suppression of transformer core triplen harmonic fluxes.

Example 19.1: ***Basic six-pulse converter based hvdc transmission***

The basic six-pulse converter dc transmission system represented by figure 19.12 connects a 230kV ac rms, 50Hz system to a 220kV, 60Hz system. The 6-pulse converters at each transmission line end are interface by a Δ_{ac}-Y_{dc} transformers of turns ratio √3:2, as shown, such that the transformer dc-side line voltage is double the ac side line voltage. The power transmission is 500MW to the inverter which is maintained at a dc voltage level of 500kV. The total dc-line resistance is 8Ω.

Determine

i. The inverter delay angle, α_i.
ii. The dc-link current, hence rectifier output voltage, thence rectifier delay angle, α_r.
iii. The rectifier input power and VAr, and inverter VAr, thence system efficiency.

Solution

Because of the transformer turns ratio, the transformer dc-side ac voltages are double the ac-side voltages.

i. The inverter delay angle is derived from equation (19.24)

$$V_{di} = \frac{3}{\pi}\sqrt{2}\,V_{LL}\cos\alpha_i$$

$$500\text{kV} = \frac{3}{\pi}\sqrt{2}\;\;2\times 220\text{kV}\times\cos\alpha_i$$

$$\text{that is } \alpha_i = 147.3°$$

ii. From $P_i = V_{di} \times I_d$, the link current is

$$I_d = \frac{500\times10^6\ \text{W}}{500\times10^3\ \text{V}} = 1000\ \text{A}$$

The rectifier output voltage is given by equation (19.23), rearranged

$$V_{dr} = -V_{di} + I_d R_{dc}$$

$$= -\ \text{-500kV + 1000A×1Ω = 508kV}$$

The rectifier delay angle is derived from equation (19.24)

$$V_{dr} = \frac{3}{\pi}\sqrt{2}\,V_{LL}\cos\alpha_i$$

$$508\text{kV} = \frac{3}{\pi}\sqrt{2}\times 2\times 230\text{kV}\times\cos\alpha_r$$

$$\text{yields } \alpha_r = 35°$$

iii. The input power is the output plus dc-link resistive losses, that is

$$P_r = I_d^2 \times R_{dc} + P_i$$
$$= 1000^2 \times 8\Omega + 500\text{MW} = 508\text{MW}$$

The input VAr from the ac side is

$$Q_r = P_r \tan\alpha_r$$
$$= 508\text{MW} \times \tan 35° = 355.7\text{MVAr}$$

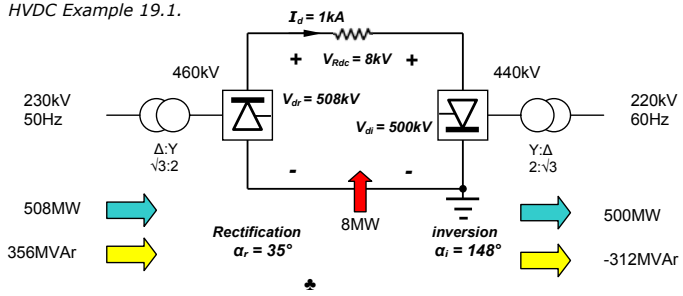
Similarly, the inverter VAr into the ac side (indicated by the negative sign) is

$$Q_i = P_i \tan\alpha_{ir}$$
$$= 500\text{MW} \times \tan 148° = -312.4\text{MVAr}$$

The efficiency is

$$\eta = \frac{P_i}{P_o}$$
$$= \frac{500\text{MW}}{508\text{MW}} \times 100 = 98.43\%$$

Figure 19.12. *HVDC Example 19.1.*



The following example is based on example 19.1.

Example 19.2: ***12-pulse hvdc transmission***

The dc transmission line represented by figure 19.9 connects a 230kV ac rms, 50Hz system to a 220kV, 60Hz system. The 12-pulse bipolar converters at each transmission line end are interface by a Y-Y transformer of turns ratio 1:1 and a Δ-Y transformer of turns ratio 1:√3, each with a converter side inductance of 1mH. The rectifier delay angle is α = 30° for 500MW power transmission and the inverter advance angle is α = 160° (in order to avoid any reactive power increase), which maintains the dc voltage level at 500kV at the inverter end.
The total line resistance is 8Ω and the dc link smoothing inductance is large enough to initially consider the dc current to be ripple free.

Determine

i. The transformer tap ratios at each end, the dc link efficiency, I^2R losses, and both terminal VAr
ii. If the rectifier tap ratio of 0.866 results in the transmission current limit giving a power of 600MW, find the delay angle and line efficiency for 500kV at the inverter.
iii. The value of the dc link inductance L_{dc} such that the link peak to peak current is 0.1pu the average load current at full load (1200A), assuming the normalised magnitude of the dc side harmonic V_{12} is 0.15pu maximum (with respect to the 50Hz supply).

Solution

i. From $P_i = V_{di} \times I_d$, the link current is

$$I_d = \frac{500\times10^6\ \text{W}}{500\times10^3\ \text{V}} = 1000\ \text{A}$$

The inverter voltage, accounting for the transformer tapping is given by an equation similar to equation (19.28), that is

$$V_{di} = 2\times\left(\frac{3}{\pi}\sqrt{2}\,a_i V_{LL}\cos\alpha_r - \frac{3\omega L_s}{\pi} I_d\right)$$
$$500\times10^3\,\text{V} = 2\times\left(\frac{3}{\pi}\sqrt{2}\,a_i\,220\times10^3\,\text{V}\cos(180°-160°) - \frac{3\times2\times\pi\times60\text{Hz}\times1\times10^{-3}\text{H}}{\pi}1000\text{A}\right)$$

which gives a transformer tap ratio at the inverter end of a_i = 0.896.

From equation (19.23), the rectifier voltage is

$$V_{dr} = V_{di} + R_{dc} I_d$$
$$= 500\text{kV} + 8\Omega \times 1000\text{A} = 508\text{kV}$$

and the necessary transformer tap ratio is derive from

$$V_{dr} = 2\times\left(\frac{3}{\pi}\sqrt{2}\,a_r V_{LL}\cos\alpha_r - \frac{3\omega L_s}{\pi} I_d\right)$$
$$508\times10^3\,\text{V} = 2\times\left(\frac{3}{\pi}\sqrt{2}\,a_r\,230\times10^3\,\text{V}\cos 30° - \frac{3\times2\times\pi\times50\text{Hz}\times1\times10^{-3}\text{H}}{\pi}1000\text{A}\right)$$

which gives a transformer tap ratio at the rectifier end of a_r = 0. 945.

The link efficiency is

$$\eta = \frac{P_i}{P_r}\times100 = \frac{V_i I_d}{V_r I_d}\times100$$
$$= \frac{500\text{kV}}{508\text{kV}}\times100 = 98.4\%$$

The I^2R losses are 1000A^2×8Ω = 8MW or (508-500)^2/8 Ω, dissipated, distributed along the line.

The rectifier reactive power is given by equation (19.14), that is

$$Q_r = P_r \tan\alpha_r$$
$$= (500\text{MW} + 1000\text{A}^2\times8\Omega)\times\tan30° = 293.3\text{MVAr}$$

This is 293.3MVAr from to rectifier ac side.

The inverter reactive power is given by equation (19.16), that is

$$Q_i = P_i \tan\alpha_i$$
$$= 500\text{MW} \times \tan160° = -182\,\text{MVAr}$$

This is 182MVAr to the ac side of the inverter.

ii. At 500kV and 600MW:

$$I_d = \frac{P_i}{V_{di}} = \frac{600\text{MW}}{500\text{kv}} = 1200\text{A}$$

Accounting for the link resistive voltage drop

$$V_{dr} = V_{di} + R_{dc} I_d$$
$$= 500\text{kV} + 8\Omega\times1200\text{A} = 509.6\text{kV}$$

The efficiency is

$$\eta = \frac{P_i}{P_r}\times100 = \frac{V_i I_d}{V_r I_d}\times100$$
$$= \frac{500\text{kV}}{509.6\text{kV}}\times100 = 98.1\%$$

The necessary rectifier angle, accounting for transformer reactive inductance, is

$$V_{dr} = 2\times\left(\frac{3}{\pi}\sqrt{2}\,a_i V_{LL}\cos\alpha_r - \frac{3\omega L_s}{\pi} I_d\right)$$
$$509.6\times10^3\,\text{V} = 2\times\left(\frac{3}{\pi}\sqrt{2}\times0.866\times230\times10^3\,\text{V}\cos\alpha_r - \frac{3\times2\times\pi\times50\text{Hz}\times1\times10^{-3}\text{H}}{\pi}\times1200\text{A}\right)$$

which gives a rectifier delay angle of α_r = 18.5°.

iii. The maximum link voltage from the 50Hz rectifier, accounting for leakage at maximum current is

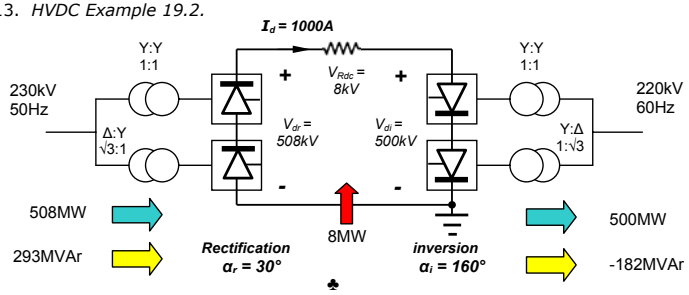
$$V_{dr} = 2\times\left(\frac{3}{\pi}\sqrt{2}\,V_{LL}\cos\alpha_r - \frac{3\omega L_s}{\pi} I_d\right)$$
$$= 2\times\left(\frac{3}{\pi}\sqrt{2}\,230\text{kV}\times\cos0° - \frac{3\times2\pi50\text{Hz}\times1\times10^{-3}\text{H}}{\pi}1200\text{A}\right)$$
$$= 620.5\,\text{kV}$$

The magnitude of the 600Hz component (12×50Hz) is 15% of 620.5kV, namely 93.1kV, which produces a ripple current of 10% of rated current, 1200A, namely 120A. Thus from $v = L di/dt$

$$L_{dc} = v_{12}\frac{\Delta t_{12}}{\Delta i_{12}}$$

$$= 93.1\text{kV} \times \frac{1}{12 \times 50\text{Hz} \times 120\text{A}} = 1.3\text{H}$$

Figure 19.13. *HVDC Example 19.2.*



19.7 VSC-Based HVDC

Voltage source converter-based (VSC) dc-transmission consists of a bipolar two-wire HVDC system with self commutatable converters connected pole-to-pole, as shown in figure 19.14. DC capacitors are used at each VSC dc-side to provide a stiff dc voltage source. The dc capacitors are grounded at their electrical centre point to establish the earth reference potential for filtering and the transmission system. The VSC is effectively mid-point grounded and DC filters and a zero-sequence blocking inductor are used to mitigate interference on any metallic communication circuits adjacent to the DC cables. There is no earth return operation. The converters are coupled to the ac system through ac phase inductors and power transformers, with harmonic filters located between the phase inductor and the transformer. The AC filters are tuned to multiples of the switching frequency, as shown in figure 19.18. This arrangement minimizes harmonic currents and avoids dc voltage stresses in the transformer, which allows use of a standard AC power transformer for matching the 50/60Hz AC network voltage to the converter AC voltage necessary to produce the desired DC transmission voltage.

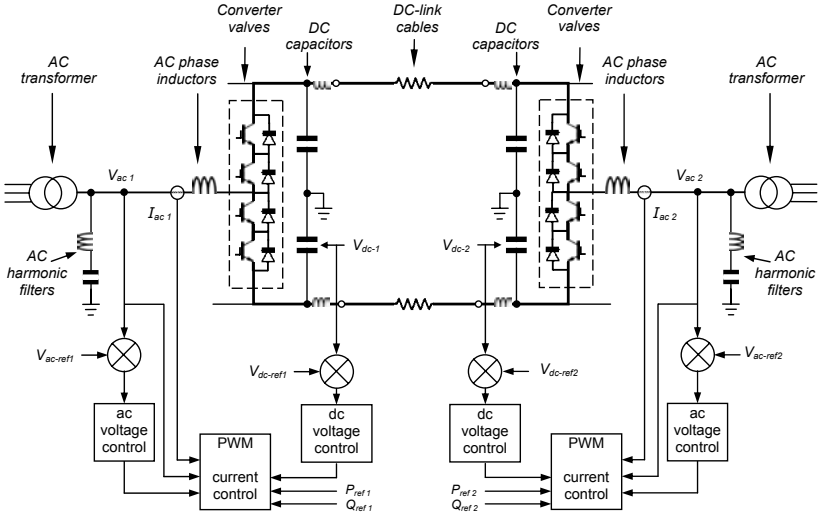


Figure 19.14. *VSC HVDC transmission, using three-level (NPC) voltage source inverters.*

The IGBT valves used in VSC converters are comprised of series-connected IGBT cells. Present technology uses 2.5kV igbt die parallel connected on a common electrically-conducting substrate in 2500A sub-modules, with 30 series connected sub-modules in a string cell. Strings are then series connected to produce the required link valve voltage requirement. The IGBT switching frequency is limited to about 2 kHz. The valves are cooled with circulating water and water to air heat exchangers. The structure is constructed to shield electromagnetic interference (EMI) radiation.

19.7.1 VSC-Based HVDC control

Power flow between the VSC and ac network can be controlled by changing the phase angle of the converter ac voltage V_S with respect to the filter bus ac voltage V_T, whereas the reactive power can be controlled by changing the magnitude of the fundamental component of the converter ac voltage V_T with respect to the filter bus ac voltage V_S. By independently controlling these two aspects of the converter voltage, operation in all four quadrants is possible. This means that the converter can be operated in the middle of its reactive power range near unity power factor to maintain dynamic reactive power reserve for contingency voltage support similar to a static VAr compensator. This also means that the real power transfer can be changed rapidly without altering the reactive power exchange with the ac network or waiting for switching of shunt compensation.

Reactive power control can be used for dynamic voltage regulation to support the ac interconnection, by synthesising a balanced set of three phase voltages. Black start (restoring system operation without an external energy source) capability is also a feature.

Independent control of the VSC ac voltage magnitude and phase relative to the system voltage decouples the active and reactive power control loops for HVDC system regulation. The active power control loop can be set to control either the active power (dc-link current) or the dc-side voltage. In a dc link, one station is selected to control the active power while the other controls the dc-side voltage. The reactive power control loop controls either the reactive power or the ac-side voltage. Either of these two modes can be selected independently at either end of the dc link. Figure 19.14 shows the characteristic ac voltage phasors including the controlled variables V_{dc}, I_d, Q, and V_L.

19.7.2 Power control concept

dc-link power

The dc-link power flow concepts are not complicated by ac phasor considerations. No reactive power is involved with dc, only the real power flows. Consider the HVDC configuration depicted in figure 19.15 involving converter #1 and converter #2.

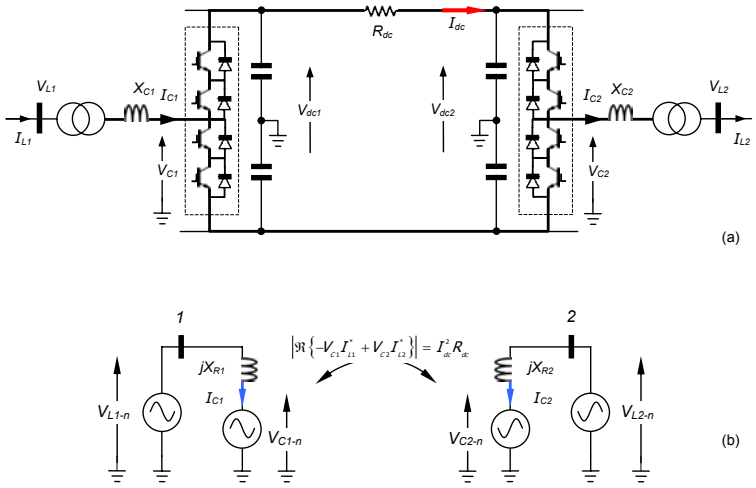


Figure 19.15. *VSC HVDC transmission dc-side, using two-level voltage source inverters.*

If one converter #1 produces an ac voltage represented by

$$V_{C1} = |V_{C1}|(\cos\delta_1 + j\cos\delta_1) = |V_{C1}|\angle\delta_1 \quad (19.29)$$

while the other converter #2 is represented by the voltage source

$$V_{C2} = |V_{C2}|(\cos\delta_2 + j\cos\delta_2) = |V_{C2}|\angle\delta_2 \quad (19.30)$$

Then if the power transmitted equals the power received, then

$$\Re\{-V_{C1}I_{L1}^* + V_{C2}I_{L2}^*\} = 0 \quad (19.31)$$

If the dc link resistive losses are incorporated equations (19.31) becomes

$$\left|\Re\{-V_{C1}I_{L1}^* + V_{C2}I_{L2}^*\}\right| = I_{dc}^2 R_{dc} \quad (19.32)$$

ac-side powers

The fundamental base apparent power S_T at the filter bus between the converter reactor and the AC filter is defined as follows (see figure 19.16):

$$S_T = P_T + jQ_T = \sqrt{3}\times V_T \times I_R \quad (19.33)$$

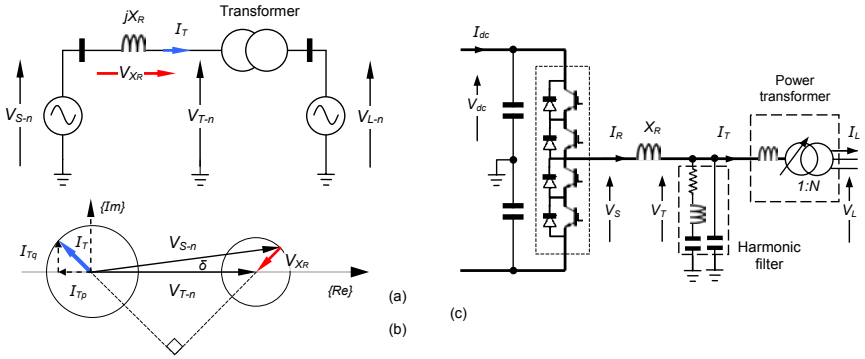


Figure 19.16. *VSC HVDC transmission ac-side.*

The active and reactive power components on the grid-side are defined as (see section 20.3 of ac power transmission):

$$P = \frac{V_T \times V_S \times \sin\delta}{\omega L} = \frac{V_T \times V_S \times \sin\delta}{X_R}$$
$$Q_T = V_T\frac{V_S \times \cos\delta - V_T}{X_R} \quad (19.34)$$

where: δ = phase angle between the filtered voltage V_T and the converter output voltage V_S
L = inductance of the converter ac line inductance

Changing the phase angle δ controls the active power flow P between the converter and the filter bus and consequently between the converter and the AC network.

As shown in figure 19.17a, for active power flows

- If the V_S phase-lags V_T, active power P flows from the AC to the DC side (rectifier).
- If the V_S phase-leads V_T, the active power P flows from the DC to the AC side (inverter).

Changing the amplitude difference between the filter voltage V_T, and the converter voltage V_S controls the reactive power flow between the converter and the AC network.

As shown in figure 19.17b, for reactive power flows

- If $V_T > V_S$, there is reactive power consumed from the ac network.
- If $V_S > V_T$, there is reactive power generated into the ac network.

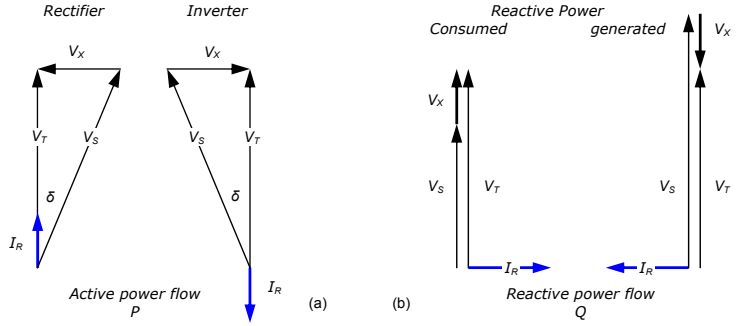


Figure 19.17. *Active and reactive power phasor diagrams.*

With the PWM (Pulse-Width-Modulation, see Section 15.2.3) controlled VSC it is possible to create any phase angle and voltage amplitude (within limits set by the dc-link voltage magnitude) by changing the PWM modulation depth and the relative phase displacement respectively, by using phase-locked-loop grid synchronised displacement. This allows independent control of the active and reactive power.

The typical P-Q diagram, which is valid within the whole steady-state AC network voltage, is shown in the figure 19.18. This figure illustrates the grid real power, P, and reactive power, Q, capability of the HVDC VSC converter terminal, measured at the interconnection point, as a function of ac system voltage.

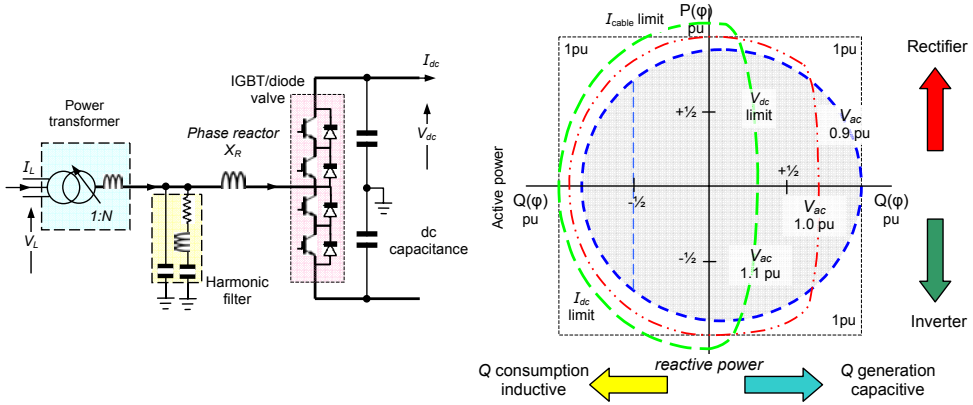


Figure 19.18. *P-Q active and reactive power control locus, showing varying voltage limits.*

The 1st and 2nd quadrants represent rectification and the 3rd and 4th inversion. A positive Q indicates delivery of reactive power to the AC network. Because the dc-link decouples the two converters, reactive power can be controlled independently at each station. There are dc-link voltage (Q generation restrictions) and current (inverter - igbt - power) limitations that have been taken into account in this typical P-Q diagram. Depending on the cable design, the cable dc current maximum may restrict the rectifier power limit. There is also a steady state minimum dc-voltage level limit, which may prevent continuous absorption of large amounts of reactive power.

The capacitive reactive power capability increases with decreasing voltage when it is needed most. Similarly, the inductive reactive power capability increases with increasing network voltage when it is needed most. For a given ac system voltage the converter can be operated at any point within its respective 'circle'. At low ac voltages the VA limit dominates. For high ac voltages, the dc-voltage limit is restrictive but it is not desirable to inject reactive power when ac voltage already is high.

19.8 HVDC Components

i. Power transformer

Because of the use of the converter ac inductors and the use of VSC PWM, the current in the transformer windings contains minimal harmonics and is not exposed to any DC voltage. The transformer is a 50/60Hz single or three phase power transformer, with taps and a tap-changer. The secondary voltage, the filter bus voltage, is controlled by the tap changer to achieve the maximum active and reactive power from the VSC, both consumed and generated. The tap changer is located on the secondary side, which has the largest voltage swing, and also to ensure that the ratio between the line winding and a possible tertiary winding is fixed. This tertiary winding feeds the station auxiliary power system and if delta connected suppresses any core triplen fluxes.
In order to maximize the active power transfer, the converter generates a low frequency zero-sequence voltage (<0.2pu) which is blocked by the ungrounded transformer secondary winding.

ii. Converter line inductors

There is one converter inductor per phase, with EMC shields to eliminate magnetic fields outside the coils. The converter line inductor is a key components in a voltage source converter with which continuous and independent control of active and reactive power is possible.
The main purposes of the converter ac inductors are:

- to provide low-pass filtering of the PWM voltage to give the desired fundamental frequency voltage. The converter generates harmonics related to the switching frequency. The harmonic currents are blocked by the converter inductor and the harmonic content on the AC bus voltage is reduced by an AC shunt filter, which diverts the harmonic currents;
- to provide active and reactive power control. The fundamental frequency voltage across the inductor defines the power flow (both active and reactive) between the AC and DC sides. The P-Q diagram in figure 19.17 shows the active and reactive power definitions;
- to limit line short-circuit currents. The short-circuit voltage of the converter inductor is typically 15%;
- to prevent dc circuit resonance; and
- minimises intermittent dc-link current flow.

The stray capacitance across the line inductor is minimised in order to minimize the harmonics coupled into the ac filter side of the inductor. The high *dv/dt* on the bridge terminals at switching results in current pulses (*i=C dv/dt*) through all stray capacitances to ground. As these currents pass through the valve, they should be minimized. The ac filter side of the inductor is effectively at ground at high frequencies and the capacitance across the inductor should therefore be low. To achieve low capacitance the converter inductors have air coils rather than magnetic cores.

iii. DC-Capacitors

The primary objective of the valve DC-link side capacitor is:

- to provide a low-inductance path for switch turn-off current;
- act as a temporary energy store; and
- to reduce the harmonic ripple on the dc-link voltage. Disturbances in the system, such as AC faults, cause DC-link voltage variations.

The high-voltage DC capacitor uses a dry, self-healing, metallised film design, as opposed to oil filled technologies. The dry capacitor design offers high capacity and low inductance, in a non-corrosion, non-radiating plastic housing.

iv. AC-filters

In a typical HVDC scheme, AC filters contain two or three grounded and/or ungrounded tuned filter branches. A typical second order filter and its frequency characteristics are shown in figure 19.19.
AC filters for VSC HVDC converters have lower ratings than those for conventional HVDC converters and are not required for reactive power compensation. These filters are permanently connected to the converter bus and not switched with transmission loading, as in conventional HVDC.

Voltage source converters can be operated with different control schemes, most of which use pulse width modulation to control the ratios (magnitude and phase) between DC and AC side fundamental frequency voltages. At the AC side converter terminal, the voltage to ground is not sinusoidal. The most frequent application of voltage source converters is as a machine drive (in industrial applications) where this is of less concern. However, connecting a voltage source converter to a transmission or distribution system requires the voltage to be sinusoidal. This necessary filtering is achieved by means of the converter inductor and AC filters.

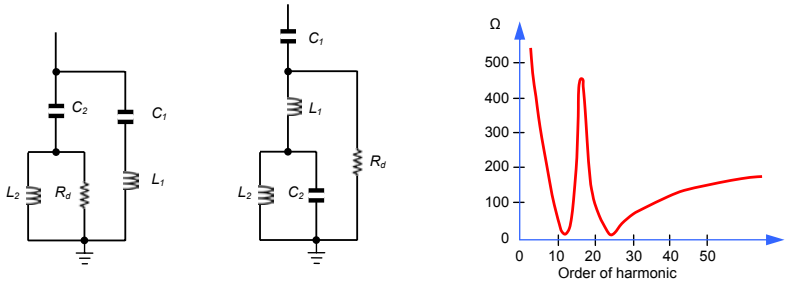


Figure 19.19. *Single phase equivalent circuit diagram and harmonic impedance characteristic of a double tuned filter.*

The distorted waveform of the converter terminal voltage can be described as a series of harmonic voltages

$$V_S = \sum_{h=1} V_h \cos\left(h\omega_o t + \alpha_h\right)$$

where V_h is the h[th] harmonic voltage.

The magnitude of the harmonic voltages vary with the DC voltage, the switching frequency (or pulse number) of the converter and converter PWM technique.
Harmonic cancellation PWM or optimal PWM utilizes a sinusoidal PWM with 3rd harmonic (triplen) injection. A 3rd harmonic is added to the fundamental frequency to increase the voltage, hence power rating of the converter, as considered in chapter 15.1.3vii.

Example 19.3: ***HVDC transmission with voltage source controlled dc-link***

The VSC dc-voltage transmission line represented by figure 19.19 connects a 130kV ac rms, 50Hz three-phase system to a 120kV, 60Hz three-phase system, some 100km apart. The voltage-controlled converter at each nominal ±80kV transmission line end are interface by a Δ-Y transformer with 25mH of line and leakage reactance referred to the grid side, and with a turns ratio to match the converter voltage requirements for the ±80kV dc link. The cable is rated at 625A with 300mm^2 of copper giving a total line resistance is 2Ω.

The 50Hz load end grid draws a total of 100MVA at a 0.75 power factor, while source end VSC draws the power requirement from the 60Hz network at unit power factor.

Determine the phasor diagram for

i. the 50Hz receiving end
ii. the 60Hz transmitting end.

Solution

i. The total active 50Hz power at the load end is

$$P_{50\text{Hz}} = S_{50\text{Hz}} \times pf_{50\text{Hz}}$$
$$= 100\text{MVA} \times 0.75 = 75\text{MW}$$

The 50Hz reactive power flow into the dc link is

$$Q = \sqrt{VAr^2 - P^2}$$
$$= \sqrt{100^2 - 75^2} = 66.1\text{MVAr}$$

The 50Hz ac line current is

$$I_{L50\text{Hz}} = \frac{S_{50\text{Hz}}}{\sqrt{3} \times V_{L50\text{Hz}}}$$
$$= \frac{100\text{MVA}}{\sqrt{3} \times 130\text{kV}} = 441\text{A}$$

The necessary phasor diagram is referenced with respect to neutral.

The reference 50Hz phase voltage is

$$V_{L-N_{50Hz}} = \frac{V_{L_{50Hz}}}{\sqrt{3}} = \frac{130\text{kV}}{\sqrt{3}} = 75.06\text{kV} \angle 0°$$

The angle between the line current and the phase voltage is

$$\theta = \cos^{-1} 0.75 = 41.4°$$

The line reactance at the 50Hz end is

$$X_{L_{50Hz}} = 2\pi f \times L = 2\pi \times 50\text{Hz} \times 25\text{mH} = 7.85\Omega$$

The 50 Hz converter line to neutral voltage (and load angle) is

$$V_{L-N\,conv50Hz} = V_{L-N_{50Hz}} + jX_L I_L = 75\text{kV} + j7.85\Omega \times 441\text{A} \angle -41.4° = 75\text{kV} + 3.45\angle 48.6° = 77.3\text{kV} + \text{j}2.59 = 77.34\text{kV} \angle 1.92°$$

The line-to-line voltage at the 50Hz VSC converter is

$$V_L = \sqrt{3} V_{L-N} = \sqrt{3} \times 77.34\text{kV} = 134\text{kV}$$

ii. The ±80kV dc-link has a pole-to-pole voltage of 160kV. From $P_i = V_{di} \times I_d$, the link current is

$$I_d = \frac{75\text{MW}}{160\text{kV}} = 469.75\text{A}$$

The cable voltage drop is

$$V_{\Delta dc} = I_d \times R_{dc} = 468.25\text{A} \times 2\Omega = 0.9375\text{kV}$$

The VSC converter dc voltage at the 60Hz transmitting end is 160kV + 0.938kV = 160.938kV such the pole voltages are ±80.469kV. These results are shown in figure 19.20.

The VSC power delivered into the link is

$$P_{dc60Hz} = I_d V_{dc60Hz} = 468\text{A} \times 2 \times 80.47\text{kV} = 75.36\text{MW}$$

With unity power at the 60Hz sending end $P = S$, the line current is

$$I_{L60Hz} = \frac{S_{60Hz}}{\sqrt{3} V_L} = \frac{P_{60Hz}}{\sqrt{3} V_L} = \frac{75.36\text{MW}}{\sqrt{3} \times 120\text{kV}} = 362.6\text{A}$$

For the 60Hz transmission end phase or diagram, the line to neutral voltage is required from the 120kV line voltage at the 60Hz end, thus

$$V_{L-N\ 60Hz} = \frac{V_{L60Hz}}{\sqrt{3}} = \frac{120\text{kV}}{\sqrt{3}} = 69.3\text{kV}$$

The 60Hz line reactance is

$$X_{L60Hz} = 2\pi f \times L_{60Hz} = 2\pi \times 60\text{Hz} \times 25\text{mH} = 9.45\Omega$$

The VSC converter side-line to neutral voltage is therefore

$$V_{L-Nconv\,60Hz} = V_{L-N60Hz} - jX_{L60Hz} I_{L60Hz} = 69.3kV - j9.45\Omega \times 362.6\text{A} \angle 0° = 69.3kV + 3.426kV \angle -90° = 69.385\text{kV} \angle -2.83°$$

The line voltage at the sending converter is

$$V_{Lconv\,60Hz} = \sqrt{3} V_{L-Nconv\,60Hz} = \sqrt{3} \times 69.385\text{kV} = 120.18\text{kV}$$

♣

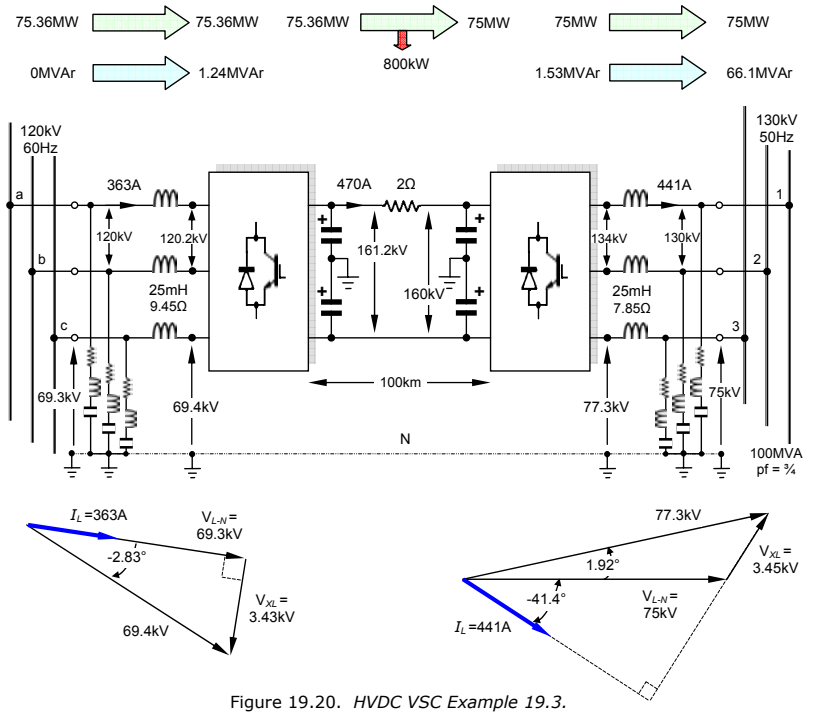


Figure 19.20. *HVDC VSC Example 19.3.*

19.9 Twelve-pulse transformer based HVDC

Figure 19.21 show a HVDC configuration that takes advantage of a star-delta, phase-shifting, transformer secondary to halve the series IGBT connection problems associate with high-voltage applications. The twelve-pulse transformer arrangement also reduces the 5th and 7th ac harmonics due to the cancellation effect because of the phase shift between the two secondary windings. These advantages are traded against to complexity associated with an extra high-voltage transformer winding. A VSC variation is to use multilevel configurations involving chained cells. The cell comprised of a capacitor of large capacitance, which is switched in and out of the series string circuit, much like the H-bridge based cascaded multilevel inverter (see chapter 15.3.3) but only half a bridge is used.

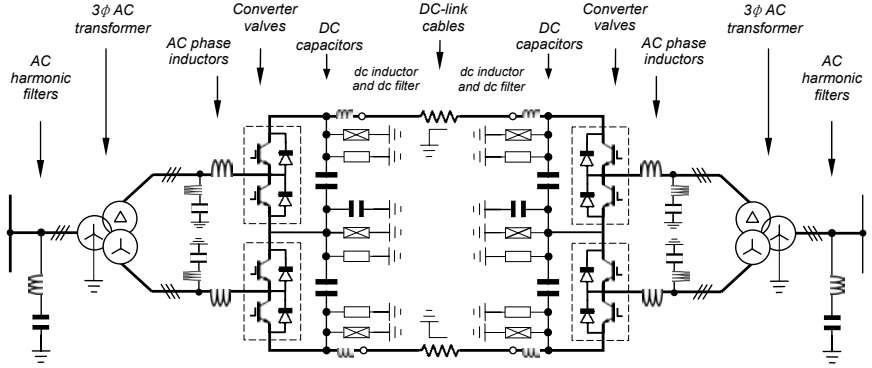


Figure 19.21. *HVDC NPC VSC using twelve-pulse transformer arrangement.*

19.10 HVDC VSC Features

The attributes of HVDC voltage source converters allow for simpler ac system integration, plus the following system technological features.

- Independent, continuous control of active and reactive power at each terminal, that is, four-quadrant operational control, down to very low power levels.
- Steady-state reactive power capability offers voltage control for weak AC grids.
- Independent control of reactive power at each terminal while maintaining full dc voltage for efficient transmission operation.
- Dynamic reactive power reserve capability from the voltage source converters for contingency voltage support of the interconnected ac system. The VSC can be connected to blacked-out networks and re-energize them.
- Less filtering requirements, about 15 to 20% of rated power.
- No requirement for switched filters or shunt capacitor banks for reactive power compensation with changes in power transfer.
- No inherent fault current contribution to increase circuit breaker interrupting duty. Fault current contribution is naturally limited to maximum load current but can be reduced during faults by fast acting VSC control.
- Robust with respect to AC network faults, thereby allowing continuity of power transmission and limitation of spread of system disturbances.
- VSC HVDC is short-circuit proof for line-to-line-to-ground and line-to-line faults on the DC-side.
- Can operate in AC grids with extremely low short-circuit levels or with passive loads.
- Can operate in an unsymmetrical network (e.g. during AC network faults) and provide unbalance control to compensate asymmetrical loads.

The progression and expansion to multi-terminal HVDC systems is hampered by the fact that VSC technology cannot inherently block the ac side from the effects of dc-side faults. This uncontrolled ac to dc rectification limitation necessitates the use of dc circuit breakers, which having a dc current requirement, are technologically challenging.

19.11 Features of conventional HVDC and HVAC transmission

The following are comparison features between ac and conventional dc electrical power transmission:

- Increased power system stability since dc power flow is controlled by converter delay angles. Converter response times are of the order of ms, while the dc link-time constant is significantly longer. Thus, the dc-link short circuit current can be limited.
- The fast converter response times means power reversal can be achieved in ms, much faster than within ac systems.
- Steady-state power flow is related to dc-link resistance, as opposed to ac-line reactance.
- Transmission line inductance has zero impedance to dc, whereas inductive reactance is relatively large in ac systems. A minimal ac skin effect occurs with ac transmission.
- AC power is lost due to dielectric losses. Capacitance between conductors is open circuit to dc, but in ac systems capacitive reactance current paths exist which increase I^2R losses
- Under sea cables have high capacitance which has minimal effect on dc transmission
- DC back-to-back converters (with minimal link inductance, tens of milliHenry's) allow synchronisation between 50Hz and 60Hz systems or same frequency systems at different voltage phase angles
- DC transmission line and its towers are cheaper per unit length than for ac, but dc terminating stations are more complicated and expensive than ac terminating stations
- DC converters have limited overload capability and are less reliable because the ancillaries (filters, cooling management, etc.)
- DC transmission can have a lower visual impact, plus a narrower corridor, hence lower environmental impact than ac systems
- Electric and magnetic fields for dc transmission have lower environmental implications than ac fields
- Corona effects tend to be less significant with dc that for ac conductors
- DC transmission is more robust to lightning effects, since more effective surge protection is used.
- Polymer cables age slower, giving a longer lifetime, with dc
- Maximum voltage gives lower I^2R losses for a given transmitted power level
- No unnecessary energy transfer since no reactive VAr ($S=VI^*=P+jQ$), with dc

The general advantages of ac transmission, over dc transmission, are

- No costs associated with ac-dc-ac conversion equipment
- Transformer (and autotransformer) voltage matching
- Reactive power and harmonics readily compensated
- Not restricted to only point-to-point connection, as is HVDC (there is no fully functioning exception)
- Established system control methods
- No ac transformer dc voltage stressing due to asymmetrical phase control alignment, and no I^2R and core losses due to high harmonic currents
- ac switch gear and breakers, (and particularly vacuum circuit breakers up to 33kV) are very effective - compared with the difficulties in breaking dc current
- Lower current harmonics

Reading list

Mohan, N., *Power Electronics,* 3rd Edition,
Wiley International, 2003.

Acha, E., *et al*., *Power Electronic Control in Electrical Systems*,
Newes, 2002.

Problems

19.1 Show that if a three-phase ac transmission system is converted to a tripole hvdc system with the third conductor transmitting a square-wave current of ±1 pu, that the other two phases can alternately conduct $½\left(\sqrt{3}+1\right)$ and $½\left(\sqrt{3}-1\right)$, yet not exceed their thermal rating.

CHAPTER 20

FACTS Devices and Custom Controllers

Originally electrical power generation, transmission, and distribution systems were direct current. The advent of the three-phase induction motor and the ability of transformers to converter one ac voltage to another ac voltage level (at the same frequency), saw the unassailable rise to dominance of ac electrical power systems. But for long distance electrical power transmission, a dc transmission system is a viable possibility. The highest functional dc voltage for dc transmission, HVDC, is ±600kV over 785km and 805km transmission lines in Brazil. Each of the two bipolar dc transmission systems carry 3.15GW. Also involved are three, three-phase 765kV ac lines which are 1GVAr variable capacitor series compensated (FACTS) at two intermediate substations.

A traditional electric power system is interconnected generating units and load centres via high-voltage electric transmission lines. The system comprises generation, transmission, and distribution subsystems, which usually belonged to the same electric utility in a given area. But the electric power industry has deregulated from large, vertically integrated utilities providing power at regulated rates to an industry that incorporates competitive companies selling unbundled power at potentially lower rates. With this new structure, which includes separate generation, distribution, and transmission companies with an open-access policy, comes the need for tighter system control strategies. The strategies must maintain the reliability level that consumers benefit from and expect, even with the extensive structural changes, such as a loss of a large generating unit or a transmission line, and loading conditions, such as the continuous varying power consumption. Electricity deregulation affects all aspects of the electrical power industry, from generation, to transmission, distribution, and consumption. Transmission circuits, in particular, operate close to their thermal limits because existing transmission lines are loaded near to their stability limits and construction of new transmission circuits is complicated by environmental and political aspects. New equipment and control devices have emerged to control the power flow on transmission lines and to enhance system stability and reliability. Flexible AC transmission systems (FACTS) and FACTS controllers, which are power electronics devices used to control power flow and enhance stability, are replacing mechanical control devices. FACTS play a key role in the operation and control of electrical power systems.

20.1 Flexible AC Transmission Systems - FACTS

FACTS employ power electronics switching devices and circuits to improve and control power flow in ac transmission and distribution systems. The purpose of a FACTS controller is to continuously and rapidly:

- facilitate regulation of the supply voltage to within a specified range;
- allow the transmission line to carry more power, closer to its thermal limit; and
- improve ac system stability, reliability, availability, and security.

In essence, FACTS compensate transmission line reactance (VAr compensation) to yield unity power factor, hence maximum power for a given voltage rating and thermal limit, I^2R. FACTS also increase transmission power by minimising current harmonics, hence associated harmonic I^2R losses, thereby maximising the fundamental current. Additionally, with appropriate control, FACTS devices can actively damp system oscillations, improving network stability margins, thereby allowing increased use of the transmission infrastructure. FACTS can enhance the through-put and capacity of present, new, and upgraded power transmission lines.

20.2 Power Quality

Non-linear and discontinuous loads decrease power quality.
Other than the steady-state voltage magnitude and frequency, power quality is degraded by:

- Voltage sagging – temporarily, a low voltage distribution system user fault interacts with the rest of the network, with an effect that decreases as the distance from the disturbance increases.
- Grounding – user caused common problem due to earth loops, improper connection, and high impedance connection to ground.
- Harmonics – user created due to non-linear loads creating harmonic currents which cause harmonic power losses and harmonic voltage drops across the system impedance.
- Voltage fluctuation and flickering – caused by high power, low frequency (<50Hz) equipment.
- Transients and voltage swelling – in the voltage supply due to load switching at the supply frequency and its harmonics.

These supply problems can be mitigated by a combination of techniques including:

- Passive and active harmonic filters;
- Static and adaptive VAr compensators; and
- Uninterruptible power supplies.

20.3 Principles of Power Transmission

The phasor diagram for one phase of a three-phase transmission system of figure 20.1a is shown in figure 20.1b, where it is assumed the load has a lagging power factor angle ϕ. The ac line is represented by its lumped series impedance (Thevenin's short circuit impedance) $Z_L = R_L + jX_L$. The apparent (complex) power leaving the sending end bus and flowing to the terminal (or receiving) bus is

$$S_S = P_S + jQ_S = \mathbf{V_S I_L^*} = V_S I\cos\phi + jV_S I\sin\phi$$

$$\text{where}\quad I_L = I = I_P + jI_Q = \frac{\mathbf{V_s - V_T}}{Z_L} = \left(\mathbf{V_s - V_T}\right) Y_L = \left(\mathbf{V_s - V_T}\right)\left(G_L + jB_L\right) \tag{20.1}$$

$$\text{where}\quad Z_L = \frac{1}{Y_L} = R_L + jX_L \text{ and } G_L = \frac{R_L}{R_L^2 + X_L^2} \quad B_L = -\frac{X_L}{R_L^2 + X_L^2}$$

$$S_S^* = P_S - jQ_S = \mathbf{V_S^* I_L} = \mathbf{V_S^*}\frac{\left(\mathbf{V_S - V_T}\right)}{Z} = \frac{\mathbf{V_S^2 - V_S^* V_T}}{Z} = \left(\mathbf{V_S^2 - V_S^* V_T}\right)\left(G_L + jB_L\right) \tag{20.2}$$

where $\mathbf{V_S^* V_T} = V_S V_T\left(\cos\delta - j\sin\delta\right)$

The sending real and reactive components are

$$P_S = V_S\frac{\left(V_S - V_T\cos\delta\right)R_L}{Z_L^2} + V_S V_T\frac{X_L}{Z_L^2}\sin\delta = V_S\left(V_S - V_T\cos\delta\right)G_L - V_S V_T B_L\sin\delta$$

$$Q_S = V_S\frac{V_S X_L - V_T R_L\sin\delta}{Z_L^2} - V_S V_T\frac{X_L}{Z_L^2}\cos\delta = -V_S^2 B_L - V_S V_T G_L\sin\delta + V_S V_T G_L\cos\delta \tag{20.3}$$

where $Z_L^2 = R_L^2 + X_L^2$ and $\delta = \delta_S - \delta_T$ is the power angle.

The real and reactive components received at the terminal bus are

$$P_T = V_T\frac{\left(-V_T + V_S\cos\delta\right)R_L}{Z_L^2} + V_S V_T\frac{X_L}{Z_L^2}\sin\delta = V_T\left(-V_T + V_S\cos\delta\right)G_L - V_S V_T B_L\sin\delta$$

$$Q_T = V_T\frac{-V_T X_L - V_S R_L\sin\delta}{Z_L^2} + V_S V_T\frac{X_L}{Z_L^2}\cos\delta = V_T\left(V_T B_L - V_S G_L\sin\delta\right) - V_S V_T B_L\cos\delta \tag{20.4}$$

The real and reactive line 'losses' are the difference between the power sent and the power received:

$$P_L = P_S - P_T = \frac{\left(V_S^2 + V_T^2\right)R_L}{Z_L^2} - 2V_S V_T\frac{R_L}{Z_L^2}\cos\delta = \left(V_S^2 + V_T^2\right)G_L - 2V_S V_T G_L\cos\delta$$

$$Q_L = Q_S - Q_T = \frac{\left(V_S^2 + V_T^2\right)X_L}{Z_L^2} - 2V_S V_T\frac{X_L}{Z_L^2}\cos\delta = -\left(V_S^2 + V_T^2\right)B_L + 2V_S V_T B_L\cos\delta \tag{20.5}$$

Since the resistive voltage drop is usually much smaller than the reactive voltage, R_L is neglected (that is ratio R/X is small). The transmission line voltage drop, which is the reactive voltage $j X_L I$, leads the current phasor I by 90°. The angle between the sending voltage V_S, δ_S, and the terminal (or receiving) voltage V_T, δ_T, is the transmission *load angle*, δ.

The power flow equation, with $R_L = 0$, becomes

$$P_S = P_T = P = \frac{V_S V_T}{X_L} \sin\delta \qquad (20.6)$$

and the reactive power, with $R_L = 0$, for the terminal and sending ends are

$$Q_T = V_T \frac{V_S \cos\delta - V_T}{X_L} \qquad (20.7)$$

$$Q_S = V_S \times \frac{V_S - V_T \cos\delta}{X_L} \qquad (20.8)$$

The average reactive power flow is

$$Q_{ST} = ½(Q_S + Q_T) = ½\frac{V_S^2 - V_T^2}{X_L} \qquad (20.9)$$

While the line inductance absorbed reactive power is $Q_{XL} = Q_S - Q_T = I^2 X_L = V_L^2 / X_L$.
The transmission line midpoint voltage (for compensation) is given by

$$V_M = ½(V_T + V_S) \quad \angle ½\delta \qquad (20.10)$$

Reactive and real power dependence on load angle is shown in figure 20.1c, where from equation (20.6) maximum power of $\hat{P} = V_S V_T / X_L$ occurs at $\delta = 90°$. Maximum VA (generating) of $\hat{Q}_S = V_S \times (V_S + V_T) / X_L$ occurs at the sending end for δ = 180°, with a maximum (absorbing) at the terminal end, where $|Q_T|_{max} = V_T \times (V_S + V_T) / X_L$.

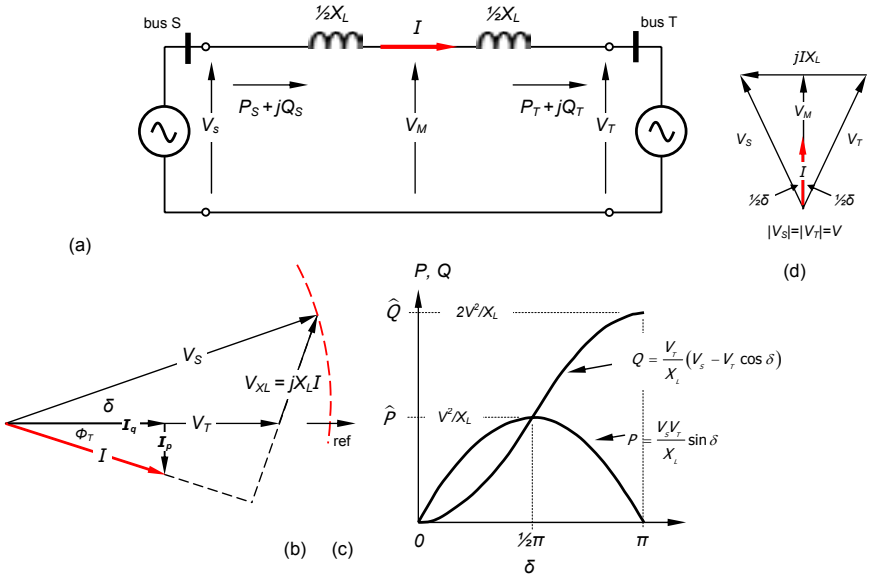


Figure 20.1. *Basic HVAC transmission system: (a) circuit diagram; (b) phasor diagram; (c) power and VAr versus load angle load δ (maximum when $|V_S|=|V_T|$); and (d) phasor diagram for $|V_S|=|V_T|$.*

For simplicity, if $|V_S| = |V_T| = V$, then, using the midpoint voltage V_M as reference (see figure 20.1d), equations (20.6) to (20.10) simplify to

$$\mathbf{I} = \frac{\mathbf{V_S} - \mathbf{V_T}}{X_L} = \frac{2V}{X_L} \sin ½\delta \quad \angle ½\pi \qquad \mathbf{V_M} = \frac{\mathbf{V_S} + \mathbf{V_T}}{2} = V \cos ½\delta \quad \angle 0$$

$$P_T = |\mathbf{V_M}||\mathbf{I}| = \frac{V^2}{X_L} \sin\delta \qquad Q = Q_S = -Q_T = V|\mathbf{I}| \sin ½\delta = \frac{V^2}{X_L}(1 - \cos\delta) \qquad (20.11)$$

As shown in figure 20.1c, the maximum active power is $\hat{P} = V^2 / X_L$ at δ = 90° whilst the maximum reactive power is $\hat{Q} = 2V^2 / X_L$, at δ = 180°.

Equation (20.6) can be used to specify how power and line current flow can be controlled (increased).

- Increase the voltage magnitudes at either or both ends, with voltage support or applying a shunt voltage V_M at the midpoint,
- Reduce the line reactance X_S, by (series) line compensation
- Increase the power load angle, by inserting series variable voltage to give a phase shift

The power flow can be reversed by changing the sign of the power angle; that is, a positive power angle $\delta > 0$ corresponds to a power flow from the sending end to the receiving bus, whereas a negative power angle $\delta_R > \delta_S$ corresponds to a power flow from the receiving to the sending bus.

Similarly, from equation (20.9), both voltage magnitudes and line reactance affect the reactive power. If both voltage magnitudes are the same, that is, a flat voltage profile, each bus will send half of the reactive power absorbed by the line. The power flow is in a sending to receiving direction when $V_R < V_S$. Hence, the four parameters that affect real and reactive power flows are V_S, V_R, X_L, and δ. At the receiving end T, equations (20.6) and (20.8) can be combined:

$$P^2(\delta) + \left(Q_T(\delta) + \frac{V_T^2}{X_L}\right)^2 = \left(\frac{V_S V_T}{X_L}\right)^2 \qquad (20.12)$$

This equation represents a circle, centred $(0, -V_T^2 / X_L)$, radius $V_S V_T / X_L$. It relates real and reactive powers received at bus T to the four parameters: V_S, V_T, δ, X_L.

Similarly, the relationship between the real and reactive powers sent to the line from the sending bus S can be expressed as

$$P^2(\delta) + \left(Q_S(\delta) - \frac{V_S^2}{X_L}\right)^2 = \left(\frac{V_S V_T}{X_L}\right)^2 \qquad (20.13)$$

Equations (20.7) and (20.8) show that any change in active power, changes the reactive power requirements of both the sending and terminal ends.

Example 20.1: ***AC transmission line VAr***

If the line reactance of X_L = 0.05pu in conjunction with operating conditions result in a terminal voltage which is 5% less than the sending end, for 1pu power flow calculate:

i. The load angle δ
ii. The sending and terminal end reactive power requirements, Q_S, Q_T
iii. The line current, hence the line reactive power, Q_{XL}
iv. The midpoint shunt voltage that can be inserted that does not change operating conditions

Solution

i. From equation (20.6) rearranged, the load angle is

$$\sin\delta = P_T \times \frac{X_L}{V_S V_T} = 1.0 \times \frac{0.05}{1.0 \times 0.95} = 0.0526$$

That is δ = 3.0° and cosδ = 0.9986.

ii. From equation (20.8), the sending end VAr is

$$Q_s = V_S \times \frac{V_S - V_T \cos\delta}{X_L}$$

$$= 1.0 \times \frac{1.0 - 0.95 \times 0.9986}{0.05} = 1.026\text{pu}$$

Since Q_S is positive, the sending end is generating VAr's, that is, the power factor is lagging at the sending end (I is lagging V_S).
The terminal end VAr is given by equation (20.7)

$$Q_T = V_T \frac{V_S \cos\delta - V_T}{X_L}$$
$$= 0.95 \times \frac{1.0 \times 0.9986 - 0.95}{0.05} = 0.923\text{pu}$$

Since Q_T is positive, the terminal end is absorbing VAr's, that is, the power factor is lagging at the terminal end (I is lagging V_T).

iii. The current can be evaluated by equating equation (20.1) with equations (20.6) and (20.7)

$$I = I\cos\phi + jI\sin\phi = I \times e^{j\phi}$$

$$I\sin\phi = \frac{V_S\cos\delta - V_T}{X_L} \qquad I\cos\phi = \frac{V_S\sin\delta}{X_L}$$
$$= \frac{1.0\times0.9986 - 0.95}{0.05} \qquad = \frac{1.0\times0.0526}{0.05}$$
$$= 0.972\text{pu} \qquad = 1.0526\text{pu}$$

$$I = 0.972 + \text{j}1.0526\,\text{pu} = 1.433\angle 42.3^\circ\,\text{pu} \quad \text{wrt} \quad V_T$$

The line reactive power is given by $Q_S - Q_T$ = 1.026 - 0.923 = 0.103pu. Alternatively, the line reactive power can be calculated from $I^2X_L = 1.433^2 \times 0.050 = 0.103$pu.

iv. The midpoint voltage is given by equation (20.10)

$$V_M = \tfrac{1}{2}(V_T + V_S) \quad \angle\tfrac{1}{2}\delta$$
$$= \tfrac{1}{2}(0.95 + 1.0) = 0.975\text{pu} \quad \angle 1.5^\circ \quad \text{wrt } V_s$$

If a voltage source with this magnitude and angle, with respect to the sending end, is shunt connected at the midpoint, then no current flows, hence no active or reactive power change occurs.

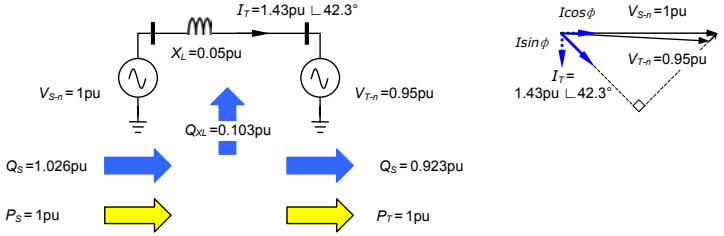


Figure 20.2. *Example 20.1.*

♣

20.4 The theory of instantaneous power in three-phase

Figure 20.3 shows a three-phase three-wire system in the *a-b-c* coordinates, where no zero-sequence voltage need be included in the three-phase three-wire system ($e_a + e_b + e_c = 0$ and $i_a + i_b + i_c = 0$). Three-phase voltages and currents in *a-b-c* coordinates, shown in figure 20.3, can be transformed into the two-phase voltages and currents in *α-β* coordinates, assuming the *a* and *α* axes coincide ($\theta_{a\alpha}$=0), as follows:

$$\begin{bmatrix} e_\alpha \\ e_\beta \end{bmatrix} = \sqrt{\tfrac{2}{3}} \begin{bmatrix} 1 & -\tfrac{1}{2} & -\tfrac{1}{2} \\ 0 & \tfrac{1}{2}\sqrt{3} & -\tfrac{1}{2}\sqrt{3} \end{bmatrix} \begin{bmatrix} e_a \\ e_b \\ e_c \end{bmatrix} \tag{20.14}$$

$$\begin{bmatrix} i_\alpha \\ i_\beta \end{bmatrix} = \sqrt{\tfrac{2}{3}} \begin{bmatrix} 1 & -\tfrac{1}{2} & -\tfrac{1}{2} \\ 0 & \tfrac{1}{2}\sqrt{3} & -\tfrac{1}{2}\sqrt{3} \end{bmatrix} \begin{bmatrix} i_a \\ i_b \\ i_c \end{bmatrix} \tag{20.15}$$

The instantaneous real power, p, either in *a-b-c* coordinates or in *α-β* coordinates is defined by

$$p = e_a i_a + e_b i_b + e_c i_c = e_\alpha i_\alpha + e_\beta i_\beta = dW/dt \tag{20.16}$$

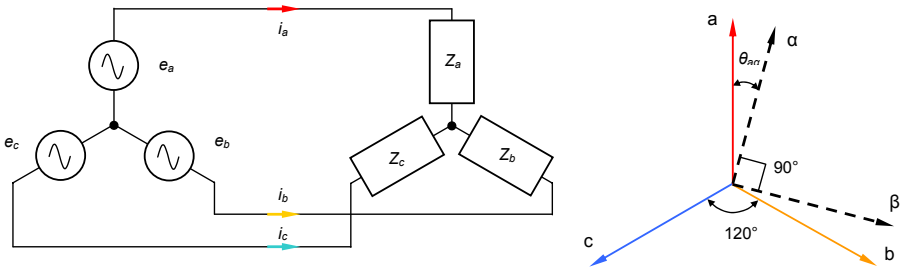


Figure 20.3. *Phase voltages and line currents in a three-phase three-wire system and transformation of a-b-c coordinates into α-β-0 coordinates.*

The three-phase instantaneous imaginary power, q, is defined by

$$q = e_\alpha i_\beta - e_\beta i_\alpha = \tfrac{1}{\sqrt{3}}\left(i_a(v_c - v_b) + i_b(v_a - v_c) + i_c(v_b - v_a)\right) \tag{20.17}$$

In matrix form, the two powers are:

$$\begin{bmatrix} p \\ q \end{bmatrix} = \begin{bmatrix} e_\alpha & e_\beta \\ -e_\beta & e_\alpha \end{bmatrix} \begin{bmatrix} i_\alpha \\ i_\beta \end{bmatrix} \tag{20.18}$$

$e_\alpha \cdot i_\alpha$ and $e_\beta \cdot i_\beta$ are the instantaneous real powers in the *α*-phase and the *β*-phase because both are defined as the product of the in-phase instantaneous voltage in one phase and the instantaneous current in the same phase. $e_\alpha \cdot i_\beta$ and $e_\beta \cdot i_\alpha$ are the instantaneous reactive powers because defined by the product of the instantaneous voltage in one phase and the instantaneous current in the other phase, which is at quadrature

Since e_α and e_β are at quadrature, the determinant of the voltage matrix in equation (20.18) is non-zero, hence the inverse of equation (20.18) always exists:

$$\begin{bmatrix} i_\alpha \\ i_\beta \end{bmatrix} = \begin{bmatrix} e_\alpha & e_\beta \\ -e_\beta & e_\alpha \end{bmatrix}^{-1} \begin{bmatrix} p \\ q \end{bmatrix} = \frac{1}{e_\alpha^2 + e_\beta^2} \begin{bmatrix} e_\alpha & -e_\beta \\ e_\beta & e_\alpha \end{bmatrix} \begin{bmatrix} p \\ q \end{bmatrix} \tag{20.19}$$

The instantaneous currents in *α-β* coordinates, i_α and i_β, can be separated into two instantaneous current components:

$$\begin{bmatrix} i_\alpha \\ i_\beta \end{bmatrix} = \begin{bmatrix} e_\alpha & e_\beta \\ -e_\beta & e_\alpha \end{bmatrix}^{-1} \begin{bmatrix} p \\ 0 \end{bmatrix} + \begin{bmatrix} e_\alpha & e_\beta \\ -e_\beta & e_\alpha \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ q \end{bmatrix}$$
$$\equiv \begin{bmatrix} i_{\alpha p} \\ i_{\beta p} \end{bmatrix} + \begin{bmatrix} i_{\alpha q} \\ i_{\beta q} \end{bmatrix} \tag{20.20}$$

Let the instantaneous powers in the *α*-phase and the *β*-phase be p_α and p_β, respectively, and are given by the conventional definition:

$$\begin{bmatrix} p_\alpha \\ p_\beta \end{bmatrix} = \begin{bmatrix} e_\alpha i_\alpha \\ e_\beta i_\beta \end{bmatrix} = \begin{bmatrix} e_\alpha i_{\alpha p} \\ e_\beta i_{\beta p} \end{bmatrix} + \begin{bmatrix} e_\alpha i_{\alpha q} \\ e_\beta i_{\beta q} \end{bmatrix} \tag{20.21}$$

From equations (20.20) and (20.21), the three-phase instantaneous real power, p, is:

$$p = p_\alpha + p_\beta = e_\alpha i_{\alpha p} + e_\beta i_{\beta p} + e_\alpha i_{\alpha q} + e_\beta i_{\beta q}$$
$$= \frac{e_\alpha^2}{e_\alpha^2 + e_\beta^2} p + \frac{e_\beta^2}{e_\alpha^2 + e_\beta^2} p + \frac{-e_\alpha e_\beta}{e_\alpha^2 + e_\beta^2} q + \frac{e_\alpha e_\beta}{e_\alpha^2 + e_\beta^2} q \tag{20.22}$$
$$= p_{\alpha p} + p_{\beta p} + p_{\alpha q} + p_{\beta q}$$

where $i_{\alpha p}$, $p_{\alpha p}$ are the instantaneous active current and active power on the *α* axis,
$i_{\beta p}$, $p_{\beta p}$ are the instantaneous active current and active power on the *β* axis,
$i_{\beta p}$, $p_{\beta p}$ are the instantaneous reactive current and reactive power on the *α* axis,
$i_{\beta p}$, $p_{\beta p}$ are the instantaneous reactive current and reactive power on the *β* axis,

The sum of the third and fourth terms on the right-hand side in equation (20.22) is always zero. From equations (20.21) and (20.22):

$$p = e_\alpha i_{\alpha p} + e_\beta i_{\beta p} \equiv p_{\alpha p} + p_{\beta p} \quad (20.23)$$

$$0 = e_\alpha i_{\alpha q} + e_\beta i_{\beta q} \equiv p_{\alpha q} + p_{\beta q} \quad (20.24)$$

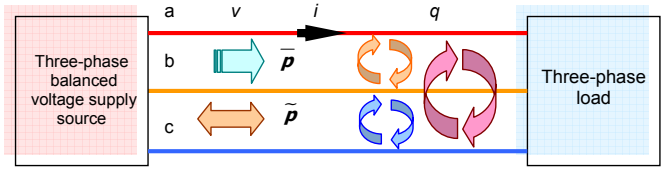


Figure 20.4. *Graphical representation of the three phase power components decomposed into p-q power components.*

Equations (20.23) and (20.24) imply the following conclusions:

- The sum of the power components, $p_{\alpha p}$ and $p_{\beta p}$, is the three-phase instantaneous real power, p, given by equation (20.16). Therefore, $p_{\alpha p}$ and $p_{\beta p}$ are referred to as the α-phase and β-phase instantaneous active powers.
- The other power components, $p_{\alpha q}$ and $p_{\beta q}$, cancel, making no contribution to the instantaneous power flow from the source to the load. Therefore, $p_{\alpha q}$ and $p_{\beta q}$ are referred to as the α-phase and β-phase instantaneous reactive powers.
- Thus, for example, in the case of a shunt active filter without energy storage, instantaneous compensation of the current components, $i_{\alpha q}$ and $i_{\beta q}$ or the power components, $p_{\alpha q}$ and $p_{\beta q}$ can be achieved. The theory based on equation (20.18) reveals the components that can eliminate from the α-phase and β-phase instantaneous currents, i_α and i_β or the α-phase and β-phase instantaneous real powers, p_α and p_β.

When the load is non-linear and unbalanced the real and imaginary powers can be split into average and oscillating components, as follows:

$$p = \bar{p} + \tilde{p} = \bar{p} + \tilde{p}_h + \tilde{p}_{2f1}$$

and

$$q = \bar{q} + \tilde{q} = \bar{q} + \tilde{q}_h + \tilde{q}_{2f1}$$

where $\bar{p}$ and $\bar{q}$ are average components
$\tilde{p}_h$ and $\tilde{q}_h$ are oscillating components
h harmonic
$\tilde{p}_{2f1}$ and $\tilde{q}_{2f1}$ oscillating components with *2f1* being twice the fundamental frequency

The corresponding three phase currents, associated with the average and oscillating power components, decompose into

$$i = i_{\bar{p}} + i_{\tilde{q}} + i_h + i_{2f1}$$

From these power components, the current components in *a-b-c* coordinates can be calculated from equation (20.19). Instantaneous power theory as presented should not be used with unbalanced or distorted supply voltages. When a linear load is supplied with distorted periodic voltage, the distortions caused by the supply voltage harmonics are still present in the source current after compensation.

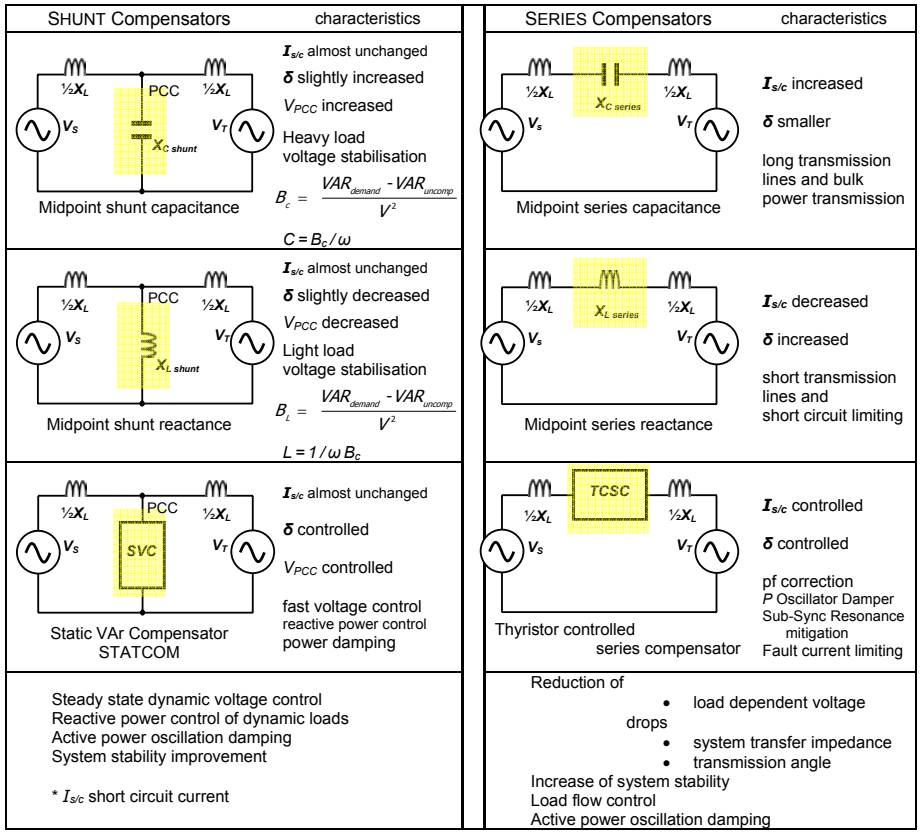


Figure 20.5. *HVAC transmission system reactive power shunt and series compensation methods.*

20.5 FACTS Devices

FACTS technology may be divided into two principal families:

i. line-commutated, thyristor based devices and
ii. self-commutated IGBT/IGCT devices.

i. Line-commutated FACTS

Line commutated FACTS may be considered as providing a means of inserting variable impedance either in parallel (static VAr compensator, SVC) or in series (thyristor switched and thyristor controlled series compensation TSSC/TCSC). The use of thyristor technology readily achieves high-power handing capability with low losses and robust overload capability.
These devices have relatively slow response times, of the order of several ac cycles, due to the limits of line frequency switching and the inherent time constant of the thyristor controlled reactive component. Line frequency switching imposes a need for filters and damping networks to eliminate harmonics and low multiples of the power frequency. The FACTS response allows for compensation of sub-cycle transients but they do not have the bandwidth to compensate for higher frequency disturbances.

ii. Self-commutating FACTS devices

Unlike line-commutated devices, self-commutating FACTS act as controlled energy sources which are capable of injecting voltage or current at the point of common coupling (PCC). This mechanism provides better decoupling between the compensation function and network conditions. Such FACTS devices employ switching devices capable of switching at high multiples of the power frequency (typically in the

range of 1 to 2kHz). This allows elimination of the low order harmonics associated with line-commutated systems. If required this increased bandwidth may be used to achieve active management of harmonics and transients at frequencies above the power frequency (active power filters, APF). Self-commutating FACTS operate as controlled sources that may inject shunt current (STATCOM) or series voltage (dynamic voltage restorer, DVR). Simultaneous series and shunt compensation may be achieved through the integration of both shunt and series devices (unified power flow controller, UPF).

Systems generally use pulse-width modulated (PWM) voltage source inverter (VSI) technology, similar to that employed in variable speed drives. However, since FACTS do not contribute real power, no external power source is required.

VSI based FACTS devices achieve faster response times, improved transient response and reduced size relative to thyristor based systems. The size reduction results from the reduction in mains frequency rated reactive components. The use of PWM semiconductor switched devices increases losses both as a result of increased device conduction loss (relative to thyristors) and the increased loss associated with a high PWM switching frequency. Although low-frequency power harmonics are absent from the output spectrum, the output does contain harmonics at the switching frequency which must be removed using passive filters. These filters are smaller than those required for thyristor systems, however they may contribute to system resonances and incur damping loss.

Advances in self commutating FACTS devices

Since FACTS devices do not contribute to the principal power flow, there is the option for transformer matching between the network voltage and the ratings of power semiconductors. This allows the use of conventional two-level VSI technology, which differs from HVDC where there is a basic requirement for high-voltage conversion systems.

Raising the operating voltage of self-commutating FACTS has benefits in terms of increased VAr capability and direct transformerless connection. Increased operating voltage is achieved though series connection of semiconductors or by means of multi-level converters.

Multi-level converters synthesis an output voltage comprised of a number of discrete steps, each of which is within the voltage rating of each individual power semiconductor device. This technique extends the achievable operating voltage, resulting in significant improvements in waveform quality, reduced filter size, and decreased losses. Intermediate voltage levels are provided by capacitors in a similar manner to the dc link capacitor of a conventional two-level inverter. However these capacitors require continuous charge balancing and must be sized according to the principal fundamental current; unlike a conventional two-level inverter, where the capacitance experiences only the switching frequency and unbalance components. Power circuits and control of multi-level converters are more complex than those of two level systems.

Use of FACTS devices to improve network stability

FACTS devices have the ability to damp network oscillations through the selective sourcing and sinking of reactive power. To achieve this, the ratings of the FACTS reactive storage components must be sized such that sufficient energy may be stored and released in anti-phase with the oscillation.

20.6 Static Reactive Power Compensation

In the steady state FACTS devices are used to manage power flows by manipulating the reactive power and impedance seen at different points on the network. There are a number of basic modes to affect static VAr compensation in a transmission system:

i. shunt compensation -
   - thyristor controlled reactor (TCR)
   - thyristor switched capacitor (TSC)
   - hybrid parallel connect TCR and TSC, termed a static VAr compensator (SVC)
   - Voltage Source Inverter (STATCOM)

ii. series compensation -
   - thyristor switched series capacitor (TSSC)
   - thyristor controlled series capacitor (TCSC)
   - hybrid parallel connected TCR and C, termed a static series VAr compensator (SVC)

iii. static phase shift compensator (SPSC)

iv. dynamic voltage restorer (DVR)

v. combined shunt and series compensation, the unified power flow controller (UPFC)

Parallel compensation is defined as any reactive power compensation utilising either switched or controlled devices, which are shunt connected at a selected network node, called the point of common coupling, (PCC) of the transmission system.

Series compensation is defined as any reactive power compensation utilising either switched or controlled devices, which are series connected into the transmission line at a selected node, called the point of common coupling, (PCC) of the transmission system.

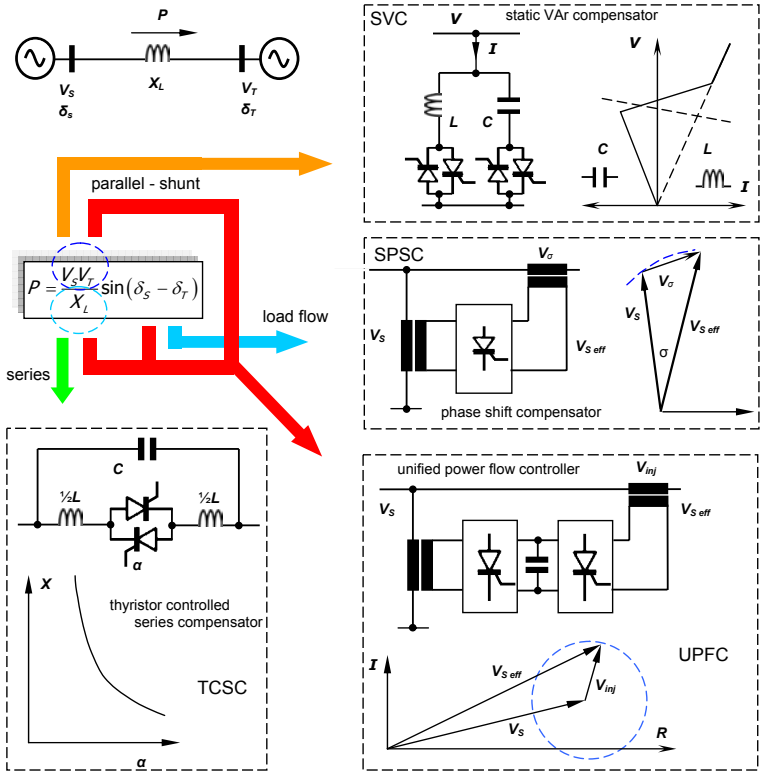


Figure 20.6. *The influence of FACTS devices on the power transfer equation.*

Compensators make use of capacitors, inductors and/or power electronic devices, and offer a higher transmission flexibility. Ideal compensators are lossless since their terminal voltage V_T and current I are in quadrature. Shunt compensators tend to have minimal effect on the fault short circuit current level, $I_{s/c}$.

Figure 20.5 shows the basic transmission line connection of each type of static VAr compensator, and summaries their main characteristics. Their influence on the power transferred is shown in figure 20.6, which is based on equation (20.6) The relative terminating and sending angle difference is arranged into absolute terms as follows

$$P = \frac{V_S V_T}{X_L} \sin\delta = \frac{V_S V_T}{X_L} \sin(\delta_S - \delta_T) \qquad (20.25)$$

20.7 Static Shunt Reactive Power Compensation

The objective of shunt compensation is to supply reactive power so as to increase the transmittable power by reducing the line voltage under light load conditions and increasing it under higher load conditions. The ideal compensator is lossless. It is located at the transmission line reactance midpoint and maintains the midpoint voltage such that $|V_S| = |V_T| = |V_M|$. Characteristically the generated and absorbed reactive powers are increased.

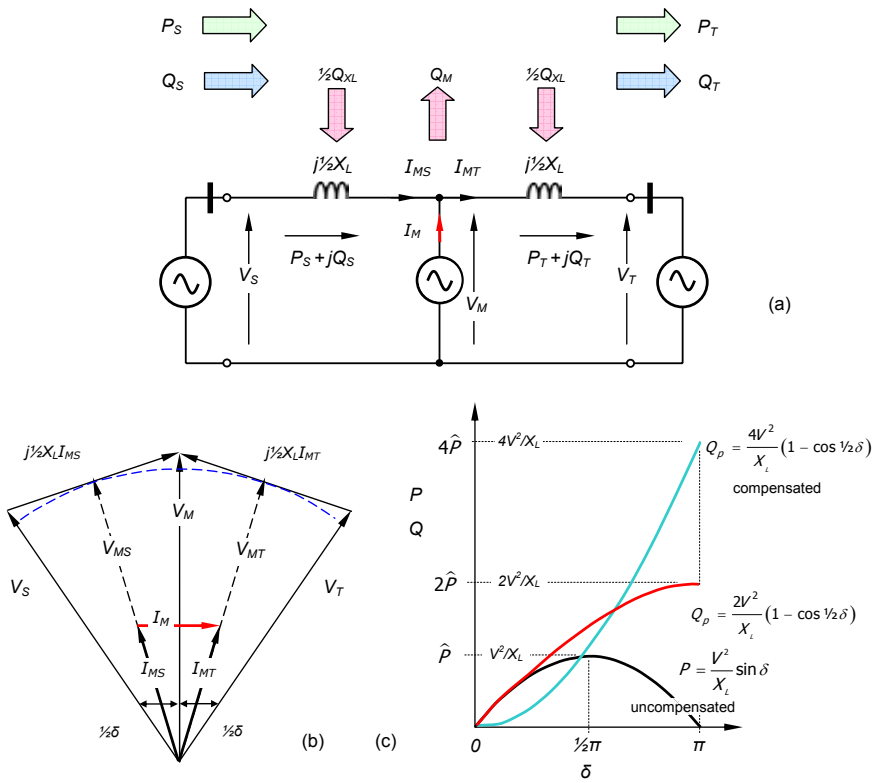


Figure 20.7. *Midpoint shunt compensation: (a) two source power system model; (b) phasor diagram for $|V_S| = |V_T| = |V_M| = V$; and (c) power versus load angle.*

Principle of shunt compensation

Ideally the sending, receiving, and midpoint voltage magnitudes are equal as shown in the compensated phasor diagram in figure 20.7b. The transmission line is then analysed as two independent halves. From this phasor diagram, with $V = V_M$, the two newly formed midpoint current and voltage magnitudes are

$$V_{MS} = V_{MT} = V\cos ¼\delta \tag{20.26}$$

$$I_{MS} = I_{MT} = I = \frac{4V}{X_L}\sin ¼\delta \tag{20.27}$$

With compensation, the transmitted active power is

$$\begin{aligned} P_p &= V_{MS} I_{MS} = V_{MT} I_{MT} = V_M I\cos ¼\delta = V\, I\cos ¼\delta \\ &= \frac{4V^2}{X_L}\sin ¼\delta \times \cos ¼\delta = \frac{2V^2}{X_L}\sin ½\delta \end{aligned} \tag{20.28}$$

The reactive power Q_s generated at the sending end and absorbed by the terminal end, Q_T is

$$Q_s = -Q_T = VI\sin ¼\delta = \frac{4V^2}{X_L}\sin^2 ¼\delta = \frac{2V^2}{X_L}(1 - \cos ½\delta) \tag{20.29}$$

The reactive power Q_p provided by the shunt compensator is

$$Q_p = |Q_S| + |Q_T| = 2VI\sin ¼\delta = \frac{8V^2}{X_L}\sin^2 ¼\delta = \frac{4V^2}{X_L}(1 - \cos ½\delta) \tag{20.30}$$

As shown in figure 20.7c, after compensation the maximum transmittable power is doubled, when δ=180º (which represent 90º across each half of the line) but at the expense of greatly increased VAr requirements, as seen in equations (20.29) and (20.30).

Shunt static VAr compensators

20.7.1 - Thyristor controlled reactor TCR

The basic phase angle controlled TCR is shown in figure 20.8b. If the thyristors are used purely as on/off switches with integral cycle control (without phase angle control) then the inductive arrangement is termed a thyristor switch reactor (TSR). Both modes are inductive thus are always associated with reactive power absorption.

Principle of TCR operation
The back-to-back connected thyristors conduct symmetrically on alternate half cycles of the ac supply.

$$\begin{aligned} i &= \frac{\sqrt{2}V}{\omega L_{sh}}(\cos\alpha + \cos\omega t) && \alpha < \omega t < \alpha + \sigma \\ &= 0 && \alpha + \sigma < \omega t < \alpha + \pi \end{aligned}$$

where $\sigma = 2(\pi - \alpha)$.

The operation of this configuration has been treated extensively in Chapter 13.1.1ii, where it was shown that continuous conduction occurs at a delay angle of 90º and partial symmetrical decreasing current (decreasing inductive VAr's) results for delay angles increasing from 90º to 180º, as shown in figure 20.8c. As the delay angle increases the fundamental current component decreases from a maximum, with the introduction of harmonics.

The power factor of the fundamental component lags by 90º, always absorbing reactive power. The odd order rms harmonics shown in figure 20.8d vary with delay angle according to

$$I_n = \frac{4}{\pi}\frac{V}{X_{Lsh}}\left[\frac{\sin(n+1)\alpha}{2(n+1)} + \frac{\sin(n-1)\alpha}{2(n-1)} - \frac{\sin n\alpha}{n}\cos\alpha\right] \quad \text{for} \quad n = 3, 5, 7... \tag{20.31}$$

and the 90º lagging fundamental rms is given by

$$\begin{aligned} I_1 &= \frac{2}{\pi}\int_{-(\pi-\alpha)}^{\pi-\alpha}\frac{V}{\omega L_{sh}}(\cos\alpha + \cos\omega t)\cos\omega t\; d\omega t \\ &= \frac{2}{\pi}\frac{V}{X_{Lsh}}\left[½\sin 2\alpha + \pi - \alpha\right] = B_L(\alpha)V \qquad \text{where} \quad X_{shL} = \omega L_{sh} \end{aligned} \tag{20.32}$$

for ½π ≤ α ≤ π with respect to zero voltage cross-over.

If the delay angles of both thyristors are not equal, even harmonics are produced, including a dc component. The total harmonic distortion is increased.
As the delay angle increases the current conduction angle σ decreases and the current decreases, as if the inductance were increasing, so that the TCR effective acts like controllable shunt susceptance.

$$L_{eff} = \frac{V}{\omega I_1} \qquad Q_1 = VI_1 = \frac{V^2}{\omega L_{eff}} \tag{20.33}$$

As the delay angle α increases and the current decreases, the thyristor and inductor conduction losses decrease. The maximum fundamental rms current component of $V/\omega L_{sh}$ occurs at α=½π.

If the three-phase TCR is configured in a delta arrangement, then the third harmonic current does not appear in the source line voltage. If two separate reactors are used in each phase as in figure 20.8a, then conduction up to 360º is possible resulting in the maximum possible fundamental with lower total harmonics, although energy cycling between the two inductors occurs. Alternatively, if transformer coupling is used, then the 5th and 7th order current harmonics can be eliminated if two three-phase delta connected TCR are used with a 12 pulse star-delta transformer secondary arrangement. (Alternatively, two discrete transformers can be used.) The use of a transformer means that voltage levels can be matched (usually a voltage step-down transformer in HVAC systems so that series semiconductor thyristor device connect is avoided). High leakage inductance minimizes the necessary TCR discrete inductance required. Two discrete TCR's also offers redundancy possibilities. Further, a transformer offers the possibility of reducing the inrush current if used in conjunction with the capacitive TSC.

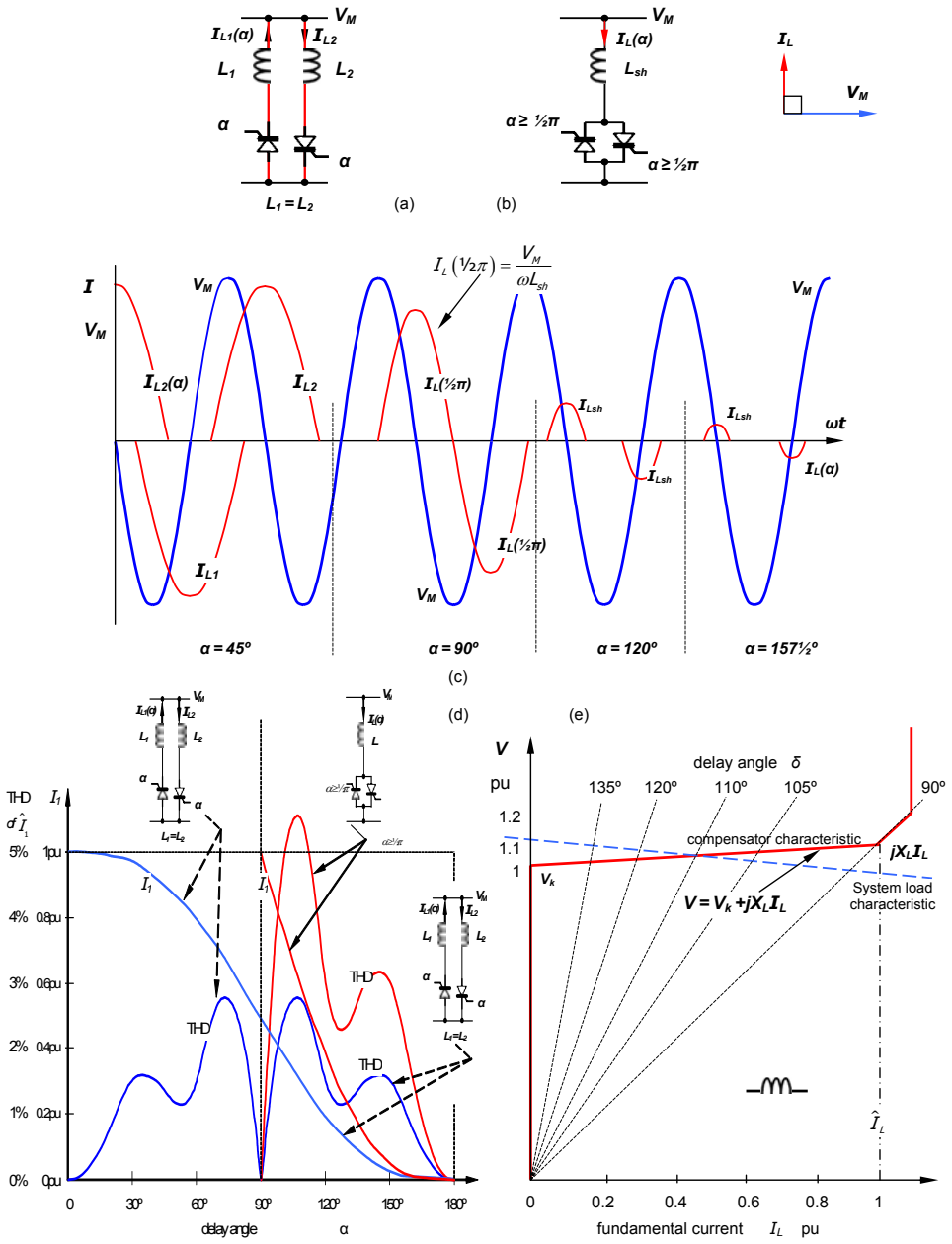


Figure 20.8. *TCR compensation: (a) dual reactor TCR compensator; (b) single reactor TCR compensator; (c) line voltage and current waveforms for delay angles α=45º, 90º, 120º and 157½º; (d) harmonics (delta connected - no triplens); and (e) fundamental I-V TCR characteristics.*

20.7.2 - Thyristor switched capacitor TSC

The basic shunt phase angle controlled TSC is shown in figure 20.9a, where the thyristors are usually operated either continuously conducting or off. Normally capacitor banks are switched in parallel to give line susceptance discrete level adjustment, since phase angle control is not possible because of the uncontrolled capacitive turn-on currents that would result. Beneficially, no harmonics are produced with continuous thyristor conduction. Transformer coupling can be used for voltage matching, the leakage of which helps control the initial current inrush. The capacitive VAr produced is determined by the capacitive current and the resultant system midpoint voltage, V_M:

$$I = \frac{V_M}{X_c} = \omega C V_M = B_c V_M \qquad (20.34)$$

$$VAr = Q_p = V_M I = -\omega C V_M^2 = -B_c V_M^2$$

At thyristor turn-on, α, with inevitable series inductance, the current into the formed LC circuit is given by

$$i(t) = \frac{\hat{V}}{X_c}\frac{n^2}{n^2-1}\cos(\omega t + \alpha) - \frac{n}{X_c}\left[V_{Co} - \frac{n^2}{n^2-1}\hat{V}\sin\alpha\right]\sin n\omega t - \frac{\hat{V}}{X_c}\cos\alpha\cos n\omega t \qquad (20.35)$$

where $n = \frac{\omega_0}{\omega} = \sqrt{\frac{X_c}{X_{Lsh}}}$ and ω is the supply frequency and $\omega_o = 1/\sqrt{LC}$.

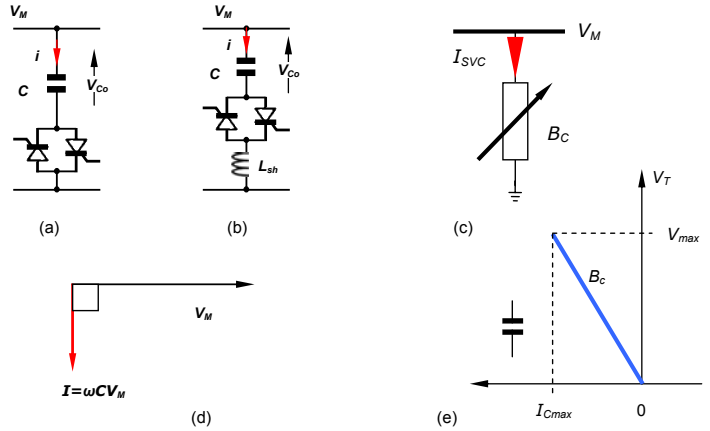


Figure 20.9. *Thyristor switched capacitor compensation: (a) ideal capacitor TSC compensator; (b) capacitor TSC compensator with line/leakage inductance; (c)variable susceptance representation; (d) I-V TSC phasor characteristics; and (d) I-V TSC susceptance characteristics.*

Equation (20.35) can be used to determine the necessary phase angle condition for transient free switch-in of the capacitors.
The oscillatory components, the second and third terms in equation (20.35), are zero when

$$\cos\alpha = 0, \text{ hence } \sin\alpha = \pm 1 \qquad (20.36)$$

$$\text{then } V_{Co} = \mp\frac{n^2}{n^2-1}\hat{V} \qquad (20.37)$$

The first condition implies thyristor turn-on at either ac peaks or troughs. The second condition implies that the capacitors be pre-charged, then the start up current is given by (first term in equation (20.35))

$$i(t) = \frac{\hat{V}}{X_c}\frac{n^2}{n^2-1}\sin\omega t \qquad (20.38)$$

Such initial conditions are usual impractical, and a turn-on angle compromise is used which results in acceptable oscillatory transient currents.

For capacitor disconnection, when the anode current reaches zero, the thyristors are no longer triggered, the system reactive energy changes abruptly and each capacitor retains a voltage

$$V_{Co} = \pm\hat{V}\frac{n^2}{n^2-1} \qquad (20.39)$$

20.7.3 - Shunt Static VAr compensator SVC (TCR//TSC)

A static VAr compensator is comprised of a thyristor controlled reactor compensator and a thyristor switched capacitor compensator as shown in figure 20.10a. The leading reactive power is provided in discrete equal steps (or 2^n steps) by banks of thyristor switched capacitor compensators (TSC) and precise continuous VAr adjustment is affected by a thyristor controlled reactor compensator (TCR). The maximum lagging current from the TCR is equal to the incremental capacitive leading current, such that the two can cancel to zero giving zero net reactive VA. As the phase angle of the TCR is increased, the net leading VAr increases. At zero TCR conduction, a capacitive bank is decremented and the TCR starts with full conduction, that is zero delay angle.
Ideally, no active power is drawn from the system and the reactive power depends on the net fundamental impedance of the parallel capacitor-reactance combination, which is TCR delay angle dependent.

$$P_{SVC} = 0$$
$$Q_{SVC} = -\frac{V_M^2}{X_{SVC}} = -V_M^2 B_{SVC} \qquad (20.40)$$

The SVC is usually transformer coupled for voltage matching of the thyristors. The compensator bus usually incorporates permanent LC notch filters to minimise the injection of 5th and 7th order harmonics, produced by the TCR, back into the HV system.
An advanced SVC that uses a voltage source inverter is called at static compensator, or STATCOM.

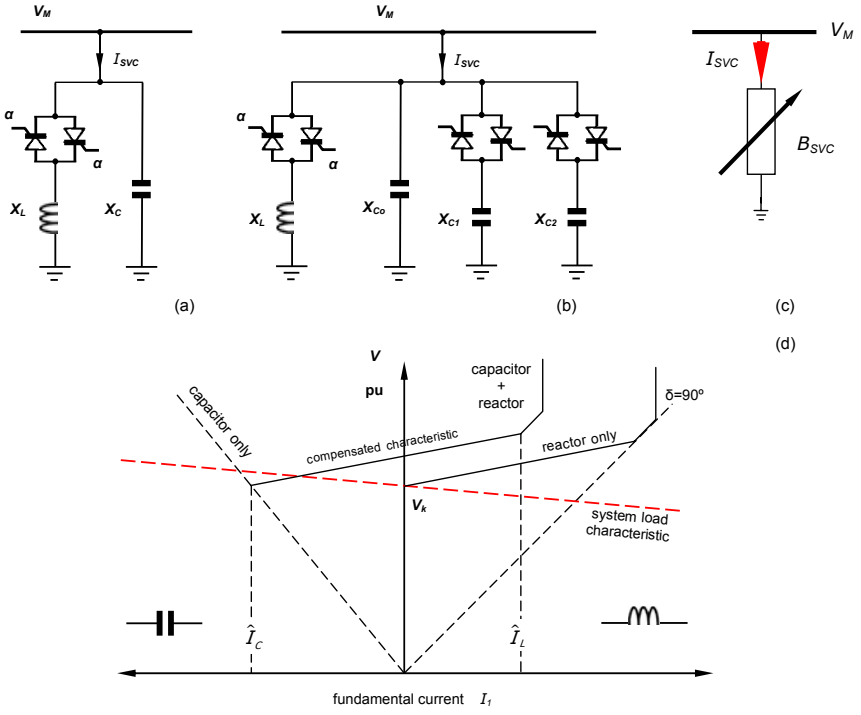


Figure 20.10. *Static VAr compensator (SVC): (a) basic SVC; (b) SVC with capacitor banks; (c) variable susceptance model representation; and (d) I-V SVC characteristics.*

Example 20.2: ***Shunt thyristor controlled reactor specification***

A 50Hz 400V ac transmission line has line reactance of $X_L = 2.2\ \Omega$ and is delivering 100kW. Calculate

i. the load angle δ
ii. the line current
iii. the TCR and line reactive powers
iv. the TCR current and reactance and inductance at this current (with $V_M = 400V$)
v. the 50Hz reactance, thence inductance if the maximum TCR current is 100A (1.0 pu)
vi. the TCR triggering delay angle, hence thyristor conduction period, if the reactor current is 0.5pu
vii. the effective reactance, inductance and the reactive power at a TCR current of 0.5pu
viii. delay angle α for $V_M = 400V$, if the TCR inductance is 5mH

Solution

i. Rearrangement of equation (20.28) gives the transmission load angle

$$\delta = 2\sin^{-1}\left(\frac{P_p X_L}{2V^2}\right) = 2\sin^{-1}\left(\frac{100\text{kW}\times 2.2\Omega}{2\times 400\text{V}^2}\right) = 86.86°$$

ii. Equation (20.27) gives the transmission line current

$$I = \frac{4V}{X_L}\sin ¼\delta = \frac{4\times 400V}{2.2\Omega}\sin(¼\times 88.7°) = 269\text{A}$$

iii. The reactive power given by equation (20.30) is

$$Q_p = |Q_S| + |Q_T| = \frac{4V^2}{X_L}(1-\cos ½\delta) = \frac{4\times 400V^2}{2.2\Omega}(1-\cos ½\times 88.7°) = 79,647\text{VAr}$$
$$Q_S = -Q_T = ½Q_p = 39,823\text{VAr}$$

iv. The TCR current is

$$I_{Q_p} = \frac{Q_p}{V} = \frac{82.445\times 10^3\,\text{VAr}}{400\text{V}} = 199\text{A}$$
$$X_p = \frac{V}{I_{Q_p}} = \frac{400\text{V}}{207.2\text{A}} = 2\Omega \quad\Rightarrow\quad L = \frac{X_p}{2\pi f} = \frac{1.93\Omega}{2\pi 50\text{Hz}} = 6.37\text{mH}$$

v. At 100A the TCR inductance at the fundamental frequency, 50Hz, is

$$X_p = \frac{V}{I_Q} = \frac{400\text{V}}{100\text{A}} = 4.0\Omega \quad\text{thence}\quad L_p = \frac{X_p}{2\pi f} = \frac{4.0\Omega}{2\pi\times 50\text{Hz}} = 12.7\text{mH}$$

vi. Solving equation (20.32) gives the transmission angle for a TCR current of 50A

$$I_1 = \frac{2}{\pi}\frac{V}{X_L}\left[½\sin 2\alpha + \pi - \alpha\right]$$
$$50A = \frac{2}{\pi}\frac{400V}{4.0\Omega}\left[½\sin 2\alpha + \pi - \alpha\right]$$
$$0 = ½\sin 2\alpha - \alpha + ¾\pi \quad\Rightarrow\quad \alpha = 113.8°$$

vii. The effective inductance and reactive power are given by equation (20.33)

$$L_{eff} = \frac{V}{2\pi f\, I_1} = \frac{400\text{V}}{2\pi\times 50\text{Hz}\times 50\text{A}} = 25.5\text{mH}$$
$$Q_1 = VI_1 = \frac{V^2}{\omega L_{eff}} = 400\text{V}\times 50\text{A} = 20\times 10^3\ \text{VAr}$$

vii. From part iv, the effective inductance for 400V is 6.14mH, or 1.93Ω with 207.2A at the fundamental frequency.

$$L = L_{eff}\frac{2}{\pi}\left[½\sin 2\alpha + \pi - \alpha\right]$$
$$5\text{mH} = 6.37\text{mH}\times\frac{2}{\pi}\left[½\sin 2\alpha + \pi - \alpha\right] \quad\Rightarrow\quad \alpha = 99.8°$$

♣

20.8 Static Series Reactive Power Compensation

Transmission line capability can be increased by installing series compensation in order to reduce the transmission line net series reactance. Effectively the apparent transmission line length is varied. The insertion of inductance decreases transmission capability, but may be used to limit fault levels or to divert power flow, while the insertion of series capacitance acts to cancel the series inductive voltage drop, reducing the net line impedance, thus:

- increases power flow capability and stability margins;
- reduces the transmission load angle;
- increases the virtual load; and
- provides a means of damping power oscillations.

Normally, series compensation is capacitive. Since distributed compensation along the line is impractical, as with shunt compensation, series compensation is normally inserted at the reactance midpoint. Series compensation is normally only used on very long ac transmission lines, thereby making long distance ac transmission viable.

Principle of series compensation

The ideal series compensator is effectively pure reactance, without any power loss. The ideal series line compensation of a transmission line is shown in figure 20.11a, where the compensator voltage is at quadrature to the line current. The line resistance is neglected. The effective line reactance is given by

$$X_{eq} = X_L - X_{sc} = X_L(1-k) \tag{20.41}$$

where $k = X_{sc}/X_L$ is the degree of series compensation. If the compensation is inductive the reactance is negative and k is negative ($k<0$), while k is positive for capacitive compensation ($k>0$). Assuming $V_S = V_T = V$, then from equation (20.11), the line current, midpoint voltage, transmitted power and reactive power are

$$\begin{aligned}
\mathbf{I} &= \frac{2V}{X_L(1-k)}\sin\tfrac{1}{2}\delta \quad \angle 0 \\
\mathbf{V_m} &= V\cos\tfrac{1}{2}\delta \quad \angle 0 \\
P_{sc} &= \frac{V^2}{X_L(1-k)}\sin\delta \\
Q_{sc} &= \frac{2V^2}{X_L}\times\frac{k}{(1-k)^2}(1-\cos\delta)
\end{aligned} \tag{20.42}$$

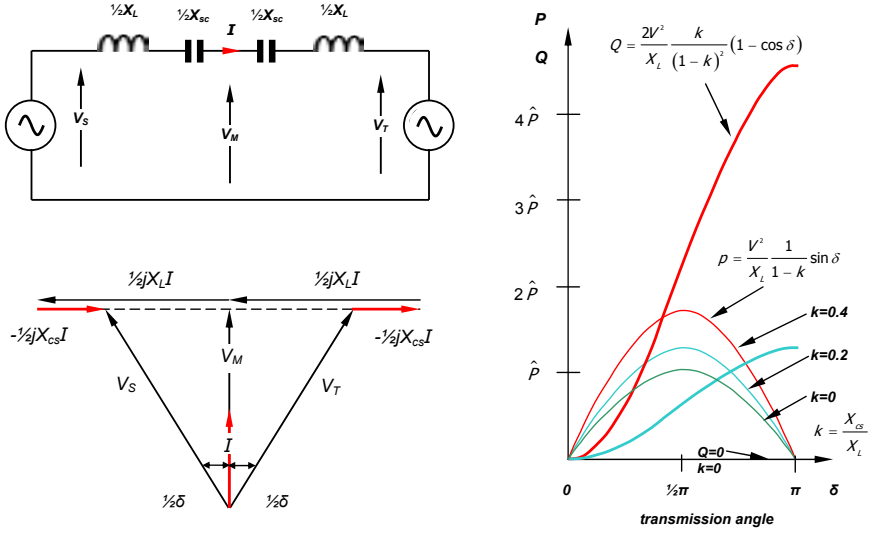


Figure 20.11. *Midpoint static series compensation: (a) two source power system model; (b) phasor diagram for $|V_S| = |V_T| = V$; and (c) power versus load angle.*

The voltage across the compensation element, as shown in the phasor diagram in figure 20.11b, at quadrature to the line current, is

$$V_{sc} = IX_{sc} \tag{20.43}$$

The power equations are shown plotted in figure 20.11c. This figure shows that increased capacitance ($k>0$), increases the transmittable power and the reactive power. Maximum power is transmitted with a load angle of $\delta = \frac{1}{2}\pi$, when $k \to 1$, that is when $\omega L_L = 1/\omega C_{sc}$ (line resonance at frequency ω).

Series static VAr compensators

20.8.1 - Thyristor switched series capacitor TSSC

A thyristor switched series capacitor compensator TSSC consist of a least one series capacitor, each shunted by a back-to-back pair of anti-parallel connected phase control thyristors, as shown in figure 20.12a. The thyristors when continuously triggered, provide a path for the line current to by-pass the series compensating capacitors. The thyristor are taken out of circuit when the gate triggering is removed and natural turn-off commutation occurs at the subsequently line current reversal, that is, the thyristors are line or naturally commutated. With this commutation process, the series capacitor charges with a dc bias as shown in figure 20.12b. Subsequent thyristor turn-on should only occur at the line zero current points in order to avoid high initial anode *di/dt* currents.

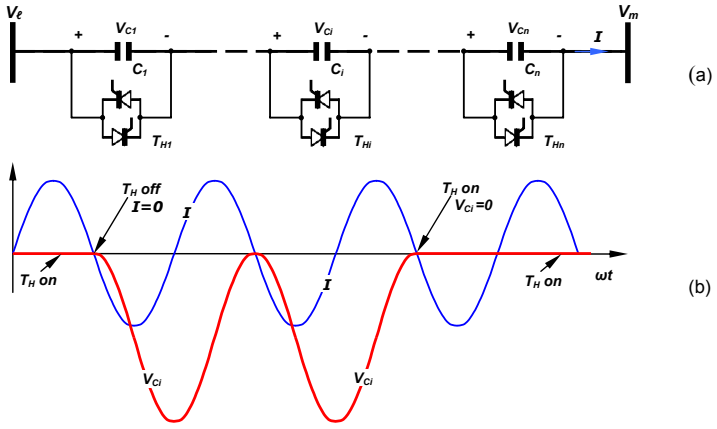


Figure 20.12. *Thyristor switched series capacitor compensation TSSC: (a) series connected capacitors and (b) zero current activation and zero voltage deactivation.*

20.8.2 - Thyristor controlled series capacitor TCSC

Better capacitor series voltage control is obtained if the thyristors in figure 20.12a are self-commutatable, such as with symmetrical voltage blocking IGCThyristors. This series TCSC compensator is the dual to the shunt TSR in figure 20.8b.

- Instead of thyristors in series with inductance, thyristors are in parallel with capacitance.
- Instead of uni-directional voltage blocking, naturally commutating switches, the capacitive series compensator uses bidirectional voltage blocking, self-commutatable switches.
- In the series compensator, compensation occurs when the series thyristors are on, while compensation is active in the series compensator case when the parallel IGC Thyristors are off.
- The shunt compensator supports a sinusoidal voltage and produces current harmonics, while the series compensator conducts the sinusoidal line current and produces voltage harmonics.

Typical series TCSC waveforms are shown in figure 20.13c, while the harmonics produced are shown in figure 20.8d. It will be noted that the same equations as in section 20.2.4i for the TCR hold, except that voltages and currents are interchanged, and capacitive reactance is used instead of inductive reactance. Specifically, if the line current is

$$i = I_M \sin\omega t = \sqrt{2}I\sin\omega t \tag{20.44}$$

Then the capacitor voltage is given by

$$v_c(t) = I_M X_C(\cos\alpha + \cos\omega t) = \frac{I_M}{\omega C}(\cos\alpha + \cos\omega t) \qquad (20.45)$$

The power factor of the fundamental voltage component lags I by 90º, always producing reactive power.

The odd order rms (total) harmonics shown in figure 20.8d vary with delay angle according to

$$V_n = \frac{4}{\pi} I X_C \left[\frac{\sin(n+1)\alpha}{2(n+1)} + \frac{\sin(n-1)\alpha}{2(n-1)} - \frac{\sin n\alpha}{n}\cos\alpha\right] \quad \text{for} \quad n = 3, 5, 7... \qquad (20.46)$$

and the 90º lagging fundamental rms voltage is given by

$$V_1 = \frac{2}{\pi}\frac{I}{\omega C}\left[\tfrac{1}{2}\sin 2\alpha + \pi - \alpha\right] = \frac{2}{\pi} I X_C\left[\tfrac{1}{2}\sin 2\alpha + \pi - \alpha\right] \quad \text{where} \quad X_C = 1/\omega C \qquad (20.47)$$

for ½π ≤ α ≤ π with respect to zero current cross over.
If the delay angles of both thyristors are not equal, even voltage harmonics are produced, including a dc voltage component. The total harmonic distortion is increased.

As the delay angle increases the voltage period angle σ decreases and the voltage decreases, as if the capacitance were increasing, so that the series TCSC effective acts like controllable capacitive susceptance.

$$C_{eff} = \frac{I}{\omega V_1}$$
$$Q_1 = V_1 I = \frac{I^2}{\omega C_{eff}} = I^2 X_{eff} \qquad (20.48)$$

Also, as the delay angle α increases and the voltage decreases, thyristor conduction increases, hence thyristor losses increase.

As with the shunt TCR, operation below 90º is possible if two capacitors are used as shown in figure 20.13b. Extra semiconductors (diodes) are needed, but the IGC Thyristors only need forward voltage blocking properties. Consequently, capacitors with uni-directional voltage properties can be used. The voltage harmonics are lower but at the expense of extra devices and losses.

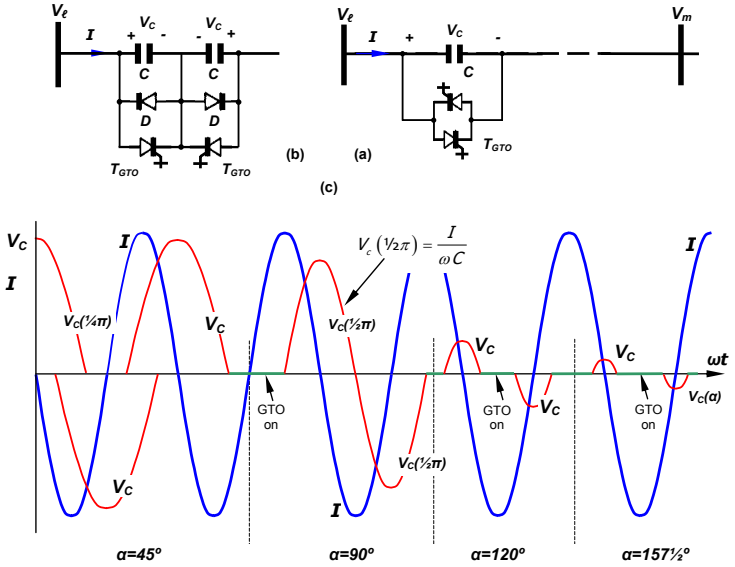


Figure 20.13. *Thyristor controlled series capacitor compensation TCSC: (a) series connected capacitors with shunt self-commutable GTOs for α≥90º; (b) for α≥0º; and (c) line current and current waveforms for delay angles α=45º, 90º, 120º and 157½º.*

20.8.3 - Series Static VAr compensator SVC (TCR//C)-TCSC

The TCR//C consists of a line series compensating capacitor in parallel with a thyristor controlled reactor (TCR), as shown in figure 20.14. By varying the delay angle of the TCR thyristors, the capacitive reactance can be decreased, since the fundamental reactance of the parallel combination is given by

$$X_{eff}(\alpha) = \frac{X_C X_{L1}(\alpha)}{X_C - X_{L1}(\alpha)} \qquad (20.49)$$

where, from equation (20.32), the reactance at the fundamental frequency is

$$X_{L1}(\alpha) = \frac{\tfrac{1}{2}\pi}{\tfrac{1}{2}\sin 2\alpha + \pi - \alpha} X_L \qquad \text{where} \quad X_L = \omega L \qquad (20.50)$$

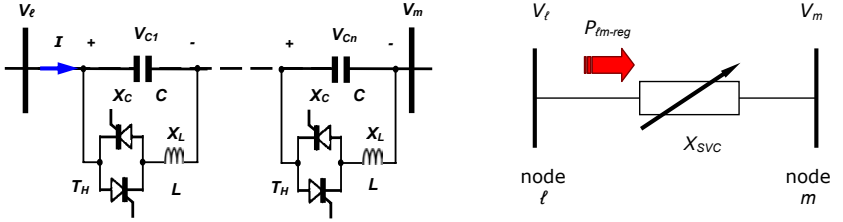


Figure 20.14. *Thyristor controlled reactance and series connected capacitance, SVC compensation and variable reactance model representation.*

The voltage harmonics produced by the reactor tend to be trapped in the parallel connected capacitor due to its the low capacitive reactance X_C which is inversely proportion to harmonic frequency (relative to line reactance X_s which increases proportional to harmonic frequency).
Accounting for the line reactance X_s and compensator fundamental reactance X_{eff}, the active and sending reactive powers are given by equations (20.6) and (20.8), that is

$$P_T = \frac{V_S V_T}{X_L + X_{eff}}\sin(\delta_S - \delta_T)$$
$$Q_S = V_S \times \frac{V_S - V_T\cos(\delta_S - \delta_T)}{X_L + X_{eff}} \qquad (20.51)$$

The signs in these equations are appropriately changed for capacitive operation.

The capacitor and inductor voltages and currents can be define during the period when the thyristors block and when a thyristor conducts. If the rms line current is I_M then

- when the thyristors block:

$$V_c(t,\alpha) = \sqrt{2} I_M X_C\left[\sin\alpha\left[1 - \sin(\omega t - \alpha)\right] - \cos\alpha\cos(\omega t + \alpha)\right] + V_{C\,t=\alpha+\sigma}$$
$$I_C(t,\alpha) = \sqrt{2} I_M \sin\omega t \quad (= \text{line current}) \qquad (20.52)$$
$$I_{Th} = I_L = 0 \quad \text{and} \quad V_L(t,\alpha) = 0$$

- when a thyristor conducts:

$$V_L(t,\alpha) = V_c(t,\alpha) = \sqrt{2} I_M X_L \frac{\omega k^2}{1-k^2}\cos\alpha\left[\begin{array}{l}\left[\cos(\omega t - \alpha) - k\sin\left(\omega_o t - k(\alpha - \tfrac{1}{2}\pi)\right)\right] \\ \quad -\left[\sin(\omega t - \alpha) - \cos\left(\omega_o t - k(\alpha - \tfrac{1}{2}\pi)\right)\right]\end{array}\right] + V_{C\,t=\alpha}\cos\left(\omega_o t - k(\alpha - \tfrac{1}{2}\pi)\right) \qquad (20.53)$$

$$I_L(t,\alpha) = \sqrt{2} I_M \frac{k}{1-k^2}\left[\begin{array}{l}\sin\alpha\left[\sin\left(\omega_o t - (\alpha - \tfrac{1}{2}\pi)/k\right) - k\cos(\omega t - \alpha)\right] \\ \quad -\cos\alpha\left[\cos\left(\omega_o t - (\alpha - \tfrac{1}{2}\pi)/k\right) + \sin(\omega t - \alpha)\right]\end{array}\right] \qquad (20.54)$$
$$I_C(t,\alpha) = I_L(t,\alpha) + \sqrt{2} I_M \sin\omega t$$

where $\omega_o = \frac{1}{\sqrt{LC}} = k\omega = 2\pi f$, that is, $k = \frac{\omega_o}{\omega}$

Example 20.3: ***Series thyristor controlled reactor specification – integral control***

A 50Hz 230kV three-phase ac transmission line has line reactance of X_L = 52 Ω per phase and a maximum thermally limited line current of 2000A. The line voltage can vary by ±5% at each end and the load angle between the ends varies between 5° and 10°, where the load is lagging.
Series TCSC SVC connected at the midpoint, comprised of four compensating three-phase modules has a capacitive reactance of 10Ω with 1.66Ω of switchable parallel inductance.
Calculate

i. the nominal power
ii. the line current and powers under worse case conditions, before series compensation
iii. the effective module reactance when the impedances are parallel connected
iv. the effective line impedance at worse case, if 50% of rated power is the transmission objective, and the resultant transmission powers

Solution

i. The nominal maximum power is given by

$$P = \sqrt{3} V_L I_L$$
$$= \sqrt{3} \times 230\text{V} \times 2000\text{A} = 796\text{MW}$$

ii. Worse case power delivery conditions are when the sending end is 5% below the nominal ac voltage, while the receiving end is 5% above the nominal, at the highest load angle, δ=10°.
The current can be evaluated by equating equation (20.1) with equations (20.6) and (20.7)

$$I = I\cos\phi + jI\sin\phi = I \times e^{j\phi}$$

$$I\sin\phi = \frac{V_S\cos\delta - V_T}{X_L} = \frac{Q_T}{V_T} \qquad I\cos\phi = \frac{V_S\sin\delta}{X_L} = \frac{P}{V_T}$$
$$= \frac{218.5\text{kV}\times\cos 10° - 241.5\text{kV}}{52\Omega} \qquad = \frac{218.5\text{kV}\times\sin 10°}{52\Omega}$$
$$= -0.5051\text{kA} \qquad = 0.730\text{kA}$$

$$I = -505.1 + \text{j}730.0 = 887.7^{-j55.3°}\text{ pu wrt } V_T$$

The real power flow is:

$$P = \frac{V_T \times V_R}{X_L}\sin\delta$$
$$= \frac{218.5\text{kV}\times 241.5\text{kV}}{52\Omega}\sin 10°$$
$$= 176.2\text{MW}$$

The sending end reactive power is

$$Q_s = V_S \times \frac{V_S - V_T\cos\delta}{X_L}$$
$$= 218.5\text{kV} \times \frac{218.5\text{kV} - 241.5\text{kV}\times\cos 10°}{52\Omega} = -81.2\text{MVAr}$$

The terminal end reactive power is

$$Q_T = V_T\frac{V_S\cos\delta - V_T}{X_L}$$
$$= 241.5\text{kV} \times \frac{218.5\text{kV}\times\cos 10° - 241.5\text{kV}}{52\Omega} = -122.2\text{MVAr}$$

The line reactive power is given by Q_s - Q_T = -81.2MVAr + 122.2 MVAr = 41.0MVAr.
Alternatively, the line reactive power can be calculated from $I^2X_L = 887.7^2 \times 52 = 41.0\text{MVAr}$

iii. The inductor j1.66Ω in parallel with the capacitor –j10Ω give a parallel combination impedance of

$$X_{cell} = \frac{-jX_{cap}\times jX_{ind}}{-jX_{cap} + jX_{ind}}$$
$$= \frac{-j10\Omega \times j1.66\Omega}{-j10\Omega + j1.66\Omega} = j2\Omega$$

iv. Worse case power delivery conditions are when the sending end is 5% below the nominal ac voltage, while the receiving end is 5% above the nominal, at the highest load angle, *δ*. That is, the

necessary line reactance, for half rated power, is given by

$$P = \frac{V_T \times V_R}{X_L}\sin\delta$$
$$= \frac{95\%V_{Nom}\times 105\%V_{Nom}}{X_L}\sin\delta$$
$$50\%\text{ of }796\text{MW} = \frac{95\%230\text{kV}\times 105\%230\text{kV}}{X_L}\sin 10°$$
$$\Rightarrow X_L = 23\Omega$$

Figure 20.15 shows that with only one module activated, the line reactance can be compensated to 24Ω. The real power flow and reactive powers are:

$$P = \frac{V_T \times V_R}{X_L}\sin\delta$$
$$= \frac{218.5\text{kV}\times 241.5\text{kV}}{24\Omega}\sin 10°$$
$$= 381.8\text{MW}$$

The sending end reactive power is

$$Q_s = V_S \times \frac{V_S - V_T\cos\delta}{X_L}$$
$$= 218.5\text{kV} \times \frac{218.5\text{kV} - 241.5\text{kV}\times\cos 10°}{24\Omega} = -176\text{MVAr}$$

The terminal end reactive power is

$$Q_T = V_T\frac{V_S\cos\delta - V_T}{X_L}$$
$$= 241.5\text{kV} \times \frac{218.5\text{kV}\times\cos 10° - 241.5\text{kV}}{24\Omega} = -264.8\text{MVAr}$$

The line reactive power is given by Q_s - Q_T = -176MVAr + 264.8 MVAr = 88.8MVAr.

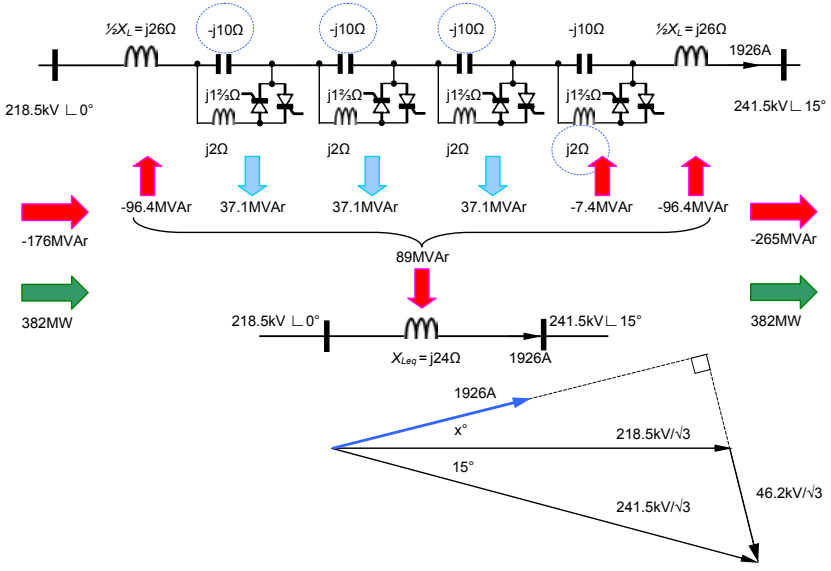


Figure 20.15. *Example 20.3.*

The current is

$$I = I\cos\phi + jI\sin\phi = I\times e^{j\phi}$$

$$I\sin\phi = \frac{V_S\cos\delta - V_T}{X_L} = \frac{Q_T}{V_T} \qquad I\cos\phi = \frac{V_S\sin\delta}{X_L} = \frac{P}{V_T}$$

$$= \frac{218.5\text{kV}\times\cos 10^\circ - 241.5\text{kV}}{24\Omega} \qquad = \frac{218.5\text{kV}\times\sin 10^\circ}{24\Omega}$$

$$= -1.10\text{kA} \qquad = 1.58\text{kA}$$

$$I = -1100 + \text{j}1581 = 1926^{-j55.2^\circ}\text{ pu wrt } V_T$$

The line reactive power can be calculated and confirmed from $I^2X_L = 1926^2\times 24\Omega = 89.0\text{MVAr}$

The sending power factor is

$$P = \sqrt{3}V_sI_L\cos\phi$$
$$= \sqrt{3}\times 218.5\text{kV}\times 1.926\text{kA}\times\cos 52.5^\circ$$
$$= 443.7\text{MW}$$

♣

Example 20.4: ***Series thyristor controlled reactor specification – Vernier control***

A 50Hz 400V ac transmission line has line reactance of X_L = 2.2 Ω and is delivering 100kW at a load angle of 80º. The TCSC comprising C=30μF and L=3.53mH is operated at a load angle of 80º. Calculate

i. the degree of compensation k
ii. the compensating capacitive reactance
iii. the line current I
iv. the reactive power Q
v. the TCSC delay angle if the effective capacitive reactance is 200Ω

Solution

i. From equation (20.42)

$$k = 1 - \frac{V^2}{X_LP_{sc}}\sin\delta = 1 - \frac{400^2}{2.2\Omega\times 100\text{kW}}\times\sin 80^\circ = 0.284$$

ii. From equation (20.41), the compensation reactance is

$$X_{sc} = kX_L = 0.284\times 2.2\Omega = 0.624\Omega$$

iii. From equation (20.42)

$$\mathbf{I} = \frac{2V}{X_L(1-k)}\sin\tfrac{1}{2}\delta = \frac{2\times 400\text{V}}{2.2\Omega\times(1-0.284)}\sin\tfrac{1}{2}80^\circ = 326.5\text{A}$$

iv. From equation (20.42)

$$Q_{sc} = \frac{2V^2}{X_L}\times\frac{k}{(1-k)^2}(1-\cos\delta) = \frac{2\times 400^2}{2.2\Omega}\times\frac{0.284}{(1-0.284)^2}\times(1-\cos 80^\circ) = 66,586\text{ VAr}$$

v. The compensator capacitive reactance is

$$X_c = \frac{1}{\omega C} = \frac{1}{2\pi f\,C} = \frac{1}{2\pi 50\text{Hz}\times 30\mu\text{F}} = 106.1\Omega$$

The compensator inductive reactance is

$$X_L = \omega L = 2\pi f\,L = 2\pi 50\text{Hz}\times 3.53\text{mH} = 1.11\Omega$$

From equations (20.49) and (20.50)

$$X_{eff}(\alpha) = \frac{X_CX_{L1}(\alpha)}{-X_C + X_{L1}(\alpha)}$$

$$200\Omega = \frac{106.1\Omega\times X_{L1}(\alpha)}{-106.1\Omega + X_{L1}(\alpha)} \quad\Rightarrow\quad X_{L1}(\alpha) = 32.0\Omega$$

Then

$$X_{L1}(\alpha) = \frac{1/2\pi}{1/2\sin 2\alpha + \pi - \alpha}X_L$$

$$32.0\Omega = \frac{1/2\pi}{1/2\sin 2\alpha + \pi - \alpha}\times 1.11\Omega \quad\Rightarrow\quad \alpha = 167^\circ$$

♣

20.8.4 Static series phase angle reactive power compensation/shift SPS

Phase compensation is a specific case of series compensation, as shown in figure 20.16, where the phase angle change is used to control the power flow. Where as series reactive control is usually located at the line reactance midpoint, phase angle compensation is performed at the sending end of the transmission line. The compensator is an ac voltage source V_ε of controllable magnitude and phase angle. The effecting sending end voltage $V_{S\,eff}$ becomes

$$V_{S\,eff} = V_S + V_\varepsilon \qquad (20.55)$$

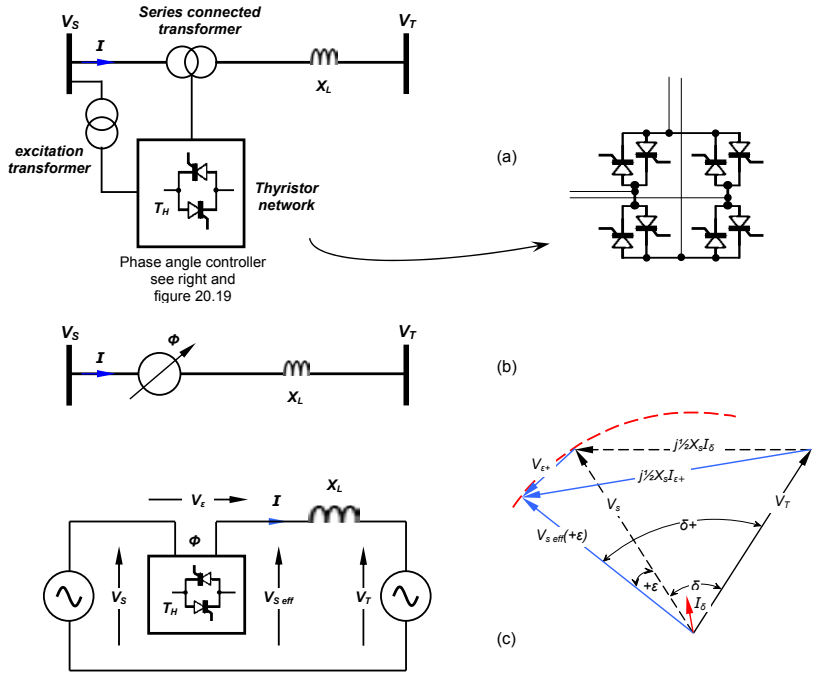


Figure 20.16. *Transformer series phase angle compensation: (a) series transformer with ac tap changing thyristor network; (b) variable phase angle representation; and (c) two port series phase angle compensator system.*

The compensator can function in one of two ways.

- The load angle is varied maintaining a voltage magnitude $V_{S\,eff}$ the same as the sending voltage V_S

$$|V_{S\,eff}| = |V_S| = V_{S\,eff} = V_S = V \qquad (20.56)$$

- The compensator phase angle is maintained at quadrature to the sending voltage

$$|V_{S\,eff}| = V_{S\,eff} = \sqrt{V_S^2 + V_\varepsilon^2} \qquad (20.57)$$

In both cases, power flow control is achieved at the expense of consuming reactive power from the network. The system transfer admittance has V_s replaced by $V_{s\,eff}$, that is

$$\begin{bmatrix} I_{s\,eff} \\ I_T \end{bmatrix} = \frac{1}{X_L}\begin{bmatrix} 1 & \cos\phi + j\sin\phi \\ -(\cos\phi - j\sin\phi) & 1 \end{bmatrix}\begin{bmatrix} V_{s\,eff} \\ V_T \end{bmatrix} \qquad (20.58)$$

Phase shifting (Φ<½π) with a constant voltage magnitude $|V_{s\,eff}| = |V_s|$

Figure 20.17a shows the resultant phase diagram for the effects of a series phase compensator when the regulator output voltage magnitude is the same as the sending voltage magnitude. Figure 20.17b shows that the effect of the phase shift is to shift the power transmission dependence on load angle. Since $|V_{s\,eff}| = |V_S| = |V_T| = V$, the active power transferred through the phase shifter is

$$P_{S-T}^{comp} = \frac{|V_s||V_T|}{X_L}\sin(\delta - \varepsilon) = \frac{V^2}{X_L}\sin(\delta - \varepsilon) \qquad (20.59)$$

By controlling the compensator angle ε, the output power can be controlled independent of the transmission load angle δ. The peak power can be shifted from a load angle δ = ½π to any desired load angle, by maintaining the phase shifter angle such that $\delta - \varepsilon$ = ½π is maintained. The transmitted reactive power is

$$Q_{comp} = \frac{2V^2}{X_L}(1 - \cos(\delta - \varepsilon)) \qquad (20.60)$$

Equations (20.59) and (20.60) show that

- The maximum power and VAr are unchanged, only the load angle at which they occur can be controlled.
- Unlike other series and shunt compensators, the phase compensator needs to handle both real power and VAr.

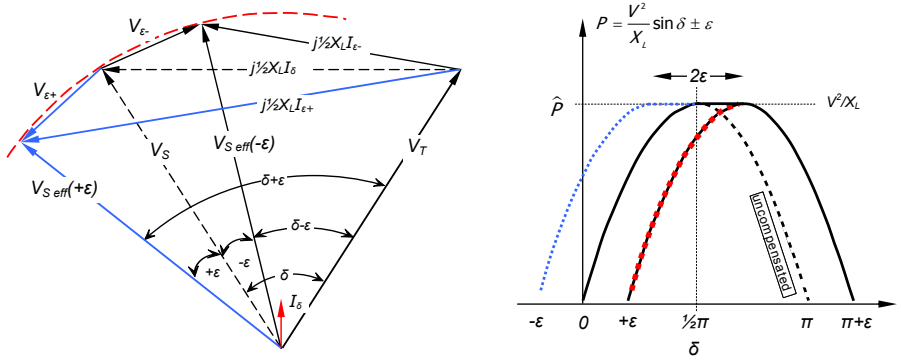
From the phasor diagram in figure 20.17a, the shift compensator terminal voltage and current are

$$V_\varepsilon = 2V\sin ½\varepsilon$$
$$I = \frac{2V}{X_L}\sin ½\delta \qquad (20.61)$$

The apparent power of the compensator is therefore

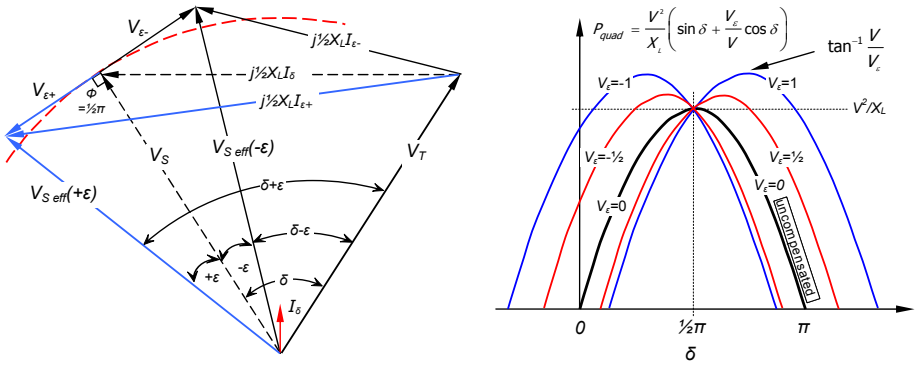
$$S_{comp} = V_\varepsilon I = \frac{4V^2}{X_L}\sin ½\varepsilon \sin ½\delta \qquad (20.62)$$

If the compensation angle is negative, by effectively reversing the terminals of the compensator, then maximum power can be attaining for load angles of less than ½π, as indicated by the dashed sine curve portion in figure 20.17b.



(a) (b)

Figure 20.17. *Transformer series phase angle compensation: (a) phasor diagram for phase shift ±ε and (b) transmission power versus load angle.*



(a) (b)

Figure 20.18. *Series phase angle compensation: (a) phasor diagram for phase shift ±ε giving quadrature boosting, Φ=½π and (b) transmission power versus load angle.*

Phase shifting with a quadrature phase shift voltage $\sqrt{V_S^2 + V_\varepsilon^2}$

Figure 20.18a shows the phasor diagram when the phase shift voltage V_ε is maintained at quadrature to the sending end voltage V_S, such that

$$|V_{S\,eff}| = V_{S\,eff} = \sqrt{V_S^2 + V_\varepsilon^2}$$

The power transmitted is show in figure 20.18b and is given by

$$P_{quad} = \frac{V^2}{X_L}\left(\sin\delta + \frac{V_\varepsilon}{V}\cos\delta\right) \qquad (20.63)$$

The effective transmission end voltage is increased, hence the possible transmitted power is increased. This increased power mode of phase shift compensation is called quadrature boost.

The Quadrature Boost phase shift compensator

The phase shifting configuration shown in figure 20.19a uses a delta connect line transformer to feed a naturally commutating thyristor tap changing circuit which transfers power to the line through the series line transformer. The transformer phase arrangement ensures that the series injected voltage is always at quadrature to the line voltage. Since the effective line voltage $V_{S\,eff}$ is now greater than the line sending voltage V_S, as shown in the phasor diagram in figure 20.19b, the converter forms a quadrature boost compensator. The tap changer is not reversible, that is, power can only flow from the shunt excitation transformer to the series compensating transformer in the transmission line.
The series compensator can be controlled in two ways

- Phase angle control of the thyristors in one bridge
- Fewer harmonics are generated by either switching in or out the different excitation windings shown in figure 20.19a, which gives 27 different output voltage possibilities.

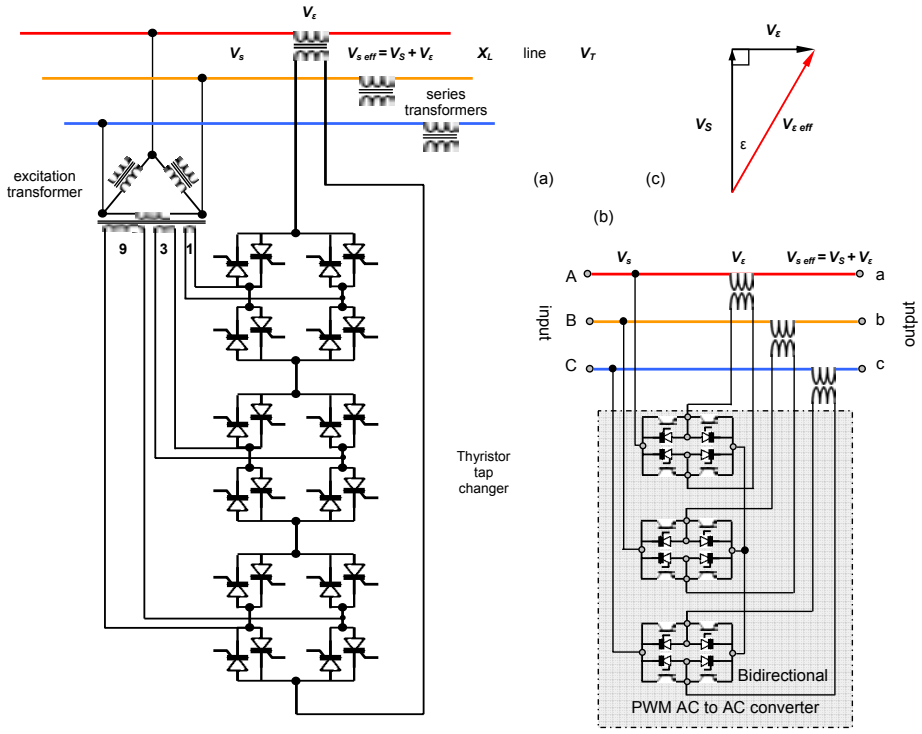


Figure 20.19. *Series phase angle compensator with quadrature boosting: (a) thyristor transformer tap changer; (b) bidirectional PWM ac to ac converter; and (c) phasor diagram for phase shift ±ε giving quadrature boosting, Φ=½π.*

Continuous Voltage Regulators, CVR

AC/AC Continuous Voltage Regulators (CVR) or PWM (or self commutated) AC voltage controllers (or stabilizers), are fast and tolerant power-electronics arrangements without moving parts and with no need for tap changing. The AC choppers are realized on the basis of single fully bidirectional switches with turnoff capability, thus the CVR topology in Figure 20.19c, with bipolar PWMD, allows bipolar regulation of the phase output voltage.

20.9 Self Commutating FACTS Devices - Custom Power

Reactive power compensators are essential transmission system devices used for voltage regulation, stability enhancement, and for increasing power transfer. FACTS devices are aimed at the transmission level, while compensators at the distribution level are term custom power controllers. FACTS compensators are based on the more robust (but slower responding) thyristor natural (or line) commutation technology. Custom power, as shown in figure 20.20 for single-phase compensators, although possibly operating on the same principles as FACTS devices, is based on less robust (but faster responding) self commutation devices (GCThyristor and IGBT), hence tends to be associated with the lower voltages of the distribution system, where more sensitive equipment and embedded, distributed generation may be connected.

Traditionally the mature technology of passive harmonic notch and low pass L-C based filters were used to control harmonic pollution and switched capacitor banks were used for power factor correction.
FACTS transmission level devices considered, based on natural commutated thyristor configurations, include the TSC, TCR, etc. Custom power distribution level devices to be considered are static synchronous compensators, which include the dynamic voltage restorer - DVR, shunt compensator -

STATCOM, and the unified power flow controller - UPFC. They are based on GCThyristor or IGBT PWM inverter/converter bridge topologies which use a large reactive dc link, intermediate energy source (a dc-side capacitor or a dc-side inductor). The topologies are commonly called active filters even though they can compensate for reactive and active power (at the fundamental frequency 50/60Hz) as well as perform harmonic filtering functions (at multiples of the fundamental frequency).

Active (self-commutating GCThyristor and IGBT) inverter based topologies provide compensation for

- Voltage and current harmonics
- Reactive power
- Neutral current
- Unbalanced loads
- Unbalanced phase voltages

The PWM inverter approach adopted can be

- Inductive dc-link, current source PWM inverter, CSI, which acts as a controllable sinusoidal current source to compensate for non-linear load harmonics
  - requires ac output shunt capacitance to filter current harmonics from the generated square-wave current waveform
  - self-supporting, large dc-link reactor
  - high reliability
- Capacitive dc-link, voltage source PWM inverter, VSI, which acts as a controllable sinusoidal voltage source
  - requires output series inductance to filter voltage harmonics from the generated square-wave voltage waveform
  - self-supporting, large dc-link capacitor
  - applicable to multilevel inverter topologies for better system voltage matching

The CSI or VSI inverter based compensators can be connected to the system, usually through voltage matching transformers (where in the case of the VSI, the transformer leakage inductance may provide voltage output harmonic filtering), in any of three ways, as shown in figure 20.20a.

- Series inverter compensator – called a dynamic voltage restorer, DVR – compensating for
  - Line voltage harmonics
  - Spikes, notches and sags and swells
  - Balance and regulate load or line terminal voltage
  - Static VAr generator to stabilise and improve voltage profile
- Shunt inverter compensator – called a STATCOM – compensating for
  - Line current harmonics
  - Reactive power compensation
  - Balancing of three-phase loads
- Series plus shunt inverters/converters – called a unified power flow controller, UPFC
  - Performs the functions of both the shunt and series compensators
  - Two inverters share a common dc current or voltage link

Each of these three compensators belong to the generic family of static synchronous compensators.

The classification matrix for custom power devices is shown in figure 20.20. For ease of comparison and understanding, single-phase versions are show, with specific three-phase configurations considered in the sections to follow. Three-wire (floating neutral) and four-wire (connected neutral) PWM inverter topologies are extensions to the basic single-phase topologies considered.

Given the switching frequency limitations of 6.5kV IGBT and IGCThyristor technology (<1 kHz), synchronous selective harmonic elimination SHE is used for switching modulation rather than standard PWM or SVM. If transformer coupling is adopted, low-voltage IGBTs (3.3kV) allow a higher switching frequency (2.5kHz). The trade-off is that the transformer core must transmit with minimal attenuation, the highest required compensation harmonic. Thinner transformer laminations (0.1mm and 0.05mm, as opposed to 0.3mm for 50/60Hz) allow a higher operating frequency, but at the expense of reduced flux density and higher core losses/kg. The best trade-off is 0.1 mm lamination thickness compensating for harmonics up to and including the 17th and 19th. Higher harmonics (the 23rd, 25th and higher) are more effectively attenuated with passive *L-C* notch (with long term drift problems) and low pass filters. When only harmonic distortion correction is required, as opposed to any 50/60Hz VAr compensation, nano-crystal amorphous core materials are an effective alternative the silicon grain orientated steels.

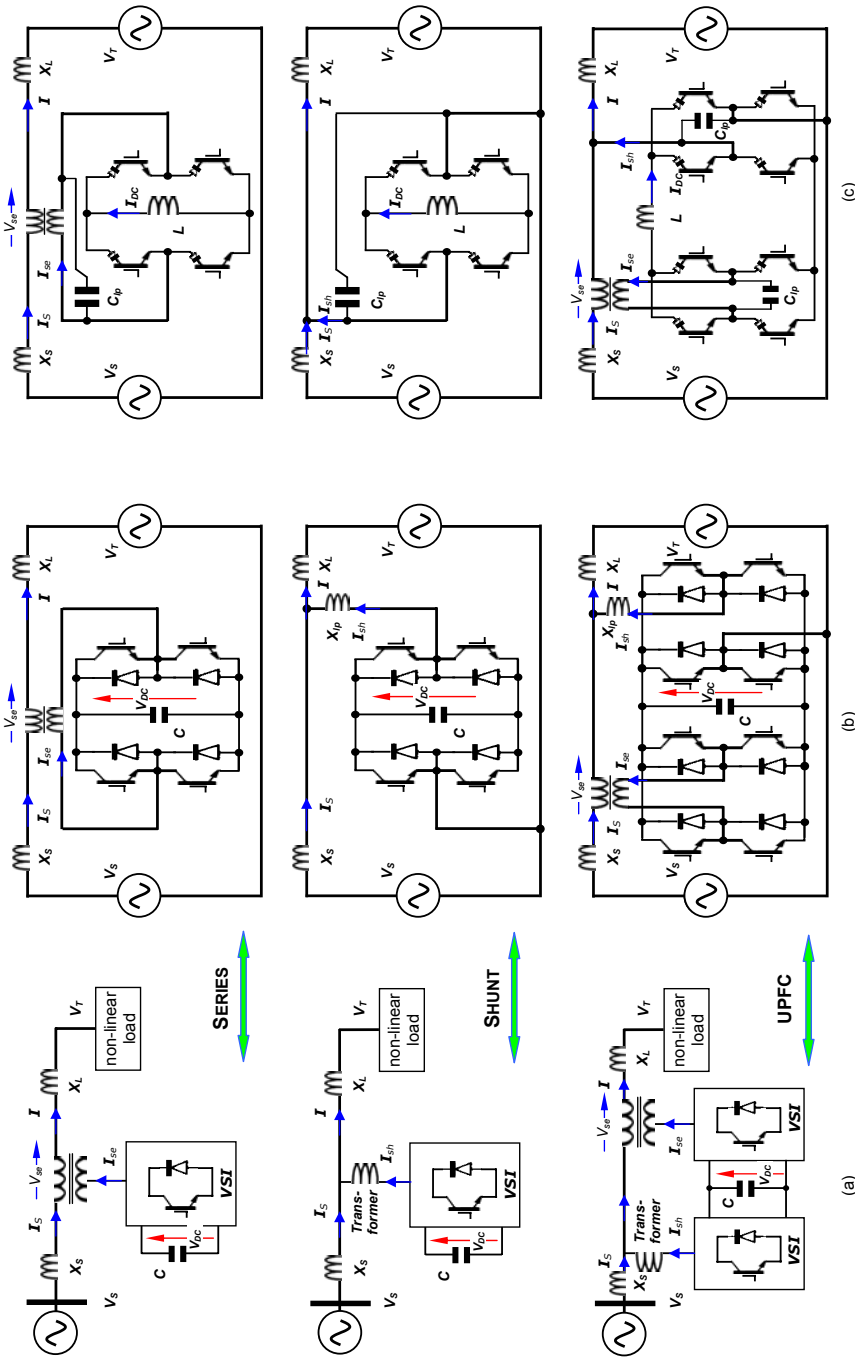


Figure 20.20. *Static synchronous compensator family: (a) transmission schematic of voltage source compensators (as power filters) transformer coupled to the ac network; (b) single-phase static synchronous compensators using dc-link capacitor voltage sources; and (c) static synchronous compensators using dc-link inductor current sources.*

BWW

20.9.1 - Static synchronous series compensator (SSSC) or Dynamic Voltage Restorer - DVR

The static synchronous series compensator (is a dynamic voltage restorer without active power transfer) is a transformer coupled, PWM voltage source inverter that functions in series with the distribution line as shown in figure 20.21. Theoretically, it draws no power from the line since it uses a capacitor on its dc link which provides only reactive power. This makes the DVR a versatile regulating compensator. In steady-state it functions as a series phase shifter SPS, injecting a variable magnitude and angle voltage at one line end in order to control both the active and reactive power flow. The phase angle (which controls the real power if a suitable bidirectional dc-link source exits) is controllable between 0 and 2π, as shown in the phasor diagram in figure 20.21c. The magnitude of $\boldsymbol{V_{DVR}}$ is controlled by the inverter PWM modulation depth. As well as being a SPS it also functions as a variable series impedance compensator.

Generally, from figure 20.21a

$$\begin{aligned} V_{DVR} &= V_T - V_S \\ &= |V_{DVR}|(\cos\varphi + j\sin\varphi) = V_d + jV_q \end{aligned} \qquad (20.64)$$

If the DVR does not involve any active power source and the only real power drawn from the ac line is that necessary to maintain the capacitor voltage so as to compensate for inverter and coupling transformer power losses, then $V_d = 0$ and V_q is in quadrature to the compensator/line current. By varying the magnitude of V_q, the DVR performs the function of a variable reactance compensator, where

$$V_d \approx 0$$
$$\text{if } \begin{cases} V_q > 0 \text{ the DVR is capacitive} \\ V_q < 0 \text{ the DVR is inductive} \end{cases} \qquad (20.65)$$

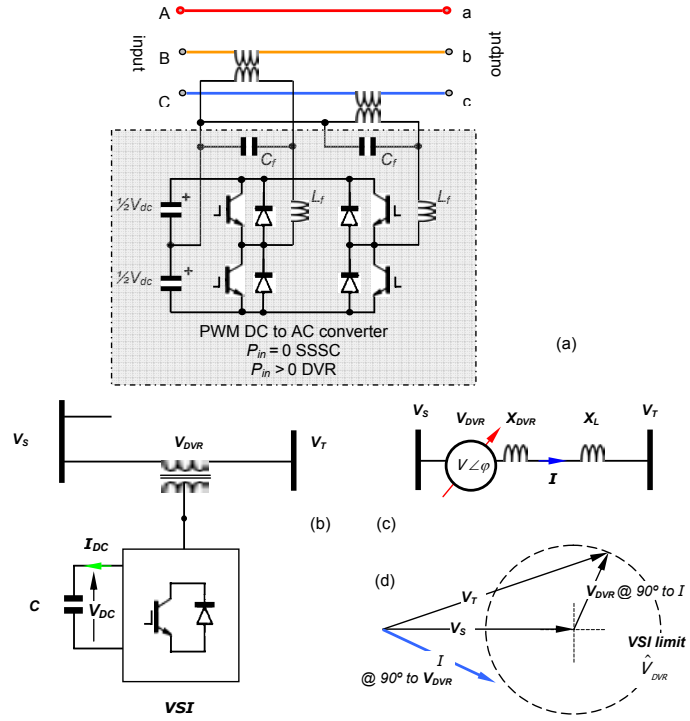


Figure 20.21. *Static synchronous series compensator (or dynamic voltage restorer DVR): (a) schematic and (b) block diagram of a voltage source inverter, transformer coupled in series with the ac network; (c) series connected DVR shown as a variable magnitude and phase angle voltage source SSSC; and (d) 50/60Hz operating phasor diagram of SSSC, where V_{DVR} is always perpendicular to the line current, I.*

BWW

Within its energy limits, the DVR is suited for dynamically compensating any line feeding sensitive or critical equipment for

- voltage harmonics
- power factor correction

and for a short duration

- voltage sags and swells
- voltage imbalances
- outages

In the standby mode, the output voltage is zero and the inverter losses are low since no switching occurs. By turning on all the upper (or lower but not both) switches in the VSI inverter, the three single-phase transformers and inverter are seen in the line as a short circuit (as for a current transformer). Given transformers are necessary, voltage matching of the VSI devices facilitates the use of 3.3kV IGBT technology that allow modulation frequencies above 2kHz, which is necessary for active filtering. Specific single-phase transformer coupling can be avoided if the DVR is connected at the opened star point of the main ac supply Y configured transformer or autotransformer. Alternatively, access to the transformer star point allows the use of a three-phase autotransformer rather than three single-phase transformers. A CSI is well suited for series application (with an outer voltage loop) since it is normally operated with the switches in an on-state, thereby ensuring that the DVR is seen as a short-circuit in the standby/fault mode.

Voltage harmonic cancellation

Each of the three source voltages can be expressed in terms of a fundamental frequency phase voltage and its harmonics. Inserting appropriate voltages can compensate for harmonics associated with the terminal supply voltages under balanced an unbalanced conditions in three wire systems.

Series Voltage regulation

The terminal voltage V_T in figure 20.22a draws a lagging current I_T and the series compensator V_{DVR} is to maintain the load voltage V_T constant, but at any angle with respect to V_S. From Kirchhoff's voltage law

$$V_S = jI_T X_R + V_{DVR} + V_T \qquad (20.66)$$

The series compensator can deliver any voltage up to a maximum $\hat{V}_{DVR}$, as shown by the circle outer locus with centre O for the series regulator in figure 20.22b. If V_T is held constant then the source V_S can have a magnitude and angle that lies anywhere within the circle. If V_T sags and swells (changes length) then provided the variation is within the circle, V_{DVR} can compensate to maintain a constant voltage V_S. Maximum and minimum voltage compensation needed from V_{DVR} occurs when the source V_S forms a tangent to the circle as shown in figure 20.22c. In each case the current is not in phase with the compensation voltage, hence the compensating converter must transfer real power. The effective sending voltage $V_{S\,eff}$ is phasor N-O. The phasor O-W represents the case when power is delivered from the compensator in an effort to compensate for the sagging (reduced) V_T voltage phasor N-W, while the phasor O-X represents the case when V_T has swelled to phasor N-X and power is drawn by the compensating converter whilst attempting to decrease the line voltage. The inverter in figure 20.21a creating V_{DVR} must have a bidirectional dc voltage supply maintaining the dc-link voltage.

The converter dc-link voltage can be self-supporting if no energy is lost or gained by the dc-link when the line current is at quadrature to the compensator voltage, as shown in figure 20.22d. In this case, the source voltage V_T can be compensated when its voltage phasor lies along the line W-O-X. The magnitude range of the voltage V_T that can be compensated, is reduced. The range is reduced from phasor N-W to phasor N-X in figure 20.22c converter power can be transferred, to between N-Y to N-Z when only reactive energy can be transferred by the compensating converter, as in figure 20.22d.
The series compensation is effective for a wide range of line impedances, including low impedance stiff feeders, provided the line impedance phasor is within the compensating circle. The basic series converter arrangement can also be using for voltage distortion compensation.

The phasor diagrams in figure 20.23 show the series compensator operating in three different modes, namely, capacitive and inductive compensation and in a mode of reversing the power flow. The effective sending voltage $V_{S\,eff}$ is shown as a dashed line in each case. Series line resistance has been added, giving the voltage phasor V_R in phase with the current in the parts of figure 20.23. The line voltage drop, V_{line} is therefore the line impedance voltage drop.

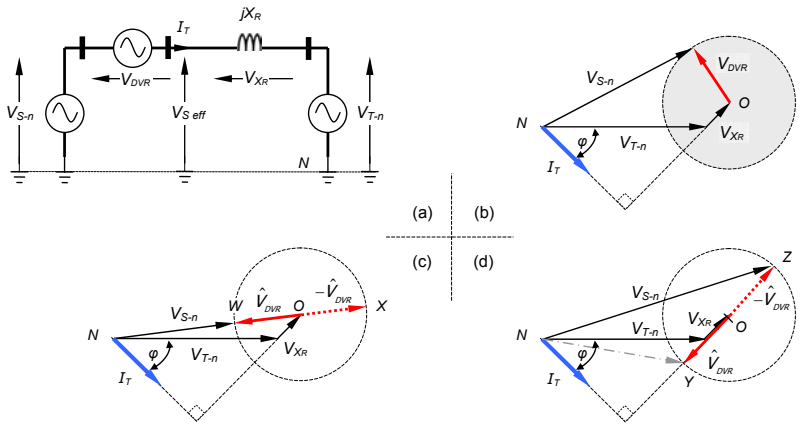


Figure 20.22. *Static series voltage compensation:*
(a) series compensated network; (b) general series voltage compensation; (c) voltage sag and swell real-power compensation; and (d) quadrature reactive-power series voltage compensation.

Capacitive compensation

SSSC capacitive compensation is shown in figure 20.23a. The phasor quadrature voltage V_{DVR} is in the opposite direction of the phasor V_{XR} and V_{DVR} lags the current phasor I by 90°. For the same line total voltage drop V_{line}, the voltage phasor associated with the line reactance V_{XR} increases and results in a transmission line current increase.

Inductive compensation

Inductive compensation is shown by the phasor diagram in figure 20.23b. The quadrature voltage is in the same direction as the phasor V_{XR} and V_{DVR} leads the current phasor I by 90°. For a constant V_{line}, the phasor V_{XR} decreases. The transmission line current reduces and results in reduced power flow.

Reverse power flow

The ability of an SSSC to reverse power flow is illustrated by the phasor diagram in figure 20.23c. Operation is similar to inductive compensation but the VSC voltage is increased until larger in magnitude than V_{XR}. For a constant V_{line}, V_{XR} reverses direction and the current phasor I reverses, thence power flow is reversed.

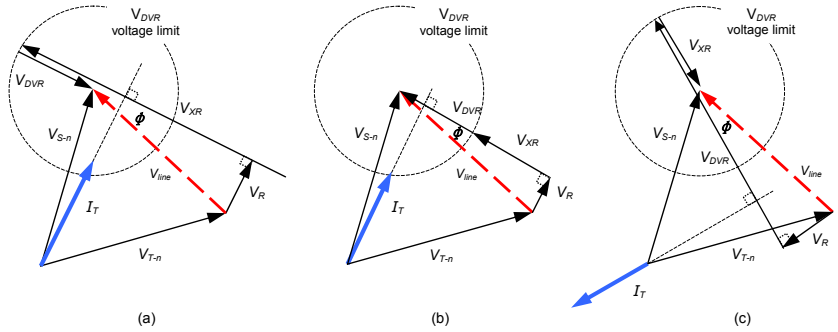


Figure 20.23. *Phasor diagrams for the three series compensator modes of operation:*
(a) capacitive compensation; (b) inductive compensation; and (c) power flow reversal.

20.9.2 - Static synchronous shunt compensator - STATCOM

The STATCOM (***stat***ic synchronous shunt ***com***pensator) is a shunt compensator comprising a current or voltage source inverter, shunt connected to the ac system through a first order passive filter, as shown in figure 20.24a for a VSI. The dc-side main reactive energy storage element is

- a dc capacitor (voltage source, VSI) in which case the interconnect filter comprises series line inductance for attenuating VSI output voltage harmonics or
- an inductor (current source, CSI) in which case the interconnect filter comprises shunt capacitance for bypassing CSI output current harmonics. Not favoured over the VSI.

No net energy is needed, except to replace the energy dissipated in the inverter and filter components. The STATCOM function is to

- regulate the line at the point of connection when functioning in a SVC mode and/or
- minimise current harmonics by anti-phase current injection action – as an active filter.

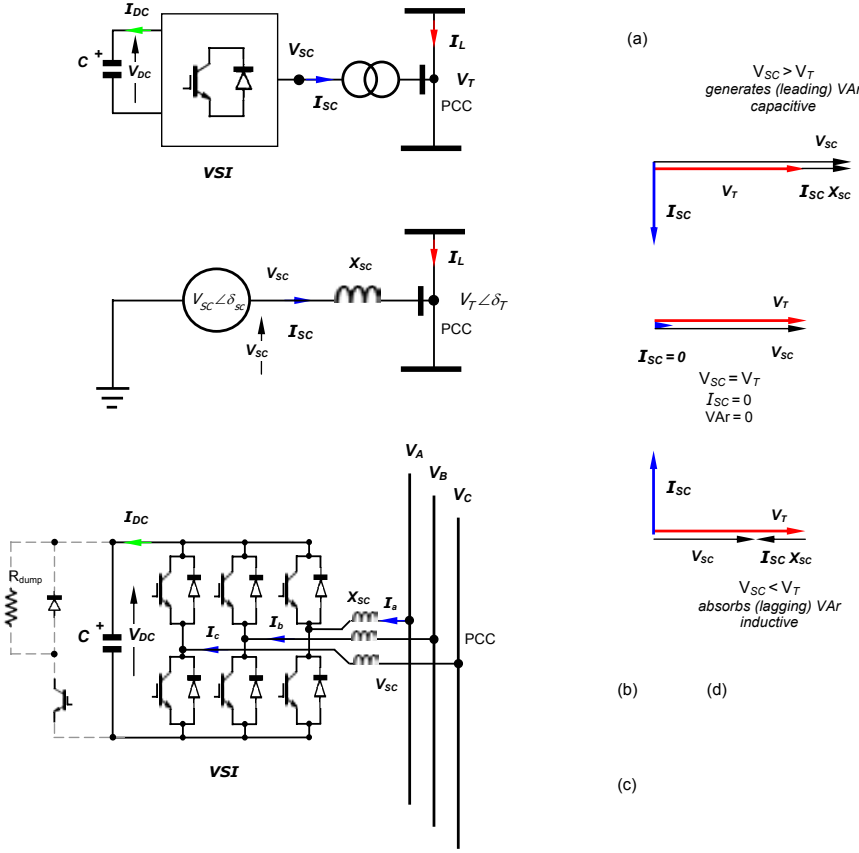


Figure 20.24. *Active shunt regulator - STATCOM: (a) a voltage source inverter VSI, inductively shunt connected (transformer coupled) to the ac network; (b) shunt connected STATCOM shown as a variable magnitude and phase angle voltage source; (c) main VSI circuit; and (d) phasor diagrams for leading (upper phasor diagram) and lagging (lower phasor diagram) modes of operation.*

Figure 20.24b shows the system model, while figure 20.24c shown the simplified VSI circuit. The series voltage harmonic filtering inductance can be the leakage inductance associated with the three single-phase line voltage matching transformers or three auto-transformers. A dc chopper, with a dumping resistor as load, may be used across the dc-link capacitor to limit VSI over-voltage during intermittent transients when the STATCOM acts as an uncontrolled rectifier, created by the VSI freewheel diodes

shown in figure 20.24c. The phasor diagram in figure 20.24d, for the line to neutral voltage, is associated with the circuit in figure 20.24b, in conjunction with the power and reactive power equations given by equations (20.6) and (20.8), can be used to explain STATCOM operating principles.

$$P = \frac{V_{SC} V_T}{X_{SC}} \sin(\delta_T - \delta_{SC}) = 0 \qquad (20.67)$$

$$Q_{SC} = V_T \times \frac{V_T - V_{SC} \cos(\delta_T - \delta_{SC})}{X_{SC}} = \frac{V_T}{X_{SC}} (V_T - V_{SC}) \qquad (20.68)$$

$$I_{sc} = \frac{V_{T-n} - V_{sc-n}}{X_{sc}} \qquad (20.69)$$

In steady-state, the inverter output voltage fundamental V_{SC} (which is controlled by the PWM modulation index) is in phase with the ac line voltage V_T ($\delta_{SC} = \delta_T$), while the STATCOM current I_{SC} always leads or lags the line voltage by 90º because of the inductive reactive coupling X_{SC} therefore $\cos(\delta_T - \delta_{SC}) = \cos 0 = 1$, Thus $P \approx 0$ is maintained as given by equation (20.67) when $\sin(\delta_T - \delta_{SC}) = \sin 0 = 0$.
From equations (20.67) and (20.68), since $\delta_{SC} = \delta_T$, only reactive power flows and

- when the STATCOM voltage V_{SC} is less than the line voltage V_T ($|V_{SC}| < |V_T|$), reactive power is absorbed (inductive) by the STATCOM, from V_T to V_{SC} (which tends to decrease the point of connection voltage), while
- when the voltage magnitudes are reversed, that is $|V_{SC}| > |V_T|$, the STATCOM generates (capacitive) reactive power (which tends to increase the point of connection voltage).
- When $V_{SC} = V_T$, the voltage V_{SC} across the connecting inductance X_{SC} is zero, so no STATCOM current flows $I_{sc} = 0$.

Thus a STATCOM behaves like a shunt inductor (I lags V) without a physical inductor or magnetic field, and like a shunt capacitor (I leads V) without a physical capacitor or electric field.
When used in a voltage regulation mode (as opposed to a VAr control mode with constant reactive power output) the STATCOM terminal I-V characteristics are as shown in figure 20.10c for the SVC.
The dc link capacitor is initially charged through the VSI freewheel diodes which form an uncontrolled three-phase line rectifier. Subsequently the STATCOM is controlled to self regulate its dc-link voltage, V_{DC}, as follows.

When the fundamental voltage of the STATCOM slightly leads the ac supply voltage, V_{SC} leads V_T, the capacitor voltage decreases resulting in $V_{SC} < V_T$, real power is transferred from the dc link to the ac line and reactive power is absorbed by the STATCOM – lagging power mode.

When the STATCOM fundamental voltage slightly lags the ac supply voltage, the capacitor voltage increases, $V_{SC} > V_T$, real power is transferred from the ac line to the dc link and reactive power is generated by the STATCOM – leading power mode.

Thus the STATCOM fundamental magnitude V_{SC} controls the reactive power, while the phase angle between the STATCOM and the ac line, $\delta_T - \delta_{SC}$, controls real power flow. In practice, when V_{SC} slightly lags V_T (δ_{SC} lags δ_T), the capacitor voltage V_{DC} is maintained whilst catering for system inverter and transformer power losses. In this way no separate dc power supply is needed to maintain the dc-link capacitor voltage.
Notice that the dc-link voltage will always be greater than the rectified ac grid voltage due to the (uncontrolled) rectification action through the six inverter bridge freewheel diodes.
Although practical limits exist on the magnitude of V_{SC}, the STATCOM power load angle δ_{SC} is continuously adjustable between 0 and 2π, but operates near the line phase angle δ_T in order to minimize real power transfer.
The SATCOM can generate more reactive power during a fault than the SVC since

- from equation (20.40), SVC capacitive reactance power decreases proportionally to voltage V_M while
- from equation (20.68), STATCOM capacitive reactive power decreases linearly with voltage V_{SC}.

Shunt voltage regulation

The terminal voltage V_T in figure 20.25a draws a lagging current I_T and the shunt compensator V_{sh} is to maintain the load voltage V_T constant, but at any angle with respect to V_S. From Kirchhoff's voltage law for the right hand loop in figure 20.25a,

$$V_T = jI_{sh} X_{sh} + V_{sh-n} \qquad (20.70)$$

The shunt compensator can deliver any current from zero up to a converter maximum $\hat{I}_{sh}$, giving, for a fixed compensation reactance X_{sh}, the circle outer locus with centre O as shown in figure 20.25b. Thus a small change in the magnitude and phase of V_{sh} will cause the shunt reactance voltage to rotate through 360°. From Kirchhoff's current law, the shunt regulator point of common contact, PCC, yields

$$I_S = I_{sh} + I_T \qquad (20.71)$$

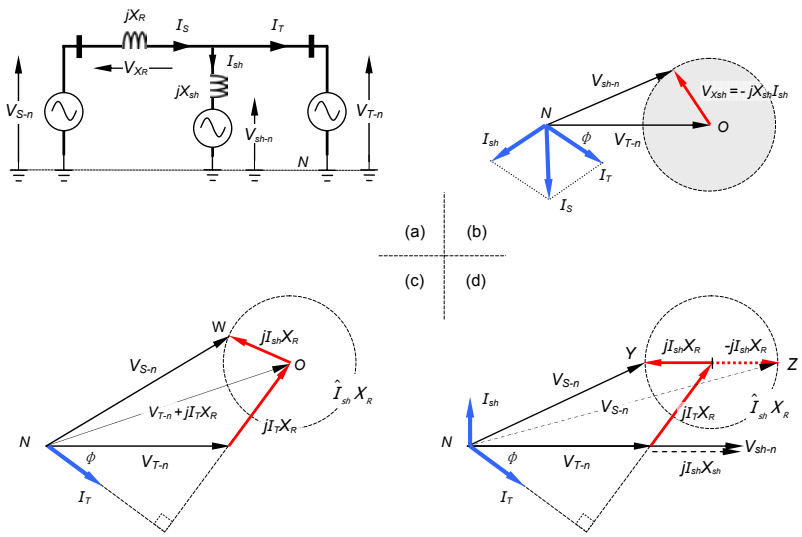


Figure 20.25. *Active shunt compensation:*
(a) shunt compensated network; (b) general shunt voltage compensation phasor diagram;
(c) shunt voltage compensation; and (d) quadrature reactive-power shunt current compensation.

These currents are shown in figure 20.25b. The outer voltage loop in figure 20.25a gives

$$V_{S-n} = jI_S X_R + V_{T-n} \tag{20.72}$$

Substituting equation (20.71) gives

$$\begin{aligned} V_{S-n} &= j\left(I_T + I_{sh}\right) X_R + V_{T-n} \\ &= \left\{V_{T-n} + jI_T X_R\right\} + jI_{sh} X_R \end{aligned} \tag{20.73}$$

Since the load network V_T in conjunction with the load network current I_T specify the load power factor, the phasor N-O, $\{V_{T\text{-}n} + jI_T X_R\}$ in figure 20.25c is fixed. Because I_{sh} can be varied between zero and $\hat{I}_{sh}$, phasor $V_{S\text{-}n}$ can lie anywhere within the circle shown in figure 20.25c and not affect the load network V_T. In figure 20.25c, the shunt regulator is delivering real power into the load network since the shunt compensating network PCC voltage $V_{T\text{-}n}$ and shunt current (angle given by phasor O-W in figure 20.25c) are not in quadrature.
Figure 20.25d shows the loci for the case when no real power is transferred by the shunt compensator, since the compensator current I_{sh} is in quadrature to the shunt voltage $V_{sh\text{-}n}$, which is in phase with the load network V_T. The allowable range of variation on the source voltage $V_{S\text{-}n}$ is a minimum for phasor N-Y (voltage-swell) and a maximum for phasor N-Z (voltage-sag), as shown in figure 20.25d. This range of possible voltage compensation is determine by the line reactance X_R, which specifies the inverter current rating I_{sh} needed to produce the necessary compensation range, namely the diameter of the circle in figure 20.25c. That is, the lower the line reactance X_R the higher the necessary compensating inverter current rating for a given voltage compensation range.

The shunt compensator operates in a type of current push-pull or sourcing-sinking mode.

- When the source voltage V_S is too high, voltage swell, the shunt draws or sinks current additional to the load current in order to increase the voltage across the line reactance X_R, thereby tending to decrease the load voltage V_T.
- When the source voltage V_S sags, the shunt compensator sources current to the load network V_T, thereby reducing the source current which decreases the voltage across the line reactance X_R, making a higher component of the source voltage available across the load network V_T.

During each mode of operation, the phasor angular relationships must be observed within this simplistic explanation.
The basic shunt converter arrangement can also be using for line current distortion compensation.

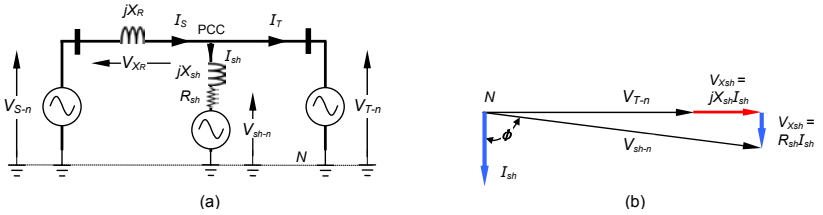


Figure 20.26. *Active shunt compensator used for power factor correction:*
(a) shunt compensated network and (b) phasor diagram.

Power factor correction
The shunt compensator can be used for power factor correction at the PCC. The compensator current I_{sh} is set to be 90° behind the PCC voltage $V_{T\text{-}n}$, with the magnitude of the current I_{sh} determining the magnitude of the compensation. This is achieved by ensuring that the load voltage and shunt regulator voltage are in phase, but the relative magnitudes are varied ($V_{sh\text{-}n} > V_{T\text{-}n}$). Since only VAr are involved from the shunt regulator, no shunt regulator dc voltage supply is needed to maintain the dc-link capacitor, except inverter losses must be accounted for. By ensuring the shunt voltage $V_{sh\text{-}n}$ slightly lags the line voltage $V_{T\text{-}n}$, the necessary inverter losses can be provided from the grid. If the inverter losses are incorporated, as represented by the resistor in figure 20.26a, then the resultant phasor diagram in figure 20.26b complies with the following output loop voltage equation.

$$V_{sh-n} = I_{sh} R_{sh} + jX_{sh} I_{sh} + V_{T-n} \tag{20.74}$$

The reactive power provided to the ac system from the shunt power factor controller is $Q = I_{sh}\, V_{T\text{-}n}$, while $P = V_{sh\text{-}n}\, I_{sh} \cos\phi$ real power is drawn from the line to cater for the inverter power losses. Since no separated dc-link voltage source is required, the shunt regulator is acting as a STATCOM, as considered earlier.

Harmonic current compensation
Figure 20.27 shows a system with a shunt active filter for harmonic current compensation of a non-linear, diode rectifier, where the active filter circuit consists of a three-phase voltage-fed PWM inverter and a dc-link capacitor, C_{dc}. The active filter is controlled to draw the compensating current, i_{AF} , from the utility that cancels the harmonic current flowing on the ac side of the diode rectifier with an inductive dc load.
Equation (20.19) represents the *α*-phase and *β*-phase compensating currents:

$$\begin{bmatrix} i_{AF\alpha} \\ i_{AF\beta} \end{bmatrix} = \begin{bmatrix} e_\alpha & e_\beta \\ -e_\beta & e_\alpha \end{bmatrix}^{-1} \begin{bmatrix} p_{AF} \\ q_{AF} \end{bmatrix} \tag{20.75}$$

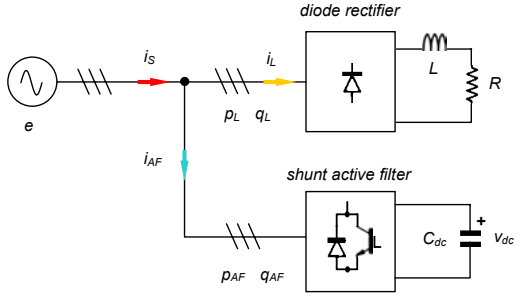


Figure 20.27. *Active shunt filter used for ac current distortion compensation.*

The powers p_{AF} and q_{AF} are the three-phase instantaneous real and imaginary power on the ac-side of the active filter, and can be extracted from p_L and q_L, which are the three-phase instantaneous real and imaginary powers on the ac-side of a harmonic-producing (non-linear) load. When the active filter compensates for the harmonic current produced by the non-linear load:

$$p_{AF} = -\tilde{p}_L \qquad q_{AF} = -\tilde{q}_L \tag{20.76}$$

where, $\tilde{p}_L$ and $\tilde{q}_L$ are the ac components of p_L and q_L, respectively. The dc components of p_L and q_L correspond to the fundamental current in i_L and the ac components correspond to the harmonic current. Two high-pass filters can be used in the control circuit to extract $\tilde{p}_L$ from p_L and $\tilde{q}_L$ from q_L.
The active filter draws and releases p_{AF} from the utility, and delivers it to the dc capacitor, assuming no loss dissipation in the active filter. Thus p_{AF} produces a voltage fluctuation on the dc capacitor. The amplitude of p_{AF} is assumed constant, then the lower the frequency of the ac component, the larger the voltage fluctuation. The dc capacitor has to absorb or release electric energy given by the integration of p_{AF} with respect to time. Thus, the relationship between the instantaneous voltage across the dc capacitor, v_{dc} and p_{AF} is:

$$½C_{dc}v_{dc}^2(t) = ½C_{dc}v_{dc}^2(0) + \int_o^t p_{AF}dt \qquad (20.77)$$

This implies that the active filter needs large dc capacitance to suppress the voltage fluctuation in order to harmonic compensate $\tilde{p}_L$. The main purpose of the voltage-fed PWM inverter is to perform an interface conversion between the utility and the dc capacitor.

The active filter draws q_{AF} from the utility, as shown in Fig. 20.27. However, q_{AF} makes no net contribution to energy transfer in the three-phase circuit. No energy source is required to the dc side of the active filter, independent of q_{AF}, whenever $p_{AF} = 0$.

Voltage distortion compensation
The shunt regulator can be used to cancel line harmonic voltages. If harmonics of the fundamental supply frequency ω are cancelled, the load voltage and current at that frequency are zero as shown in figure 20.28. Then, from figure 20.28, where $I_{S\text{-}harm} = I_{sh\text{-}harm}$

$$V_{S-n-harm} - jI_{S-harm}n\omega L_R = V_{sh-n-harm} + jI_{sh-harm}n\omega L_{sh} \qquad (20.78)$$

That is

$$V_{sh-n-harm} = -\frac{L_{sh}}{L_R} \times V_{S-n-harm} \qquad (20.79)$$

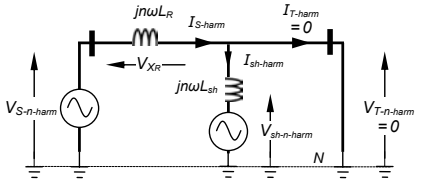


Figure 20.28. *Active shunt compensator used for voltage distortion compensation.*

The necessary cancelling voltage magnitude from the shunt regulator is dependant on the relative magnitudes of the two line reactance's, independent of harmonic frequency. If $L_{sh} > L_R$ the shunt compensator must produce an anti-phase voltage greater in magnitude than the harmonic line voltage.
Series compensation techniques are more effective for line voltage distortion compensation, while shunt compensation methods are more effective for line current harmonic compensation.

The four-quadrant *P-Q* and boundary phasor diagrams for the shunt regulator are shown in figure 20.29.

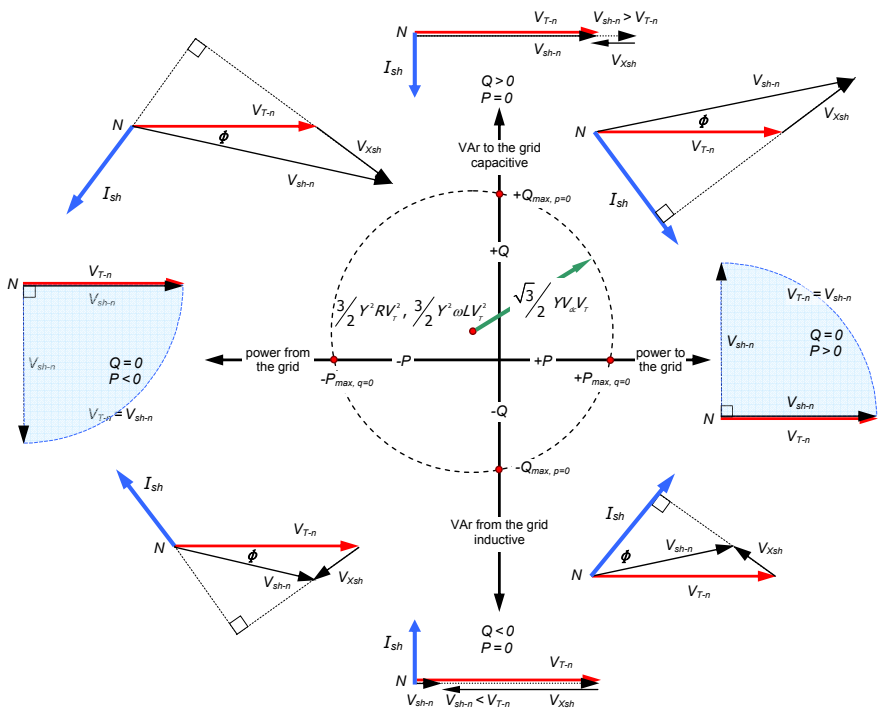


Figure 20.29. *Phasor diagrams for P and Q power exchange for the shunt compensator.*

Table 20.1: Comparison of STATCOM and SVC

property	STATCOM	SVC
I-V characteristic	Current source Good under-voltage performance	Impedance source Good overvoltage performance
Control range	Symmetrical Otherwise hybrid solution	Adjustable with cascaded TCR/TSC
Modularity	Redundancy Compensated aging degradation Common inverter to other applications	Redundancy Aging degradation TCR/TSC branches common additions to SVC
Response time	1 to 2 cycles No natural commutation delays	2 to 3 cycles Limited by supply frequency
Transient behaviour	Self protecting on critical system faults	Active before, during and after transient conditions
Volume requirements	40% to 50% of SVC	100%
On-line availability	96% to 98% of time	>99% of the time
Capital costs	120% to 150% that of SVC	100%

20.9.3 - Unified power flow controller - UPFC

The unified power flow controller shown in figure 20.30a consists of a shunt and a series static synchronous compensator, where the two compensating inverters are connected back to back, and are decoupled by sharing a common dc link energy storage element (inductor or capacitor). As such, the two converters can operate independently, giving a versatile compensator that can simultaneously perform the function of either or both of the static synchronous series and shunt compensators, namely

- Active power flow
- Reactive power flow
- Voltage magnitude control
- Voltage harmonic elimination (active power filtering, see section 20.2.8)
- Current harmonic elimination (active power filtering, see section 20.2.8)

The shunt compensator provides

- voltage regulation at the point of connection by injecting reactive power into the line and
- balance of the real power exchanged between the two compensators when providing for inverter and transformer losses and any real power transferred by the series compensator.

The series compensator is used to

- control the real and reactive power by injecting a controllable magnitude and phase compensating voltage in series with the line.

The UPFC thereby fulfils the functions of reactive shunt compensation, active and reactive series compensation, and phase shifting. Additionally, the UPFC can provide transient stability control by suppressing system oscillations.
As shown in figure 20.30, the UPFC can control simultaneously the three parameter associated with line power flow (line impedance, voltage, and phase angle). The UPFC is connected at either the sending or the terminal points of the distribution/transmission system.

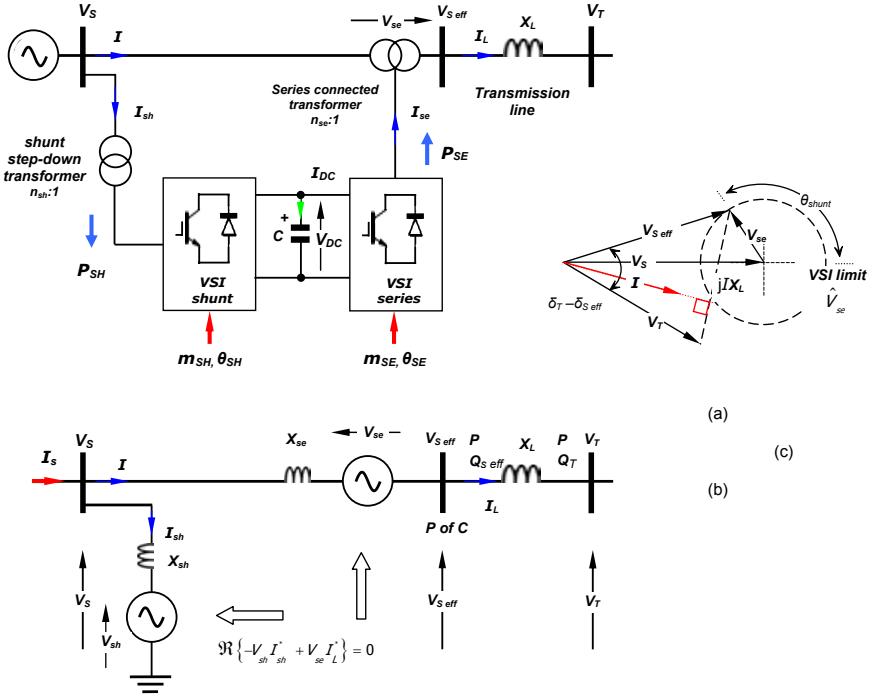


Figure 20.30. *Unified power flow controller - UPFC: (a) single line diagram of the UPFC showing decoupled back to back connected inverters and matching transformers; (b) UPFC equivalent circuit; and (c) phasor diagram for system voltages and line current, I_L.*

The series and shunt converters are operated to give point of connection voltages

$$V_{se} = |V_{se}|(\cos\theta_{se} + j\sin\theta_{se}) \qquad 0 \le \theta_{se} \le 2\pi$$
$$V_{sh} = |V_{sh}|(\cos\theta_{sh} + j\sin\theta_{sh}) \qquad 0 \le \theta_{sh} \le 2\pi \tag{20.80}$$

The magnitudes of converter voltages V_{sh} and V_{se} are controlled by the turns ratio of the matching transformers, the PWM modulation depth, and are restricted by the operational voltage limits (both upper and lower voltage limits) imposed by the inverter technology.

$$V_{SH} = m_{SH}\frac{V_{dc}}{2\sqrt{2}\,n_{SH}V_B}$$
$$V_{SE} = m_{SE}\frac{V_{dc}}{2\sqrt{2}\,n_{SE}V_B} \tag{20.81}$$

where m is the inverter modulation index, n is the coupling transformer turns ratio, V_B is the transmission side base voltage, and V_{dc} is the back to back inverter dc link voltage.
The effective sending end voltage $V_{S\,eff}$, hence power, is controlled by adjusting the series voltage V_{se}, that is

$$V_{S\,eff} = V_S + V_{se} \tag{20.82}$$

The active power drawn by the series converter should equal the active power generated by the shunt converter (minus inverter and transformer losses) and vice versa, that is

$$\Re\{-V_{sh}I_{sh}^* + V_{se}I_L^*\} = 0 \tag{20.83}$$

Because line energy can be transferred readily between both converters in compensating for converter and transformer losses, the dc-link capacitor can be small, yet be maintained at the necessary rated link voltage. A consequence of the back-to-back connection is that the dc-link capacitor decouples the two converters and the shunt and series converter reactive powers can be controlled independently. Both converters can provide reactive power, and power for the series converter can be provided through the shunt converter. Because the series converter can now provide (and absorb) real power, the injected shunt voltage magnitude and relative phase are unrestricted, within the *I*-*V* limits of the two inverters. This is shown by the circle in the phasor diagram in figure 20.30c, where unlike for the DVR, as shown in the phasor diagram in figure 20.21c, the line current I_L and the series compensation voltage V_{se} are not restricted to be at quadrature (that is, real power transfer can be involved with UPFC operation).

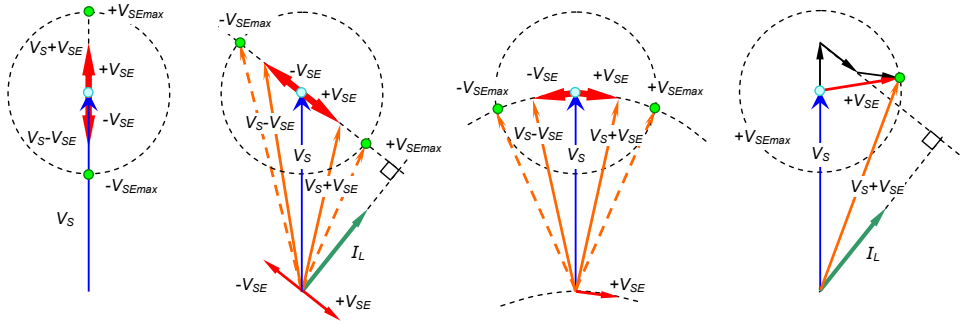


Figure 20.31. *Phasor diagrams for the UPFC series operating modes.*

The series converter can be operated in any of four modes:

- Voltage regulation - figure 20.31a. The magnitude of the sending bus voltage V_S is regulated (increased or decreased) by injecting a voltage V_{SE} of maximum magnitude V_{SEmax}, in phase (or out of phase) with V_S, thus avoiding the need for a transformer tap changer.
- Line compensation - figure 20.31b. Series reactive compensation is obtained by series injecting a voltage V_{SE} of maximum magnitude V_{SEmax}, orthogonal to the line current I_L. The effective voltage across the line impedance X_L is decreased (or increased) if the voltage V_{SE} lags the current by 90° (or leads the current I_L by 90°).
- Phase angle regulation - figure 20.31c. The required phase shift is realized by injecting a voltage V_{SE} of maximum magnitude V_{SEmax}, that shifts the phase angle of V_S by ±θ while keeping the magnitude of V_S constant.

- Power flow control - figure 20.31d. Unified simultaneous control of terminal voltage (figure 20.31a), line impedance (figure 20.31b), and phase angle (figure 20.31c) means the UPFC is able to perform multifunctional power flow control. The magnitude and the phase angle of the series injected voltage V_{SE} is selected so as to produce a line current that results in the desired real and reactive power flow on the transmission line.

The complex conjugate of the complex power at the receiving end of the line is given by

$$S^* = P - jQ = \mathbf{V}_R^* \left(\frac{\mathbf{V}_S + \mathbf{V}_{SE} - \mathbf{V}_T}{jX} \right) \tag{20.84}$$

After compensation, the real and reactive power flows between $V_{S\,eff}$ and V_T are given by

$$P = \frac{V_S V_T}{X_L} \sin\delta_T + \frac{V_{S\,eff} V_T}{X_L} \sin\left(\delta_T - \delta_{S\,eff}\right) = P_o(\delta) + P_{SE}\left(\delta, \delta_{S\,eff}\right)$$

$$Q_T = V_T \times \frac{V_{S\,eff}\cos\left(\delta_T - \delta_{S\,eff}\right) - V_T}{X_L} + \frac{V_{S\,eff} V_T}{X_L} \sin\left(\delta_T - \delta_{S\,eff}\right) = Q_o(\delta) + Q_{SE}\left(\delta, \delta_{S\,eff}\right) \tag{20.85}$$

$$Q_{S\,eff} = V_{S\,eff} \times \frac{V_{S\,eff} - V_T \cos\left(\delta_T - \delta_{S\,eff}\right)}{X_L}$$

If $V_{S\,eff} = 0$ the real and reactive power of the uncompensated system result, as given by equations (20.6) and (20.8). The maximum *P*-*Q* compensation components, $V_T V_{SE\,max}/X_L$, occur when $\delta_T - \delta_{Seff} = \frac{1}{2}\pi$. The series DVR compensator can a voltage between 0 and V_{SEmax} with a rotational angle between 0 and 360°. This circle can be defined by

$$\left(P(\delta, \delta_{SE}) - P_o(\delta)\right)^2 + \left(Q(\delta, \delta_{SE}) - Q_o(\delta)\right)^2 = \left(\frac{V_T V_{SE\,max}}{X_L}\right)^2 \tag{20.86}$$

Figure 20.32 shows a series of loci of the reactive power *Q* demanded at the receiving bus versus the transmitted real power *P* as a function of the series voltage magnitude V_{SE} and phase angle δ_{SE} at three different power angles δ, namely, δ = 0°, 45°, and 90°, with $V_S = V_T = V$, $V^2/X_L = 1$, and $V_T V_{SE\,max}/X_L = \frac{1}{2}$. Figure 20.32 shows that the UPFC can independently control real and reactive power flow at any transmission angle.

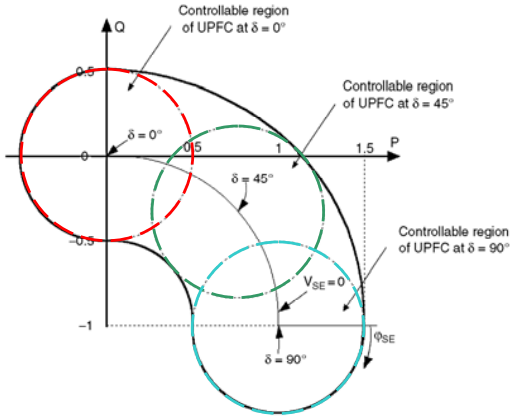


Figure 20.32. *Unified power flow controller – UPFC, P-Q relationship for a two-bus system for three power angles, δ = 0°, 45°, and 90°, with $V_S = V_T = V$, $V^2/X_L = 1$, and $V_T V_{SE\,max}/X_L = \frac{1}{2}$.*

If bidirectional transmission line power flow control is required, a shunt compensator is needed at the opposite line end to the UPFC. Therein lies the overlooked fundamental conceptual limitation of the UPFC. Ideally, shunt compensation is most effective at the line reactance midpoint, while series compensation is most effective at a transmission line end. With the UPFC, both forms of compensation, shunt STATCOM and series DVR, occur at the same single point of connection.

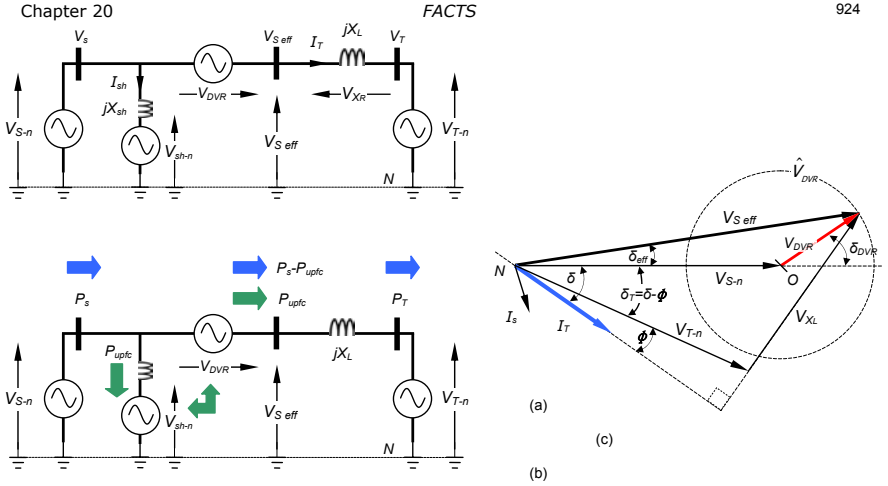


Figure 20.33. *Unified power flow controller - UPFC: (a) single line diagram of the UPFC; (b) UPFC power flow diagram; and (c) phasor diagram for system voltages and line current, I_T.*

With the aid of the phasor diagram in figure 20.33c, the following power equations can be derived. The source grid V_S delivered powers are

$$P_S = V_{S-n} I_T \cos\delta$$
$$Q_S = V_{S-n} I_T \sin\delta \tag{20.87}$$

The UPFC powers, from the phasor diagram in figure 20.33c are

$$P_{UPFC} = V_{DVR} I_T \cos\left(\delta_{DVR} - \delta\right)$$
$$Q_{UPFC} = V_{DVR} I_T \sin\left(\delta_{DVR} - \delta\right) \tag{20.88}$$

The line inductance VAr is

$$Q_{XR} = Q_{S\,eff} - Q_T = \frac{V_{XL}^2}{X_L} = I_T^2 X_L \tag{20.89}$$

Alternatively, using the phasor diagram in figure 20.33c, the terminal grid V_T received powers are

$$P_T = V_{T-n} I_T \cos\left(\phi - \delta\right)$$
$$Q_T = V_{T-n} I_T \sin\left(\phi - \delta\right) \tag{20.90}$$

From equation (20.88) four modes of UPFC control can be deduced:

- If $\delta_{DVR} = \delta$, no reactive power flow is contributed from or controlled by the series DVR, while maximum active power is contributed.
- If δ_{DVR} is at quadrature to δ, the DVR acts as a phase shifter and controls the active power but the reactive power is at a maximum.
- If δ_{DVR} is at quadrature to the line current I_T, the active power flow is controlled with the DVR acting as a controllable series reactive element.
- For other δ_{DVR} the UPFC becomes a combined phase shifter and variable series reactive compensator.

An important feature of the UPFC is that any energy for compensation at 50/60Hz and/or for harmonic compensation, is drawn from the network as sinusoidal current. This is unlike when the energy for the dc-link capacitor is provided via a rectifier feed from the ac network, where the rectification process itself can produce substantial harmonic currents in the network.

20.10 Combined Active and Passive Filters

The basic static synchronous compensators (shunt - STATCOM and series - DVR) can be used simultaneously for both 50/60Hz fundamental power quality improvement and control as well as for line harmonic filtering, by injecting current or voltage, as appropriate, at the PCC. In the harmonic filtering mode, the compensators basically inject anti-phase current and voltage harmonics. In order to do so, the PWM frequency of the compensator inverter must be at least twice that of the highest frequency harmonic to be cancelled.

20.10.1 - Current compensation – shunt filtering

As shown in figure 20.34a, the static synchronous shunt compensator can be used to shunt inject equal but opposite magnitude harmonic compensating currents such that

$$I_S = I_{shunt} + I_L \qquad (20.91)$$

The load current I_L is non-linear, as with rectification for highly inductive loads. The compensator shunt injects a current I_{shunt} such that the supply current I_s is a pure sinusoid at the fundamental frequency. The sending voltage source V_L sees the transmission system as a purely resistive load, if STATCOM normal VAr compensation is also operational.
The STATCOM output is second order *L-C* low pass filtered to prevent PWM carrier components from being injected into the ac system. A second order *L-C* high-pass shunt line filter is normally incorporate to cater for current frequency components at the modulation frequency and beyond the shunt compensator's bandwidth. The high cut-off frequency, well beyond the power frequency, results in reduced size, as well as reduced possibility of resonant effects.

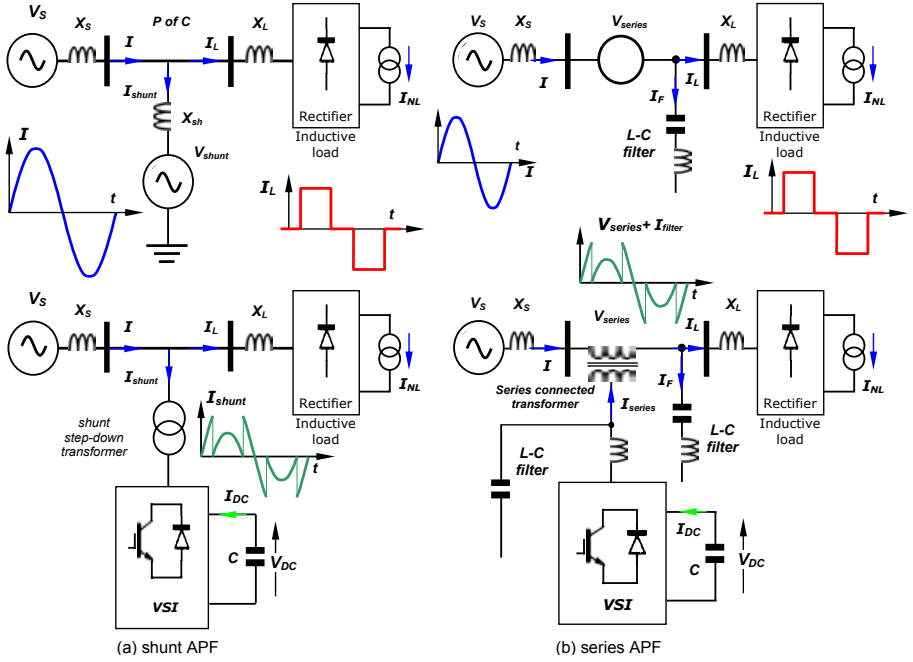


Figure 20.34. *Combined active and passive filters:*
(a) transformer voltage matched shunt APF and (b) transformer voltage matched series APF.

20.10.2 - Voltage compensation – series filtering

As shown in figure 20.34b, the static synchronous series compensator can be used to series injects equal but opposite magnitude harmonic compensating voltages on the line such that

$$V_S = V_{series} + V_L \qquad (20.92)$$

The load current I_L and voltage V_L are both non-linear, since the non-linear current associated with the rectification of highly inductive loads produces non-sinusoidal voltages across the series line inductance, normally around the peaks and troughs of the three-phase sine-waves. The compensator series injects a voltage V_{series} such that the sinusoidal supply voltage V_s delivers a more sinusoidal current into the transmission line. Since the loads still draws a non-linear current, passive notch-shunt and high-pass shunt second order *L-C* filtering are needed to provide a bypass path for the current harmonics.
The series compensator output is second order *L-C* low pass filtered to prevent PWM carrier components from being injected into the ac system.

20.10.3 - Hybrid Arrangements

STATCOM-based hybrid arrangements can be used for both voltage regulation and load-compensation. STATCOMs are usually combined with SVCs or passive harmonics filters. While both provide improvement of compensation capabilities, the former are often used for voltage regulation, while the latter are utilised for load compensation. Additionally, in hybrid topologies the rated power of STATCOM constitutes a part of a hybrid controller's rated power thus they allow the installation costs to be reduced. Shown in the parts of figure 20.35 are general topologies and *V–I* characteristics of STATCOM SVC hybrid arrangements. The parallel connected SVC part extends the current operating region of STATCOM. The combination of STATCOM and TSC (Figure 20.35a) extends the operating region towards the generation of reactive power (capacitive region). This property is important in practice, because it is often necessary in distribution systems to compensate inductive-type loads to provide terminal-voltage regulation. Extension of *V–I* characteristics of STATCOM towards absorption of reactive power (inductive region) is possible by paralleling it with TSR (figure 20.35b). The symmetrical extension of the *V–I* characteristics is provided by the hybrid arrangement shown in Figure 20.35c. In addition to improving *V–I* characteristics, a hybrid arrangement can be used to optimize losses, cost, and performance for a particular application.

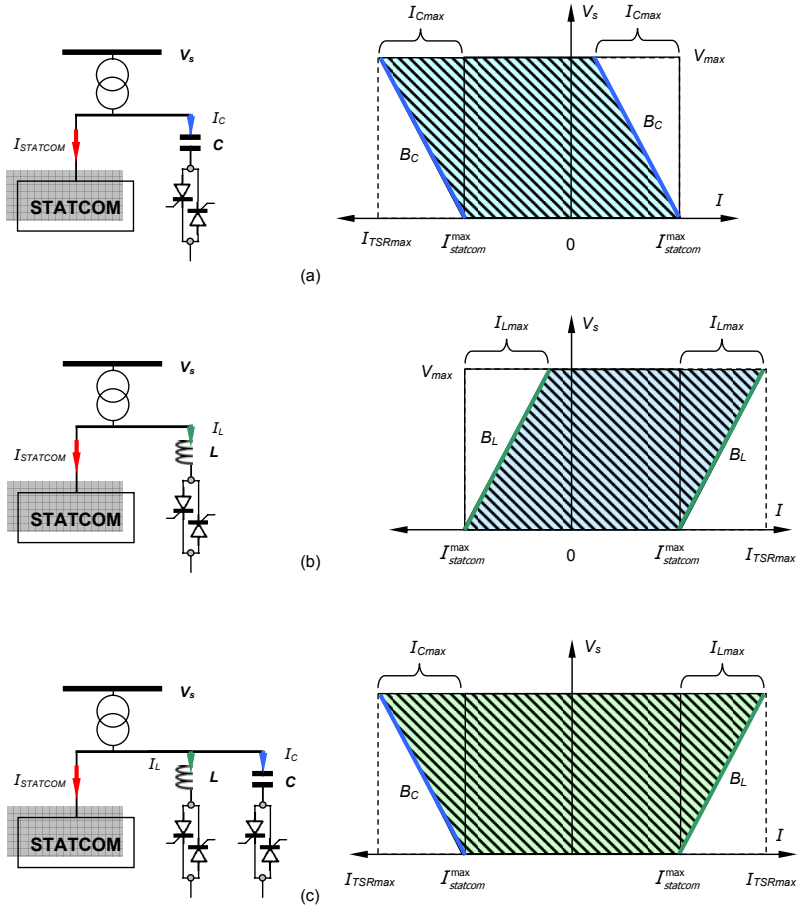


Figure 20.35. *Combined STATCOM and thyristor and static controller and hybrid I-V characteristics:*
(a) STATCOM and TSC; (b) STATCOM and TSR; and (c) STATCOM plus TSR and TSC.

20.10.4 - Active and passive combination filtering

All effective active filtering relies on the addition of passive filtering, even if only to filter compensator inverter pwm outputs.
Semiconductor voltage ratings usually prevent the direct coupling of the compensator inverter to the ac grid. At 50/60Hz, transformer coupling provides a simple and efficient interface method. But for active filtering application, the coupling transformer must have sufficient bandwidth to transmit the necessary compensating harmonic components. Normal 0.3mm silicon steel laminated transformer cores produce transformers suitable for compensation of the 5th and 7th harmonics, but attenuation at the 12th and 13th harmonics results in the inverter dc-link voltage being ineffectively utilised. Special steels (higher silicon) and thinner laminations (0.1mm and 0.05mm) cater for higher frequency operation but as well as being more costly, maximum flux density levels are decreased and core losses are increased. High permeability, amorphous metal-based soft magnetic materials offer modest high frequency losses with high flux densities properties, but are expensive. Indirect filtering methods involving the normal 0.3mm 50/60Hz steels may therefore be preferred.

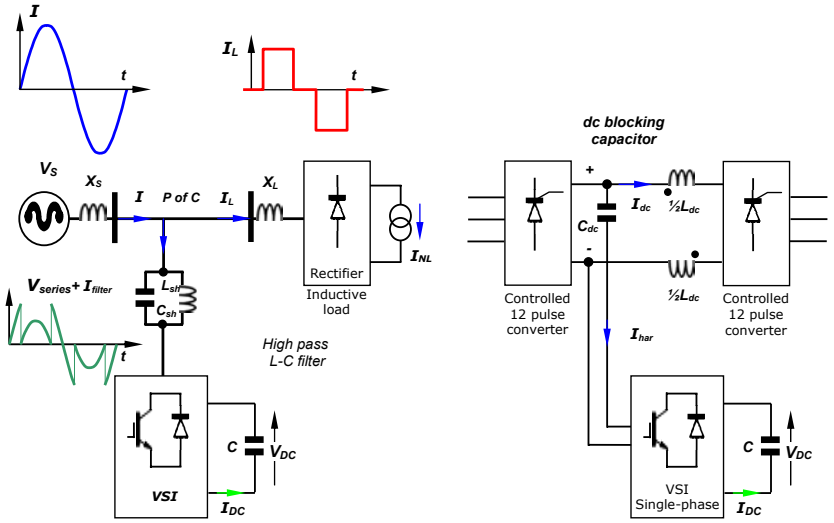


Figure 20.36. *Combined transformerless active and passive power filters: (a) 50/60Hz ac decoupled shunt APF method and (b) dc decoupled shunt APF approach.*

Figure 20.36 shows indirect filter coupling methods suitable for dc and ac lines where the active filter is dc or 50/60Hz decoupled to the transmission system. The series filter (L_{sh}/C_{sh} and C_{dc}) supports the system voltage while the inverter experiences only its own low dc-link voltage. The method is effective for ac but has limitations:

- the passive decoupling filter characteristics drift in time, namely the notch frequency and *Q*;
- the large size and weight of the filter inductor, being based on 50/60Hz design concepts; and
- the inverter is only capable of harmonic compensation without VAr compensation since the 50/60Hz decoupling filter blocks any transfer at the 50/60Hz line transmission frequency. For VAr compensation, the inverter fundamental output voltage must be of a similar magnitude as the line voltage, hence the use of a voltage-matching transformer.

These ac limitations are not relevant to dc-link filtering since dc decoupling capacitor aging characteristics do not affect the block frequency, being dc (VAr compensation is not relevant to dc systems). A single-phase inverter bridge is used for dc-link harmonic compensation. Although the dc-link current harmonics at 12*n* can be cancelled (in a symmetrically triggered 12-pulse system), cancellation of the 11th and 13th order (12*n*±1) ac side current harmonics is more problematic since the rectifying process does not necessarily ensure harmonic current flow in the correct rectifier leg. This is problematic with 12-pulse (and >12) converters.

In the case of ac transmission, transformerless series power filtering is possible since lower voltages are normally involved, but usually a separate isolated single-phase inverter is needed in each phase. The inverter default mode is to operate with all (upper or lower) switches on so that the series compensator is seen as a short circuit.

20.11 Summary of Compensator Comparison and Features

FACTS devices enhance high-voltage ac-grids by:

- increase of power transfer without adding new transmission lines
- transmission cost minimization
- steady-state and dynamic voltage control
- reactive power control of dynamic loads
- active damping of power oscillations
- increase of reliability under system contingencies
- improvement of system stability and voltage quality high flexibility for embedding of various energy sources

A shunt compensator acts like a controllable current source and can draw or inject reactive leading or lagging current at the point of connection.
Objectives of dynamic shunt compensation are

- steady state and transient voltage control
- Reactive power control of transient loads
- Damping of active power oscillations
- Increase of system stability

A series compensator is a driving voltage at the line reactance midpoint, hence is more effective than a shunt compensator for controlling current and power flow, and for damping oscillations. It can only supply or absorb reactive power. When used as a phase angle controller, at the sending or receiving ends, a real power source is required.
Objectives of dynamic series compensation are

- Reduction of load dependent voltage drops
- Reduction of system transfer impedance
- Reduction of transmission angle
- Increase of system stability
- Load flow control to specific power branches
- Damping of active power oscillations

Two points to bear in mind when transformer coupling compensation FACTS type devices.

- The transformer core must be able to transmit at the highest harmonic frequency.
- Series coupling into a dc link imposes a dc bias current, hence flux in the coupling core.

20.12 Summary of the Advantages of AC Transmission over DC Transmission

The general advantages of ac transmission, over dc transmission, are

- No costs associated with ac-dc-ac conversion equipment
- Transformer (and autotransformer) voltage matching
- Reactive power and harmonics readily compensated
- Not restricted to only point-to-point connection, as is HVDC (there is one exception)
- Established system control methods
- No ac transformer dc voltage stressing due to asymmetrical phase control alignment, and no I^2R and core losses due to high harmonic currents
- ac switch gear and breakers, (and particularly vacuum circuit breakers up to 33kV) are very effective - compared with the difficulties in breaking dc current
- Lower current harmonics

Reading list

Mohan, N., *Power Electronics,* 3rd Edition, Wiley International, 2003.

Acha, E., *et al*., *Power Electronic Control in Electrical Systems*, Newes, 2002.

CHAPTER 21

Inverter Grid Connection for Embedded Generation

Distributed Generation (DG, or embedded generation) is a back-up electric power generating unit at or near the consumer premises that primarily is used by the energy user to provide emergency power when grid-connected power is unavailable. Installation of the back-up unit close to the demand centre avoids the cost of transmitting the power and the associated transmission losses. Back-up generating units are currently defined as distributed generation to differentiate from the traditional centralized power generation model. Although the centralized power generation model is economical and a reliable source of energy production, the lack of significant increase in new build generating capacity or even in expanding existing ones to meet the needs of current demand, presents a challenge to the electrical power industry, needing a solution.

21.1 Distributed generation

A typical DG energy conversion system is comprises two main energy converting stages. The first stage is the prime fuel converting block in which the prime fuel internal energy is converted into mechanical energy, as in the case of internal combustion engines. The second stage converts the mechanical energy into electrical power using an electromechanical energy conversion device such as synchronous alternator or induction generator, which produces AC power.

Another way of converting prime fuel source into electrical energy is through a chemical or photosynthesis conversion process. Fuel cells and photovoltaic solar energy converter are examples that produce DC power. The interfacing unit is essential to convert the produced DC source into harmonized constant voltage and frequency AC power source. A DC to AC power electronic inverter system is used as the interfacing unit. The inverter must produce high quality AC power with a voltage waveform of limited supply frequency fluctuation and low THD at the point of common coupling (PCC), in accordance with the appropriate standard. The inverter must be capable of preventing the DG from islanding (anti-islanding capability) on the hosting grid. Islanding is a condition occurring when a generator or an inverter and a portion of the grid system separates from the remainder of the large distribution system and continues to operate in an energized state. Islanding may pose a safety threat or cause equipment problems; therefore cannot be permitted without system coordination.
The inverter output produced must comply with hosting grid electricity voltage and frequency standards. A coupling transformer is needed to interface the DG generator with the grid to match the distribution voltage level at the point of connection. Only when it is safe and synchronised conditions exist is the DG interconnected with the permission and coordination of the grid operator.

Another configuration normally adopted for supplying power to sensitive electrical load demand is to use DG in conjunction with a UPS unit. A UPS system normally incorporates an energy storage medium such as batteries to enable power supply continuity as well as improve power quality and reduce the influence of voltage surges, spikes and swells which could cause loss of production.

Once the interconnection is established the hosting utility assumes responsibility of DG operation and contribution and treats it as part of its generation system.

Current DG/distribution network interconnected systems practice is to revert the distribution network to its original configuration (radial or meshed distribution system) with all interconnected DG units de-energized whenever an unexpected disturbance occurs in the system. Since most distribution systems comprise radial feeders, this leads to the supply discontinuation for all the down-line customers. In this way the DG contribution is restricted to the hosting utility demand and conditions.

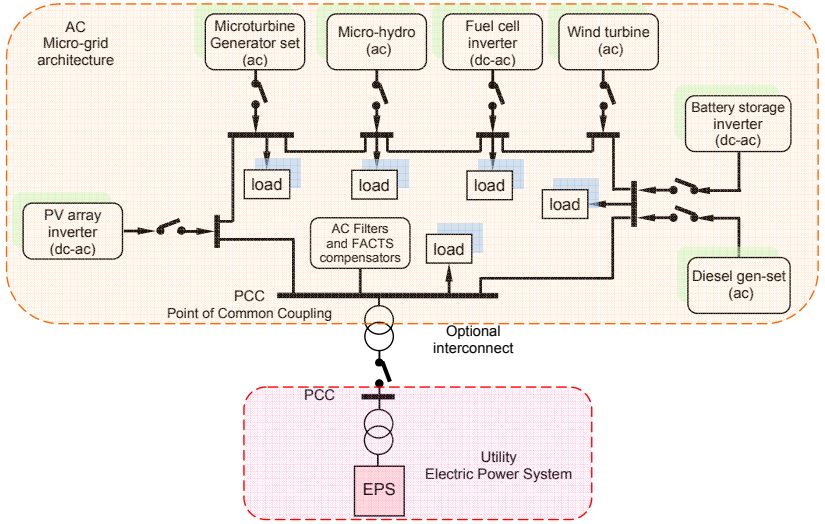


Figure 21.1. *AC microgrid architectural structure.*

21.1.1 DG Possibilities

DG is attractive for the following opportunities:

- DG can be fuelled by locally available renewable and alternative mix of fuel sources to meet current energy demand. Renewable sources are wind and solar, while alternative fuels are those produced from waste products or biomass and can be in gas, liquid or solid form. Greater independency from importing petroleum fuel can be achieved by incorporating DG that are powered by various fuel sources;
- DG can support the projected increase in demand, without investment in the expansion of existing distribution network, by installing the DG close to a new load centre;
- Installing DG within the industrial/commercial premises avoids negotiating land use and the need for rights-of-way for electric transmission and distribution, thereby minimizing further capital investment;
- DG can be used in reducing intermittent and peak supply burdens on utilities grid by injecting power as required by the controller;
- DG has the ability to support the existing energy supply when needed and in a short time (black start) without incurring capital cost;
- DG penetration in the energy market will create overall competitive pricing of energy. The current DG generation rate ($/kWh) is competitive with the centralized generation system as more efficient fuel energy conversion units such as fuel cells and micro turbines are continuously improved and diversified;
- DG can decrease the vulnerability of the electric distribution system to external threats and hidden undetected faults that may cause blackout by feeding power to the sensitive infrastructure;

- DG is flexible, being capability of being configured to operate in a stand-by mode, isolated mode, or sharing the load through integration with the electric grid.

Using DG that is fuelled by various prime alternative fuel sources will reduced fossil fuel consumption hence reduce CO_2 emissions.

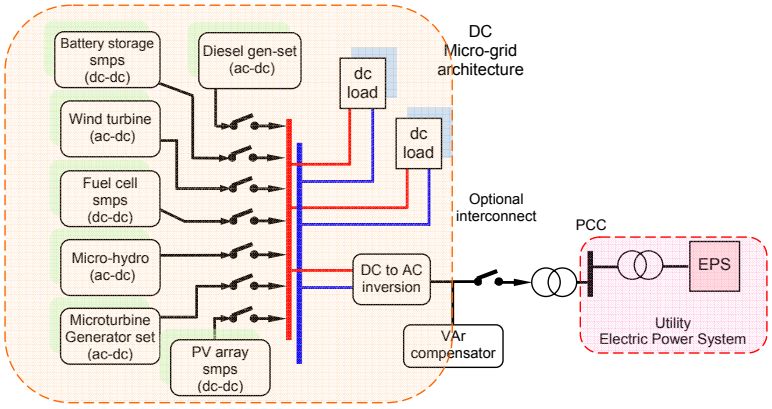


Figure 21.2. *DC microgrid architectural structure.*

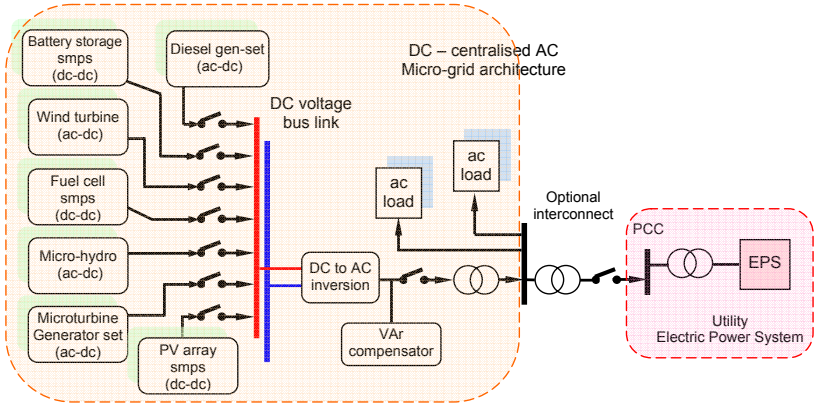


Figure 21.3. *DC – centralised ac microgrid architectural structure.*

21.1.2 Integration and Interconnection Requirements

Key elements for the reliability of distributed generation power systems are the performance of the electrical switchgear, interconnection, controls, and communication features. The main components of interconnection according to the protection functions they perform are categorized as follows:

Synchronization. Automatic sensing of the voltage and frequency can achieve fast interconnection to the hosting grid.

Islanding. Islanding protection is a mandatory feature that the hosting grid requires from the DG operator. Islanding on part of the hosting network could jeopardize the maintenance crew safety and causes malfunction of nearby coordinated protection units. Relays are normally used to provide protection at both the grid and the DG end of the connection. DG inverters should incorporated built-in features to disconnect from the hosting grid once anti-islanding conditions are violated.

Voltage and frequency tolerance. For high quality power injection, both the voltage and frequency margins should not exceed the grid tolerance specification. Both voltage and frequency detection is part of the anti-islanding protection control.

DC injection level. Under abnormal operating conditions, grid tie inverters could inject low level DC current into the hosting grid. Similarly, transformer-less grid-tie inverters may inject DC current into the grid. It is part of the inverter feedback loop to detect the presence of the DC component and adjust the triggering sequence to the switching devices to remedy the situation. A coupling transformer could be used to isolate the DC current from flowing to the AC side. A low cost solution is to incorporate a DC detection device to disconnect the inverter in case of severe DC level injection.

Grounding. Protective grounding is mainly designed to protect the operator. Grounding could also contribute to reducing the magnitude of transient over-voltages and lightning protection. Grounding components must be capable of carrying the maximum available fault current and withstanding a second strike within a few cycles after the first. Grounding cables must be connected directly to the equipment. No impedance, circuit breaker or measuring devices, *etc.*, are permitted between the grounding cable and the equipment.

Metering and monitoring. Monitored parameters can include current, voltage, real and reactive power, oil temperature, vibration, *etc*. Metered parameters also include power output, which may be used for billing that requires utility-grade metering accuracy.

Dispatch, communication, and control. These integration and communication components interface the DG units with the utility. Their functions include:

- Regional load management, work order management, and billing services;
- Distribution automation;
- Feeder switching;
- Short circuit analysis;
- Voltage profile calculations and trouble calls management.

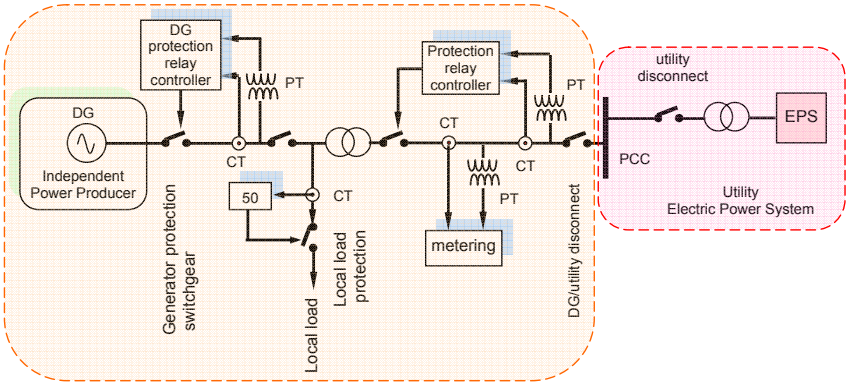


Figure 21.4. *DG-utility interconnection protection requirements.*

A typical interconnection line schematic with the protection elements between the DG and the hosting grid is shown in Figure 21.4. Typical minimum DG/utilities interconnection protection relay requirements are:

- The DG protective switchgear should include an over/under voltage trip function, an over/under frequency trip function, and a means for disconnecting the DG from the utility when a protective function initiates a trip;
- The DG and associated protective switchgear must not contribute to the formation of an unintended island;
- DG switchgear must be equipped with automatic means to prevent DB reconnection with the utility distribution system unless the distribution system service voltage and frequency is of specified settings and is stable for specified time, typically 60s;
- Circuit breakers or other interrupting devices at the PCC must be capable of interrupting the maximum available fault current.

21.2 Interfacing conversion methods

A common feature of embedded generation interfacing is voltage translation and stabilisation using the boost converter concept. The boost converter is used since its input current can be continuous thus drawing continuous energy from the energy source, for maximum source energy extraction efficiency.

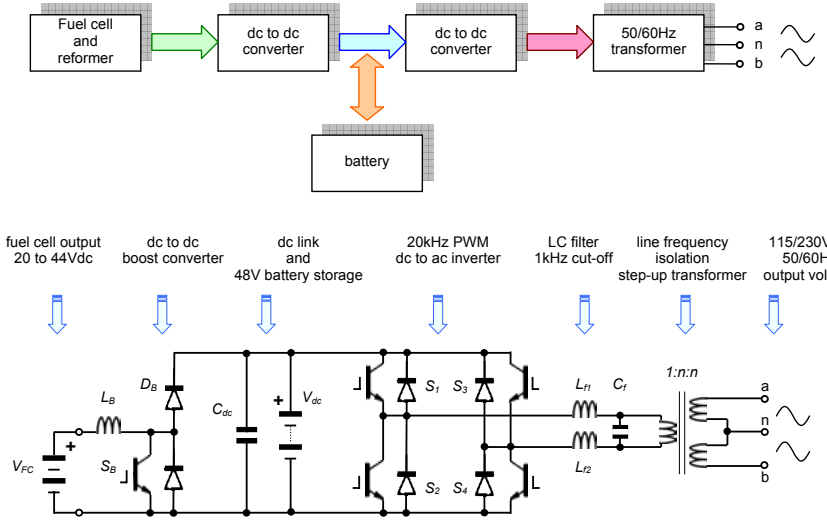


Figure 21.5. *Dc to line frequency power conditioner: (a) block diagram and (b) circuit topology.*

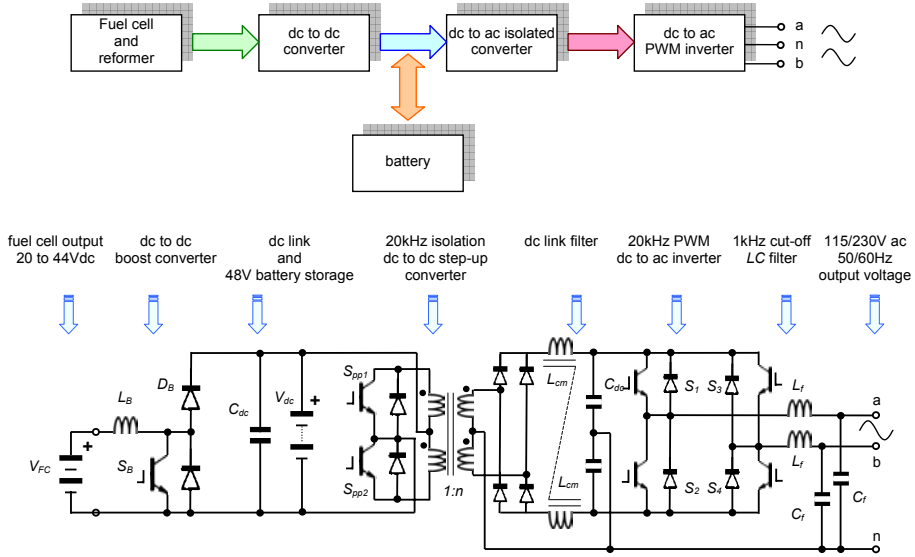


Figure 21.6. *Dc to line frequency power conditioner with high frequency isolation: (a) block diagram and (b) circuit topology.*

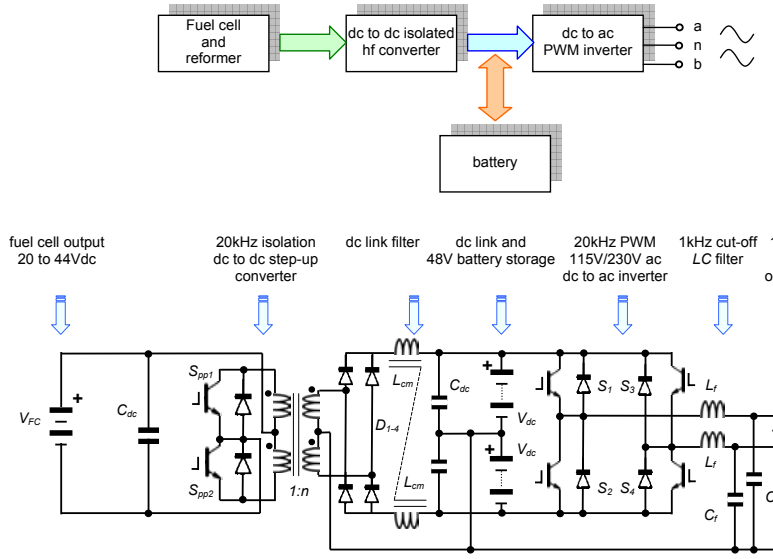


Figure 21.7. *Dc to line frequency power conditioner with high frequency isolation and two power conversion stages: (a) block diagram and (b) circuit topology.*

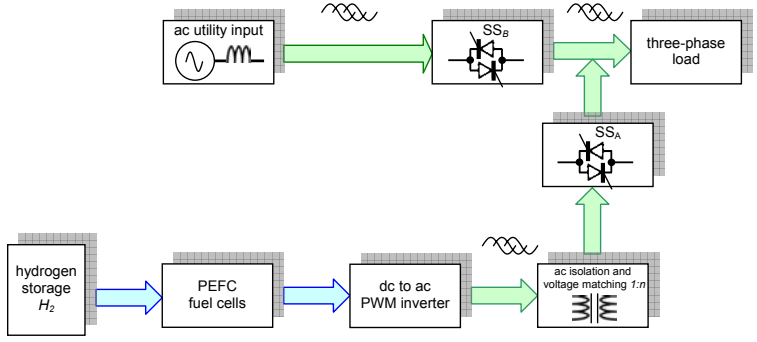


Figure 21.8. *Modular fuel cell power conversion system supplying a three phase load in parallel with the grid, via solid state circuit breaker for source isolation and islanding.*

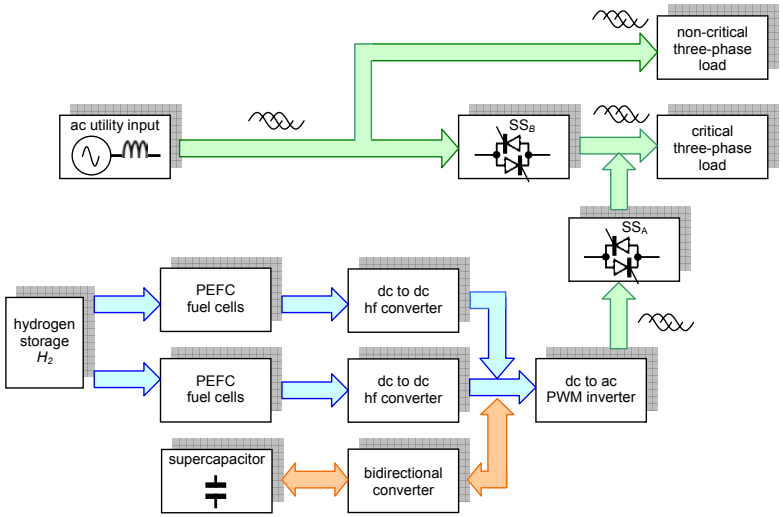


Figure 21.9. *Modular fuel cell power conversion system for grid connection with supply backup for critical loads.*

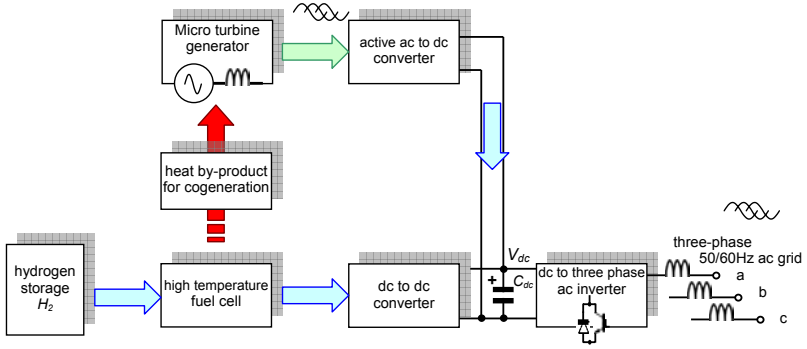


Figure 21.10. *Modular fuel cell power conversion system for three-phase grid connection, with cogeneration on to a common dc link.*

Reading list

blank

CHAPTER 22

Energy Sources and Storage -

Primary Sources

The progressive proliferation of embedded and distributed generation with renewable energy sources has spurred research into alternative energy sources and storage methods. Although this chapter is mainly concerned with primary sources, namely fuel cells and photovoltaic cells, their energy and power density capabilities can only be put into context by considering conventional energy sources, specifically, the hydrocarbons.

In electrical terms, primary and secondary energy sources are defined as follows.

Primary source is not a reversible energy source. During energy discharging, the original states are permanently changed as electrical energy is released until the original energy reactants, or any one of them, are depleted. A primary cell can be used only once.

Secondary source is reversible and the original states can be reconstituted by the application of an electrical potential which injects energy into the source. A secondary cell can source and sink energy many times.

22.1 Hydrocarbon attributes

Table 22.1 shows that the energy density associated with the hydrocarbons dwarfs all other common sources such as batteries and supercapacitors. The table highlights why petrol is so firmly entrenched, while the limitations of hydrogen are made apparent. Although hydrogen offers 3 times more energy than petrol for a giving reactant weight, its volume per kg is grossly in excess of that of petrol. One mole of hydrogen H_2 occupies 22.4 litres and weight just 2 grams. Thus the energy per unit volume for petrol is 2500 times more than that for hydrogen. Table 10.2a presents the basic properties of hydrogen.
Pressurising hydrogen mitigates the volume limitation somewhat, but liquefaction is at an impractical temperature of 20K, with a density of 0.07g/cm^3 (only helium has a lower boiling point). Pressure helps minimally since the boiling point rises to only 43K at 13 bar, with minimal temperature increase for higher pressures. On the other hand, propane has a boiling point of -42°C, but can be liquefied at 21°C and 13bar, with a gain in energy density.
The expansion ratio for hydrogen is 1:850, that is, at atmospheric temperature and pressure, hydrogen gas occupies 850 times the volume as in its liquid state. The ratio is 1:240 at 250 bar and atmospheric temperature, but cannot approach the liquid ratio at much higher pressures, as shown in Table 22.2b. Basically, hydrogen has a poor molecular packing density. A cubic metre of water contains 111kg of hydrogen, whereas a cubic metre of liquid hydrogen contains only 71kg of hydrogen. In fact, water (111kg) contains more kg/m^3 of hydrogen that methanol (100kg) and a similar weight of hydrogen as in heptane (113kg). The energy stored, in Joules, is given by mass times specific energy (calorific value).

Thus, in summary, the specific energy of pure hydrogen (the energy per kilogram) is higher than any other fuel at better than 120 MJ/kg. However, its energy density (energy per m^3) is very low, that is, it is difficult to get a large mass of hydrogen into a small volume, as shown in table 22.2. Cryo-adsorption of hydrogen into graphite at low temperature and pressures up to 5MPa, or the use of metal hydrides, provide higher hydrogen storage densities, as shown in table 22.2b.
Table 22.1, in conjunction with figure 22.1, show that both batteries and super capacitors fall significantly short of the energy that can be released from hydrocarbons, on an equivalent volume and weight basis.

Table 22.1: Energy properties of hydrocarbons and alternative energy sources

Fuel Type	Fuel	CO_2	Calorific Value				Density	
		kg/kWh	MJ/kg	kWh/kg	MJ/Litre	MJ/m^3	kg/m^3	Litres/kg
solid	charcoal		30	10	-	6250	208	-
	coal	0.32	25	8	-	37500	1500	-
	wood		16	5.4	-	8000	400-700	-
	dung cake		7	2.3	-	5000	700	-
liquid	kerosene	0.27	48	15	35	44000	820	0.73
	petrol		50	16	48	37000	751	1.36
	diesel		45	15	39	38200	850	1.19
	ethanol		30	10	30	23700	789	1.00
gas	bio gas		38	12		46	1.2	
	butane(LPG)	0.23	50	16	29	30000	600	1.72
	butane gas		50	16	29	125	2.5	
	methane		55	18	0.9kJ/mol	39	0.668	
	hydrogen		150	49		13	0.0838	
battery	Lithium-ion	0		0.2				
	Lead acid	0		0.04				
capacitor	double layer	0		0.02				
fuel cell	hydrogen	0		26				
PV cell		0						
Flywheel	FES	0		0.13				
fusion	U235	0		2.5×10^7				

Table 22.2a: Hydrogen; high gravimetric energy density but low volumetric energy density

Hydrogen property	unit	value
relative atomic mass	u	1.00794
density	kg/Nm^3	0.0899
specific HHV energy	MJ/kg	142.0
specific LHV energy	MJ/kg	120.0
HHV energy density	MJ/Nm^3	11.7
LHV energy density	MJ/Nm^3	9.9
boiling point	K (°C)	20.268 (–252.88)
melting point	K (°C)	14.025 (–259.13)
critical point	K (°C)	33.250 (–239.90)

Table 22.2b: Hydrogen; high gravimetric energy density but low volumetric energy density

Energy carrier	Form of hydrogen Storage 120 MJ/kg		Energy density by weight	Energy density by volume
			kWh/kg	kWh/l
Hydrogen	gas	20 MPa	33.3	0.53
	gas	24.8 MPa	33.3	0.64
	gas	30 MPa	33.3	0.75
	liquid	-253°C	33.3	2.36
	metal hydride		0.58	3.18
fusion	U_{235}		2.5×10^{7}	4.7×10^{9}

Comparisons
The Ragone plot in figure 22.1 shows the energy storage and power handling capacity of some alternative storage techniques. Energy and power densities, in steady-state, are related by *Energy = Power × time*.

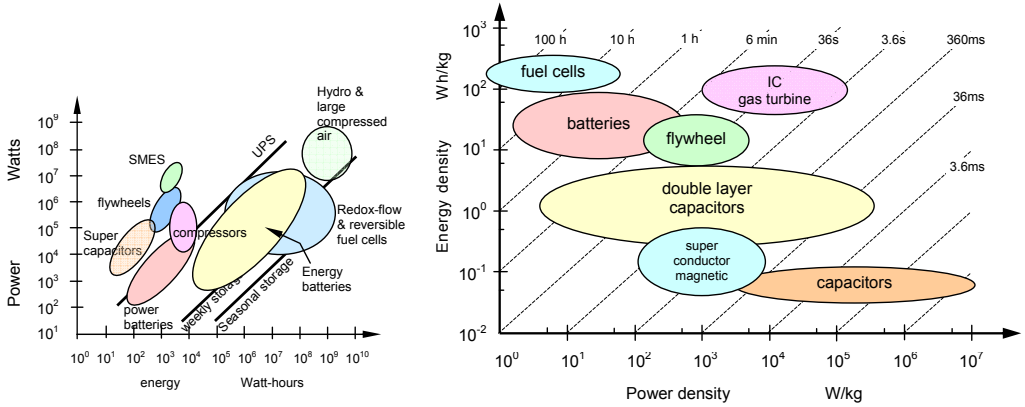


Figure 22.1. *Gravimetric energy and power densities of common energy sources.*

22.2 The fuel cell

The fuel cell is similar to a battery, in function and appearance. It produces electricity directly, using chemicals.
A fuel cell is a solid-state electrochemical device that consists of two electrodes, an anode and cathode, sandwiched around an electrolyte, with a catalyst membrane between each electrode and the electrolyte, which enhances ionization of the fuel molecules. The oxidant oxygen, usually from air, passes over the cathode electrode and typically hydrogen fuel over the anode, generating electricity, and water and heat by-products. An ion-conducting membrane separates the anode and cathode, allowing the reaction to take place without affecting the electrodes.
The fuel cell relies on a basic oxidation/reduction reaction, as with a battery, but the reaction takes place on the fuel rather than on the electrodes. As long as fuel is supplied and oxidized old fuel is disposed of, the cell will continue to generate energy, both electrical power and heat. Since conversion of the fuel to energy takes place via an electrochemical process, not combustion, the process is clean - no CO_2, quiet and highly efficient - two to three times more efficient than fuel combustion.

How the fuel cell operates
The fuel cell process can be divided into three stages; two of these stages involve the chemical reactions at each electrode and the third is ion conduction through the electrolyte.

(i) At the hydrogen electrode - anode
The reaction at the anode involves the release of electrons from the hydrogen fuel that will then be conducted by the electrode to the electrical load. The hydrogen arrives at the anode as a diatomic gas $2H_2$ where each adsorbed hydrogen molecule ionizes into four hydrogen protons, H^+ and four electrons, e^-. The rate of this process is increased with the aid of a catalyst. Because the chemical reaction at this electrode produces positive ions, it is the anode. The negatively charged electrons are then forced to flow from the conductive electrode (anode) to an external electrical load before re-entering the fuel cell at the cathode (that is, electrons diffuse naturally from a high concentration to a low concentration of electrons). However, the hydrogen ions may or may not conduct through the electrolyte since this depends on the pH and type of electrolyte used, meaning the hydrogen ions may have to temporarily remain at the anode in a receptive state.

(ii) At the oxygen electrode - cathode
Meanwhile, oxygen molecules O_2 are diffusing through to the catalytic surface of the cathode electrode which facilitates the separation of the adsorbed oxygen molecule (oxygen bonds are broken) into oxygen atoms which are held momentarily in a receptive state on the active catalyst. The entering electrons diverted externally from the anode bypassing the electrolyte, the oxidant, and cathode electrode together causes another reaction to occur where negative ions and products are produced. If H^+ ions are the free moving ions (because of the acid electrolyte used), they are attracted to the negative ions generated at the cathode and conduct through the acid electrolyte and product $2H_2O$, along with heat. If however, H^+ ions are not able to travel through the electrolyte, then the cathode produced free moving negative ions move through the alkali electrolyte to combine with the H^+ ions to complete the process at the anode.
(iii) Through the electrolyte medium
The electronically-insulated (does not conduct free electrons) electrolyte serves as the physical barrier preventing the fuel and oxidant gas streams from directly mixing allowing, only the appropriate ions to move freely through the layer. This requires that one of the reactants must be able to form the ionized specie needed to complete the process and form the primary by-product, water.

Fuel cells that use hydrogen can be thought of as an electrochemical devices that perform the reverse of electrolysis, where passing an electric current through water splits it up into hydrogen and oxygen. In the fuel cell, hydrogen and oxygen are joined together to produce water and electricity.
Although the majority of fuel cells use hydrogen as the fuel, some fuel cells work off methane, and a few use liquid fuels such as methanol.

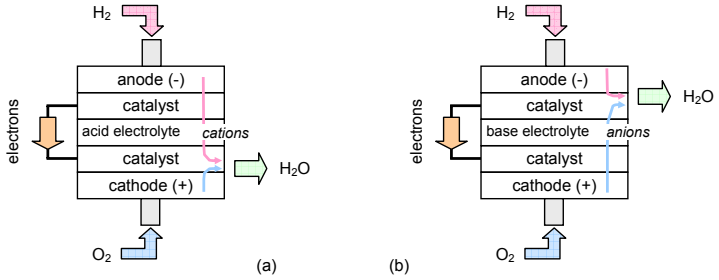


Figure 22.2. *Pictorial representation of a fuel cell with electrolyte conduction of: (a) cations and (b) anions.*

Two basic ion operating mechanism are possible. The first involves cations passing through an acid electrolytic membrane from anode to cathode, while the second mechanism involves anions passing through an alkali membrane in the opposite direction, namely from the cathode to the anode. The ion conducting membrane is non-conducting to electrons.

- ***Cation conduction***: Pressurised hydrogen H_2 gas fuel is fed into the anode of the fuel cell. The pressure forces the H_2 through the catalyst. When an H_2 molecule comes in contact with the platinum catalyst, it splits into two H^+ ions (protons) and two electrons e^-, which take different paths to the cathode. The catalyst increases the rate of this splitting process. The electrolyte membrane does not pass electrons. The electrons conduct externally from the anode, creating a current that can be utilized in an external circuit, then return to the cathode-side of the fuel cell where they are reunited with the hydrogen plus oxygen ions forming a molecule of water, as illustrated in figure 22.2a.

Meanwhile, on the cathode-side of the fuel cell, oxygen gas O_2 (or air) is being forced through the catalyst, where it is encouraged to form two oxygen atoms. Each of these atoms has a strong negative charge. This negative charge attracts two H^+ ions, protons, at the anode side through the membrane, where they combine with an oxygen atom and two of the electrons from the external circuit to form a water molecule H_2O, plus heat energy.

- ***Anion conduction***: The operating principle of an anion conducting cell is illustrated in Figure 22.2b. Oxygen supplied at the cathode (air electrode) reacts with incoming electrons from the external circuit to form oxide ions, which migrate to the anode (fuel electrode) through the anion conducting electrolyte. At the anode, anions combine with hydrogen ions (and/or carbon monoxide) in the fuel to form water (and/or carbon dioxide), liberating electrons. Electrons (electricity) flow from the anode through the external circuit to the cathode.

In principle, fuel cells can operate on many reactant combinations but most fuel cells actually operate on a narrow range of fuels in combination with oxygen from the air. As considered in section 22.9, fuels range from pure hydrogen gas, liquid alcohols and other liquid or gaseous hydrocarbons to metals and solid carbon. For the fuel cell types restricted to operating only on pure hydrogen, it is necessary to involve systems that generate hydrogen. Electrolysis and fuel reforming technologies generate hydrogen, but storage at useful energy densities is problematic. Several types of fuel cell can be operated directly on readily available fuels such as methanol, ethanol and natural gas, thereby dispensing with the infrastructure issues of hydrogen. Traditionally these fuels come predominantly from fossil sources, but they are also increasingly available from renewable bio-sources. Since the fuel cell relies on electro-chemistry and not combustion, emissions are smaller than emissions from the cleanest fuel combustion processes.
Fuel cells can be made in a wide range of sizes, from a few watts to MW. The potential power generated by a fuel cell stack depends on the number and size of the individual fuel cells that comprise the stack and the surface area of the membrane. Since the single fuel cell produces only about 0.7V, many separate fuel cells must be series connected to form a high-voltage fuel-cell stack. Bipolar plates are used to judiciously connect one fuel cell to another and are subjected to both oxidizing and reducing conditions and potentials. An issue with bipolar plates is stability. Metallic bipolar plates corrode, and the by-products of corrosion (iron and chromium ions) decrease the effectiveness of fuel cell membranes and electrodes. Low-temperature fuel cells use lightweight metals, graphite, and carbon plus high temperature thermoset composites, as bipolar plate material.

The fuel cell offers a unique combination of benefits. In addition to low or zero emissions, benefits include high efficiency and reliability, multi-fuel capability, siting flexibility, durability, scalability and ease of maintenance. Fuel cells operate silently, so they reduce noise pollution as well as air pollution and the waste heat from a fuel cell can be used to provide domestic hot water or space heating.

There are several different types of fuel cells, each using a different chemistry. Fuel cells are usually classified by their operating temperature and the type of electrolyte used.

The materials for the cell components are selected based on suitable electrical conducting properties required of these components to perform their intended cell functions:

- adequate chemical and structural stability at high temperatures encountered during cell operation as well as during cell fabrication;
- minimal reactivity and interdiffusion among different components; and
- matching thermal expansion among different components.

22.3 Materials and cell design

Like the battery, the fuel cell has few component parts although the materials may be sophisticated, involving rare earth transitional metals, high temperature composite ceramics, cermets, etc. The basic fuel cell component parts are:

- Electrodes;
- Catalysts;
- Electrolyte;
- Interconnects; and their
- Stack construction.

22.3.1 Electrodes

Fuel cell electrodes serve three functions:

- To ensure a stable interface between the reactant gas and the electrolyte;
- To catalyze the electrode reactions; and
- To conduct electrons from or to the reaction sites

An electrode forms part of the three-phase boundary, where the electrolyte, electrode, and gas all come together.

i. Cathode

The ***cathode*** is the positive electrode of the fuel cell because it is the electrode where negative ions are produced. It has etched channels that distribute the oxygen to the surface of the catalyst. It also conducts the electrons received from the external circuit to the catalyst, where they can recombine with the hydrogen ions and oxygen to form water. An integral part of the cathode is the metallic interconnect, a bipolar plate, which forms an integral part of the anode of the adjacent cell, when cells are series connected to give higher voltage output.

The oxidant gas is air or oxygen at the cathode, and the electrochemical reduction of oxygen requires a series of elementary reactions and involves the receipt of multiple electrons. The cathode must meet the requirements of:

- high catalytic activity and high surface area for oxygen molecule dissociation and oxygen reduction;
- high electronic conductivity;
- chemical and dimensional stability in environments encountered during cell fabrication and cell operation;
- thermal expansion match with other cell components;
- compatibility and minimum reactivity with the electrolyte and the interconnection; and
- must have a stable, porous microstructure so that gaseous oxygen can readily diffuse through the cathode to the cathode/electrolyte interface.

ii. Anode

The ***anode*** is the negative electrode of the fuel cell because it is the electrode where positive ions are produced. It conducts away the electrons that are released from the hydrogen molecules so that they can be used in an external electrical circuit. An integral part of the anode is the metallic interconnect, a bipolar plate, which form parts of the cathode of an adjacent series cell. It has channels etched into it that disperse the hydrogen gas uniformly over the surface of the catalyst.

The properties of the anode must include:

- an excellent catalyst (in the case of high temperature fuel cells) with a large surface area for the oxidation of fuel (hydrogen, carbon monoxide);
- stable in the reducing environment of the fuel;
- electronically conducting;
- have sufficient porosity to allow the transport of the fuel to and the transport of the products of fuel oxidation away from the electrolyte/anode interface where the fuel oxidation reaction takes place;
- matching thermal expansion coefficient with that of the electrolyte and interconnect;
- integrity of porosity for gas permeation;
- chemical stability with the electrolyte and interconnect;
- applicability to use with versatile fuels and impurities; and
- cost effective is a commercialization factor.

22.3.2 Catalyst

The ***catalyst*** facilitates and speeds up the rate of the ionization reaction of oxygen at the cathode and hydrogen at the anode. It is usually made of platinum for low temperature cells and nickel at higher temperatures. The catalyst is rough and porous so that the maximum surface area can be exposed to the hydrogen or oxygen. The catalyst faces the electrolyte and must be able to conduct electrons to/from the electrode.

The effectiveness of the catalyst is paramount to fuel cell operation. Platinum is sufficiently reactive in bonding H and O intermediates as is required to facilitate the electrode process, then effectively releases the intermediate to form the final product. For example, the anode process requires Pt sites to bond H atoms when the H_2 molecule reacts, and these Pt sites release the H atom as $H^+ + e^-$, as shown in the two equations:

$$H_2 + 2Pt \rightarrow 2Pt\text{–}H$$

$$2Pt\text{–}H \rightarrow 2Pt + 2H^+ + 2e^-$$

Platinum can be oxidized which is a problem because it compete with hydrogen oxidation, reducing the half-cell potential:

$$Pt + H_2O \leftrightarrow PtOH + H^+ + e^-$$

$$Pt + H_2O \leftrightarrow PtO + 2H^+ + 2e^-$$

$$½O_2 + 2H^+ + 2e^- \rightarrow H_2O$$

Hydrocarbon fuels, like CH_3OH, tend to poison the anode platinum catalyst with COH and CO species. The poisoning can be prevented by the addition of Ruthenium, Ru.

$$Ru + H_2O \rightarrow RuOH + H^+ + e^-$$

$$Pt_xCO + RuOH \rightarrow CO_2 + H^+ + e^- + xPt + Ru$$

Catalytic reduction action also occurs on the cathode:

$$O_2 + 2Pt \rightarrow 2Pt - O$$

$$2Pt - O \rightarrow 2Pt + 2O^{2-} - 4e^-$$

The reaction effectiveness is dependant on temperature and platinum surface area. At lower temperature reactions, platinum particles are used, 2 nanometres in diameter, resulting in a large Pt area that is accessible to the gas molecules. At elevated temperatures, over 400C, transition metals, Ni and oxides NiO, become effective, low cost catalysts, that can also act as the electrode.

$$NiO + CO_2 \rightarrow Ni^{2+} + CO_3^{2-}$$

At higher again temperatures no catalyst is necessary.

Reduction of oxygen at the cathode is less efficient than the catalytic action on hydrogen at the anode. In lower temperature fuel cell, various nanoparticle platinum alloys (AuPt and Pt_3Ni-III) are used to increase the O_2 reduction reaction activity. The tolerance to poisoning and the affects of an acidic environment are factors to be overcome with semi-formed alloy catalysts.

22.3.3 Electrolyte

In terms of chemistries, there are two main types of fuel cells, depending on the pH of the electrolyte. Common to all electrolytes is that they must not conduct free electrons.

- Acid electrolyte which allows H^+ hydrogen ion (proton) migration and free movement through the electrolyte from the anode to the cathode, producing water at the cathode, while conversely
- Alkaline electrolyte which allows anion migration, for example, OH^- hydroxyl ion migration through the electrolyte from the cathode to the anode, producing excess water at the anode.

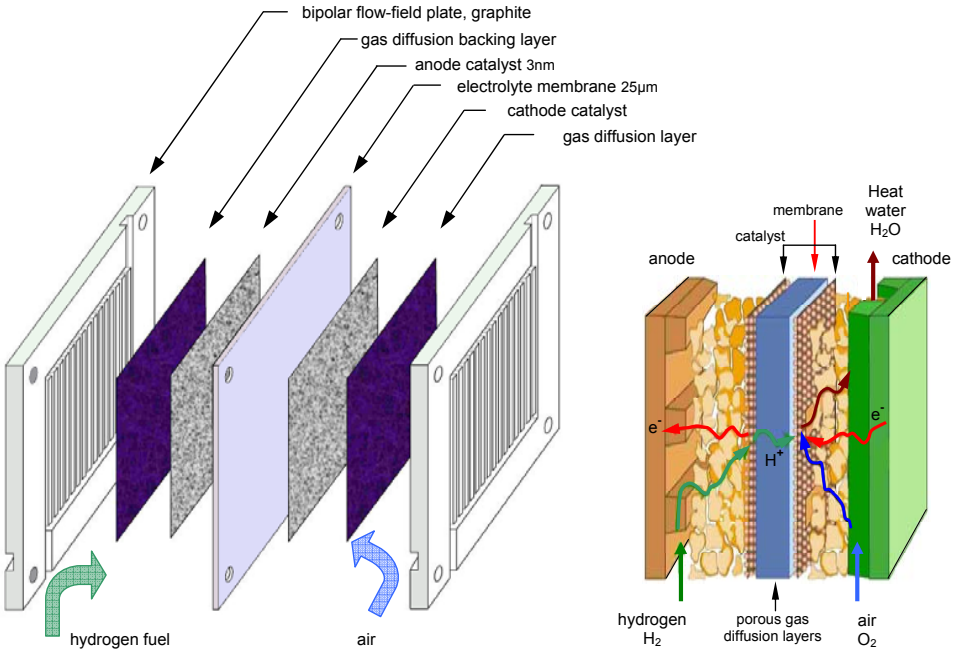


Figure 22.3. *The parts of a fuel cell/Membrane/electrode assembly with backing layers. Enlarged cross-section of a membrane/electrode assembly showing structural details.*

The ***electrolyte*** is the either a solid (< 200°C) or a liquid (> 200°C), depending on the operating temperature. The electrolyte can be acidic or alkali and must pass ions but importantly, the electrolyte does not conduct free electrons and does not react with reactant ions. There are several criteria that the electrolyte has to meet. The electrolyte must be:

- Dense and leak tight;
- Stable in reducing and oxidising environments;
- A good ionic conductor at operating temperatures;
- Non-electron conductor;
- Thin to reduce ionic resistance;
- Extended in area for maximum current capacity;
- Thermal shock resistant; and
- Economically processable.

22.3.4 Interconnect

Since a single cell only produces a voltage of less than 1V and power around 1W/cm^2, many cells are electrically connected together in a cell stack to obtain higher voltage and power. To connect multiple cells together, an interconnection is used in stacks. The requirements of the interconnection are the most severe of all cell components and include:

- High, nearly 100 percent electronic conductivity, not only through the bulk material but also in in-situ-formed oxide scales;
- strong adhesion between the as-formed oxide scale and the underlying alloy substrate;
- resistance to corrosion and surface stability in both oxidizing and reducing atmospheres at the cell operating temperature since it is exposed to duel atmospheres, with air (or oxygen) on the cathode side and fuel on the anode side;
- low permeability for oxygen and hydrogen to minimize direct combination of oxidant and fuel during cell operation;
- chemical compatibility with other materials in contact with the interconnect, such as seals and cell materials;
- mechanical reliability and durability at the cell operating temperature;
- a thermal expansion coefficient close to that of other stack components, mainly the cathode and the electrolyte (particularly for stacks using a rigid seal design); and
- non-reactivity chemical compatibility with other cell materials.

Suitable metallic alloy interconnects offer advantages such as improved manufacturability, significantly lower raw material and fabrication costs, and higher electrical and thermal conductivity.

22.3.5 Stack design

In the case of planar cell stacks, an effective seal must be provided to isolate air from the fuel. The seal must have a thermal expansion match to the fuel cell components, must be electrically insulating and must be chemically stable under the operational conditions of the stack. Also, the seal should:

- exhibit no deleterious interfacial reactions with other cell components;
- be stable under both the high temperature oxidizing and reducing operational conditions;
- be created at a low enough temperature to avoid damaging cell components (under 850°C for some materials); and
- should not migrate or flow from the designated sealing region during sealing or cell operation.

In addition, the sealing system should be able to withstand thermal cycling between the cell operation temperature and room temperature. Sealing is more of a problem with cell which operate at higher temperatures, where different sealing approaches include rigid, bonded seals (for example, glass-ceramics and brazes), compliant seals (for example viscous glasses) and compressive seals (for example, mica-based composites); multiple sealants may also be used in any given stack design between different cell components.

22.4 Fuel Cell Chemistries

The two chemical ion mechanisms involving anions and cations, will be discussed in terms of:

- an acid electrolyte and transfers cations, protons, H^+, through the electrolyte, producing water at the cathode and
- an alkaline electrolyte and transfers anions (OH^-, CO_3^{2-}, O^{2-}) through the electrolyte, producing water at the anode.

22.4.1 *Proton H^+ Cation Conducting Electrolyte*: e.g. PEMFC, DMFC, PAFC

The Acidic electrolytic Fuel Cell utilises one of the simplest reactions of any fuel cell. In low temperature PEMFC and DMFC Fuel Cells, the electrolyte is a flexible acidic membrane that when saturated offers a number of desirable features:

- Highly chemical resistant;
- Mechanically strong, when only 50μm thick;
- Acidic;
- Highly water absorptive;
- Good proton conductor, H^+ cation, when saturated, but not flooded;
- Electrically insulating, high resistance to electrons; and
- Non-diffusive the gas molecules, no H_2 and O_2 crossover.

Chemistry

In both acid electrolyte and solid polymer fuel cells the electrolyte conducts mobile H^+ ions, protons, generated at the anode. It is these H^+ ions that are key to the reactions within the cell.
At the anode, hydrogen gas is readily ionised, producing free electrons and H^+ ions:

$$2H_{2(g)} \rightarrow 4H^+ + 4e^- \qquad E^o_{½\,cell} = 0.0\text{V} \qquad (22.1)$$

The electrons from the anode conduct through the external circuit connected to the fuel cell, to the cathode. The H^+ ions migrate or permeate through the electrolyte, and also reach the cathode.
At the cathode the H^+ ions and electrons react with the oxygen atoms, producing water. At the cathode the oxygen undergoes a two-step indirect reduction reaction. The stable H_2O_2 intermediate is undesirable as it lowers the cell voltage and H_2O_2 attacks and corrodes the carbonaceous electrodes commonly used in lower temperature cells.

$$\begin{aligned} O_2 + 2H^+ + 2e^- &= H_2O_2 \\ H_2O_2 + 2H^+ + 2e^- &= 2H_2O \end{aligned} \qquad (22.2)$$

Overall cathode half reaction:

$$O_{2\,(g)} + 4H^+ + 4e^- \rightarrow 2H_2O \qquad E^o_{½\,cell} = 1.23\text{V} \qquad (22.3)$$

This movement of ions through the electrolyte, and movement of electrons through the external circuit is illustrated in the diagram in figure 22.4a.

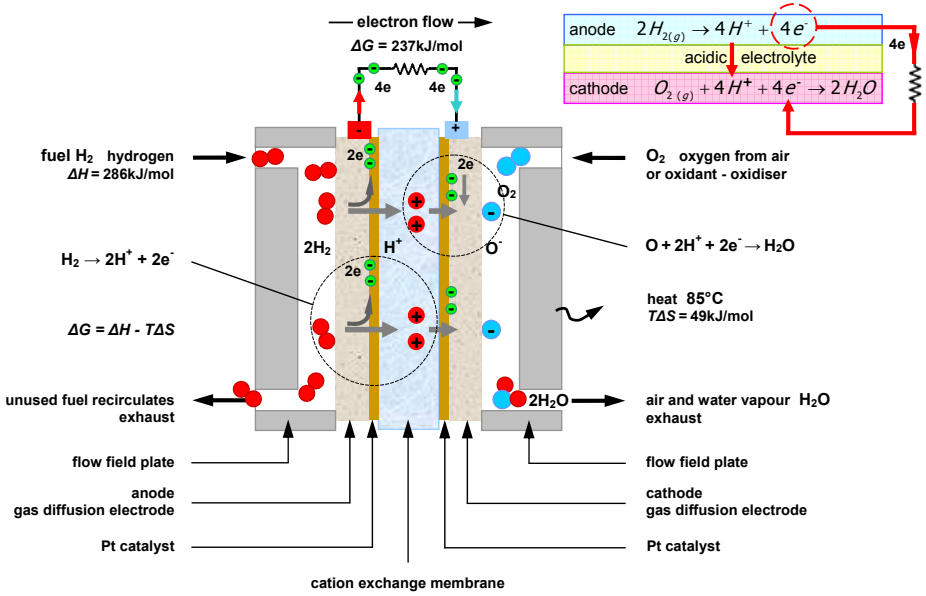


Figure 22.4. *The chemistry mechanisms of an acidic electrolyte fuel cell.*

The net fuel cell reaction, which is exothermic, is the algebraic summation of the half-cell reactions:

$$2H_{2(g)} + O_{2(g)} \rightarrow 2H_2O + energy \qquad E^o_{cell} = 1.23\text{V} \qquad (22.4)$$

The chemistry of the fuel cell gives E^o_{cell} = 1.23 - 0 = 1.23V. Standard tables give the enthalpy as 285.15kJ/mol and the number of electrons per H_2, as n = 2. The dissociation kinetics of O_2 and H_2 improve with increased temperature but the cell potential decreases with temperature.

Positive ions are produced at the anode and negative ions are produced at the cathode. The water molecules are always produced at the cathode, the production of which exceeds that which may be required (if any) at the anode.
The H^+ cations, as found in the Polymer Exchange Membrane Fuel Cell - PEMFC, Direct Methanol Fuel Cell – DMFC, and the Phosphoric Acid Fuel Cell - PAFC all behave the same within the fuel cell. That is, equations (22.1) to (22.4) are applicable to the three cation fuel cells presented.

22.4.2 *Anion (OH^-, CO_3^{2-}, O^{2-}) Conducting Electrolyte*: e.g. AFC, MCFC, SOFC

The important feature of an alkali is that there is an excess of anions, X^- ions, generated at the cathode, and these are key to the reactions within the alkali electrolyte fuel cell. An electric current is produced as a result the half-cell reaction at each electrode.
At the anode, hydrogen gas reacts with the X^- ions, anions which have migrated through the electrolyte, producing water, and releasing electrons. The reaction is:

$$2H_{2(g)} + 4X^-_{(aq)} \rightarrow 4H_20_{(l)} + 4e^- \qquad (22.5)$$

The electrons leave the fuel cell at the anode, passing through the external electrical circuit connected to the fuel cell, and reach the cathode. At the cathode the entering electrons react with the incoming oxygen, and water, producing more X^- ions to replenish those used at the anode.

$$O_{2(g)} + 2H_2O + 4e^- \rightarrow 4X^-_{(aq)} \qquad (22.6)$$

The X^- ions move through the electrolyte, and the electrons move round the external circuit. The complementary movement of ions and electrons is illustrated in the diagram in figure 22.5a. The net reaction, which is exothermic, is the algebraic half-cell reaction summation:

$$2H_{2(g)} + O_{2(g)} \rightarrow 2H_2O + energy \qquad (22.7)$$

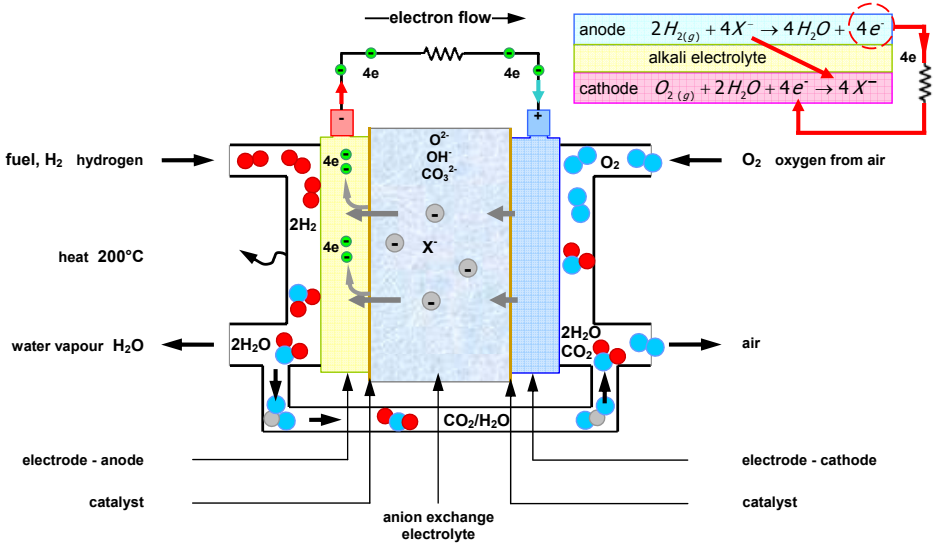


Figure 22.5. *The chemistry mechanisms of an alkaline electrolyte fuel cell.*

Positive ions are produced at the anode and negative ions are produced at the cathode. Water molecules are always produced at the anode, and at a faster rate than those (if any) being consumed at the cathode.
Anions, such as OH^-, CO_3^{2-} and O^{2-} as found in the Alkaline Fuel Cell - AFC, Molten Carbonate Fuel Cell - MCFC, and Solid Oxide Fuel Cell - SOFC, behave similarly within the fuel cell, as given by equations (22.5) to (22.7).

22.5 Six main Fuel Cells

Some of the main types of fuel cells, with increasing operating temperature, include:

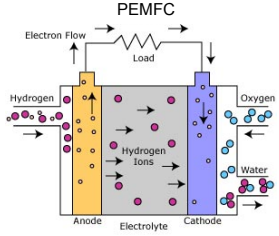
- **Low-temperature Fuel Cell types** (<200°C)
  - *Polymer exchange membrane fuel cell* (PEMFC) - H^+
  - *Alkaline fuel cell* (AFC) - OH^-
  - *Direct-methanol fuel cell* (DMFC) - H^+
- **High-temperature Fuel Cell types** (>200°C)
  - *Phosphoric-acid fuel cell* (PAFC) - H^+
  - *Molten-carbonate fuel cell* (MCFC) - CO_3^{2-}
  - *Solid oxide fuel cell* (SOFC) - O^{2-}

22.6 Low-temperature Fuel Cell types

22.6.1 *Polymer exchange membrane fuel cell* (PEMFC) (Proton Exchange Membrane FC)

The PEMFC has a high power density and a relatively low operating temperature (ranging from 60 to 80°C). The low operating temperature means that the fuel cell can warm-up and begin to generate electricity quickly, resulting in less wear on system components, leading to better durability. Cells do not use corrosive fluids like some fuel cells. However, it requires a noble-metal catalyst (typically platinum) to separate the hydrogen's electrons and protons, adding to system cost. The platinum catalyst is also hyper-sensitive to CO poisoning, making it necessary to employ an external reactor to reduce CO in the fuel gas if the hydrogen is derived from an alcohol or hydrocarbon fuel. Platinum/ruthenium catalysts are more resistant to CO. The slow kinetics of oxygen at the cathode also necessitates the use of a catalyst.
Higher-density liquid fuels such as methanol, ethanol, natural gas, liquefied petroleum gas, and gasoline can be used for fuel, but the system must have an external fuel processor to reform the methanol to hydrogen. Reforming releases minimal carbon dioxide.
The PEMFC has a high power density, high efficiency, and can adapt to varying demands. It promises a high conversion efficiency of over 60% and an energy density of 120 W/Kg, with 2kW/l and 2W/cm^2.
The PEM fuel cell is currently favoured for electric motor vehicles. Other applications include air conditioning and domestic sized electronic equipment.
Key features are:

- Fast start-up, rapid transient power response, high power density, compact
- H^+ via polymer (ionomer) membrane electrolyte, Pt, 50°C to 150°C, H_2 at < 5 atm.
- Sensitive to impurities, CO, sulphur species, and NH_3
- Anode and cathode gas sealing is simpler with a solid electrolyte therefore cheaper, no orientation restrictions, and low corrosion leads to a longer cell and stack life.
- Portable applications, 50 to 500kW – automotive, small stationary auxiliary power units
- Overall reaction ΔH = -286.0kJ/mol, with n = 2.



Anode Reactions: $2H_{2(g)} \xrightarrow{Pt} 4H^+ + 4e^-$ $E^o_{½cell}$ = 0.0V

Cathode Reactions: $O_{2(g)} + 4H^+ + 4e^- \xrightarrow{Pt} 2H_2O$ $E^o_{½cell}$ = 1.229V

Overall Cell Reactions: $2H_{2(g)} + O_{2(g)} \rightarrow 2H_2O$ E^o_{cell} = 1.229V

As shown in figure 22.3 for the PEMFC case, which is typical of the basic physical structure, there are four basic elements of the fuel cell:

- The ***anode*** is the negative electrode and conducts the electrons that are released from the hydrogen molecules so that they can be used in an external electrical circuit. The carbon or metal interconnect, which makes electrical contact to the anode carbon sheet, has channels etched into it that disperse the hydrogen gas uniformly over the back surface of the catalyst layer.

- The ***cathode*** is the positive electrode and also has etched channels that distribute the oxygen to the back surface of the catalyst layer. It also conducts the electrons received from the external circuit to the catalyst, where they can recombine with the hydrogen ions and oxygen to form water. The carbon or metal interconnect, like the anode interconnect, makes electrical contact to the cathode carbon sheet.
- The ***electrolyte*** is the proton exchange membrane. This specially treated organic material is a thin, 50µm, flexible plastic (poly-perfluorosulphonic acid) film, which only conducts positively charged ions when it is saturated with water. The membrane must be hydrated (by absorbing water) in order to function and remain stable. Importantly, the membrane blocks electrons. Being a solid, membrane helps minimise gas crossover.
- The ***catalyst*** is usually made of platinum nanoparticles thinly coated onto carbon paper or cloth, which is the effective anode or cathode. The catalyst is rough and porous so that the maximum platinum surface area can be exposed to the hydrogen or oxygen. The platinum-coated side of the carbon sheet, the catalyst, faces the polymer membrane.

The Backing Layers
The fuel cell anode and cathode assemblies are designed to maximize the current that can be obtained from an electrode and catalyst assembly. The backing layers on which the catalyst is deposited, are usually made of a porous carbon paper or carbon cloth, typically 100 to 300 µm thick, that can conduct the electrons and effectively diffuse reactant gas ions created and reacted in the platinum catalyst. The platinum catalyst particles are deposited 50µm thick on one side of each backing layer; the side that faces the electrolytic.

Diffusion refers to the flow of gas molecules from a region of high concentration, the outer side of the backing layer where the gas is flowing by in the flow fields, to a region of low concentration, the inner side of the backing layer next to the catalyst layer where the gas is consumed by the reaction. The porous structure of the backing layers allows the gas to spread out as it diffuses so that when it penetrates the backing, the gas will be in contact with the entire surface area of the catalyzed membrane.

The backing layers also assist in water management during the operation of the fuel cell. An effective backing material allows the correct amount of water vapour to reach the membrane/electrode assembly to keep the membrane humidified. The backing material also allows the liquid water produced at the cathode to leave the cell to avoid flooding. The backing layers are often wet-proofed with Teflon to ensure that most of the pores in the carbon cloth (or carbon paper) do not become saturated with water, which would prevent the rapid gas diffusion necessary for a good rate of reaction, from occurring at the electrodes.

The solid organic polymer electrolyte membrane
The polymer electrolyte membrane is essentially PTFE containing a fraction of pendant sulphonic acid groups. (Nomenclature: 'sulphonic acid group' usually refers to the un-dissociated SO_3H group, where as 'sulphonate' refers to the ionised SO_3^- group after the proton has dissociated). The ion containing component is normally given in terms of equivalent weight (that is, number of grams of dry polymer per mole of acidic groups). The useful equivalent weight for Nafion ranges from 800-1500 g/mol.
The length of and the precise nature of the side chains vary between different brands of polymer. Common to all is the PTFE based fluorocarbon 'backbone' of the polymer that has several desirable properties:

- PTFE is hydrophobic - this means the hydrophilic sulphonate groups are effectively repelled by the chains and cluster together;
- PTFE is extremely resilient to chemical attack – the environment within the membrane is hostile and acidic. Hydrocarbon-based polymers would tend to degrade rapidly; and
- PTFE is a thermoplastic with high mechanical strength – meaning very thin membranes can be produced, reducing the thickness of each cell and increasing the power density of the stack.

As shown in figure 22.6, the thin permeable electrolyte sheet of poly-perflourosulphonic acid consists of three distinct regions.

- A Teflon-like fluorocarbon backbone of hundreds of repeating chains -
  $-CF_2 - CF - CF_2 -$
- Side chains connecting the molecular backbone to the third region -
  $O - CF_2 - CF - O - CF_2 - CF_2 -$
- Ion clusters of sulphonic acid ions -
  $SO_3^- H^+$

When the membrane becomes hydrated, the hydrogen ions become mobile and bond to water molecules as they move from one SO_3^- site to another in the acid molecule.

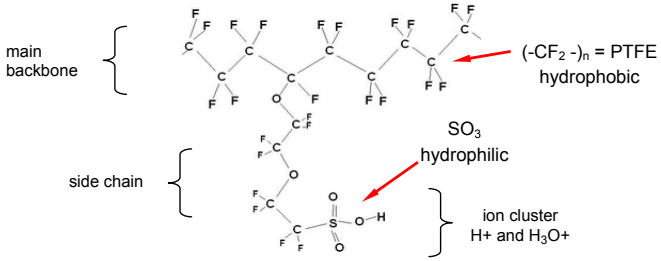
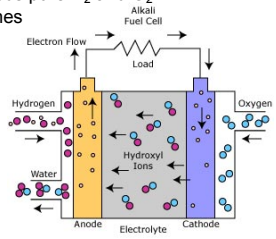


Figure 22.6. *A PTFE fluorocarbon (NAFION) used for the PEM fuel cell electrolyte.*

22.6.2 *Alkaline fuel cell* (AFC)

This is one of the oldest designs for fuel cells, with modest cell efficiencies of nearly 70%. The AFC is susceptible to contamination, so requires pure hydrogen and oxygen. It is also expensive, which is hampering commercialization.
Key features are:

- Fast start-up, but bulky, 10kW to 100kW
- OH^- via liquid or polymer electrolyte, KOH, non-Pt, 80°C to 160°C, gas or liquid fuels
- Susceptible to carbon contamination by K_2CO_3 formation, needs pure H_2 and O_2
- Portable, small stationary/transport, space vehicles, submarines
- Overall reaction ΔH = -286.0kJ/mol, with n = 2.



The anode half-cell reaction involves two stages.

$$2H_2 \xrightarrow{Ni} 4H^+ + 4e^-$$

$$4H^+ + 4OH^- \xrightarrow{Ni} 4H_2O$$

The overall cell reactions are:

Anode Reaction: $2H_{2(g)} + 4OH^- \xrightarrow{Ni} 4H_2O + 4e^-$ $E^o_{½cell}$ = -0.83V

Cathode Reaction: $O_{2(g)} + 2H_2O + 4e^- \xrightarrow{Ni} 4OH^-$ $E^o_{½cell}$ = 0.401V

Overall Net Reaction: $2H_{2(g)} + O_{2(g)} \rightarrow 2H_2O$ E^o_{cell} = 1.231V

The chemistry of this fuel cell gives E^o_{cell} = 0.401 - -0.83 = 1.231V. Standard tables give the enthalpy as 286kJ/mol and the number of electrons per H_2, as n = 2.

These fuel cells use an alkali solution of potassium hydroxide KOH in water as the electrolyte and can use a variety of non-precious metals as a catalyst at the anode and cathode. High-temperature AFCs operate at temperatures between 100°C and 250°C, with newer AFC designs operating at temperatures below 100°C. Concentration of electrolyte, KOH, which is contained in an asbestos matrix, decreases to 50% at lower temperatures. Electro-catalyst can by Ni, Ag, metal oxides, spinels, and noble metals.

Aqueous alkaline solutions do not reject CO_2. Thus a disadvantage of this fuel cell type is that it is readily poisoned by carbon dioxide CO_2, which reacts to produce K_2CO_3 that irreversibly blocks the pores in the cathode. The levels of CO_2 in the air affect cell operation, making it necessary to purify both the hydrogen and oxygen used in the cell. This purification process is costly. This susceptibility to catalyst poisoning affects the cell's lifetime and precludes the use of platinum as a catalyst.

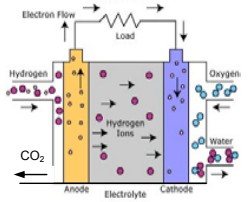
22.6.3 *Direct-methanol fuel cell* (DMFC)

The methanol fuel cell is an optimised PEMFC but is not as efficient. The efficiency is low due to the permeation of neutral methanol and water through the sulphonated organic hydrocarbon polymer membrane (e.g. Nafion), termed crossover. The low-temperature oxidation of methanol to hydrogen ions and carbon dioxide requires more active platinum catalyst, which makes these fuel cells expensive.

Other catalysts are Ni, Ag, metal oxides, spinels and noble metals. The thin proton exchange membrane is covered on both sides with a sparse layer of platinum-based catalyst.
DMFC do not use hydrogen fuel, but more convenient liquid methanol. The anode catalyst draws H_2 from liquid methanol, thus eliminating the need for a fuel reformer. They are less efficient but offer compact and convenient designs suitable for future consumer electronics applications and in automotive areas because the fuel is convenient. One litre of methanol can theoretically provide 5kWh, but 1.7kWh outputs are typical.
Key features are:

- Use with liquid alcohol - methanol canisters, compact, no reforming
- Methanol is liquid from -97°C to 65°C, with ×10 the energy/litre of highly compressed H_2
- H^+ via polymer electrolyte, Pt or alternatives (Ru), 90°C to 100°C
- Cell efficiency of 40% and increases with temperature
- Small portable, battery replacement in mobile phones and laptops, small transport
- Overall reaction ΔH = -726.6kJ/mol, with n = 6.

Direct Methanol



Anode Reaction: $2CH_3OH + 2H_2O \xrightarrow{Pt/Ru} 2CO_{2(g)} + 12H^+ + 12e^-$ $E^o_{½cell}$ = 0.2V

Cathode Reaction: $3O_{2(g)} + 12H^+ + 12e^- \xrightarrow{Pt} 6H_2O$ $E^o_{½cell}$ = 1.23V

Overall Cell Reaction: $2CH_3OH + 3O_{2(g)} \rightarrow 2CO_{2(g)} + 4H_2O$ E^o_{cell} = 1.214V

ote that H_2O is required at the anode, with an excess over the anode requirement being produced at the cathode.

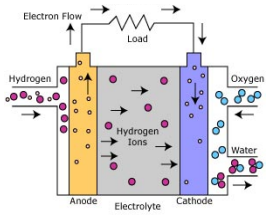
22.7 High-temperature Fuel Cell types (well in excess of 100°C)

22.7.1 *Phosphoric-acid fuel cell* (PAFC)

The liquid phosphoric-acid fuel cell has potential for use in small stationary power-generation systems up to 10MW with 50% cell efficiency and better than 80% efficiency if steam produces cogeneration, CHP. It operates at a higher temperature than polymer exchange membrane fuel cells, so it has a longer warm-up time, restricting its application areas, and it must be continuously operated since H_3PO_4 electrolyte effloresces irreversibly and solidifies at 40°C. A silicon carbide matrix is used to retain the concentrated H_3PO_4 electrolyte and highly dispersed Platinum and Pt alloy anode catalysts particles. CO_2 does not affect the electrolyte or cell performance, which can therefore be operated with reformed fossil fuel. Simple construction, low electrolyte volatility, and long-term stability are additional advantages.
Key features are:

- First commercially available, best suited to large size
- H^+ via liquid or polymer electrolyte, Pt, 150-250°C, with near 100% concentration H_3PO_4
- Carbon paper electrodes Pt coated, Pt unaffected by CO (<1½%) or CO_2 impurities in H_2
- Sulphur must be removed from petrol if used
- Medium-large scale stationary, large mobile applications
- Overall reaction ΔH = -286.0kJ/mol, with n = 2

Phosphoric acid
fuel cell



Anode Reaction: $2H_{2(g)} \xrightarrow{Pt} 4H^+ + 4e^-$ $E^o_{½cell}$ = 0.0V

Cathode Reaction: $O_{2\,(g)} + 4H^+ + 4e^- \xrightarrow{Pt} 2H_2O$ $E^o_{½cell}$ = 1.229V

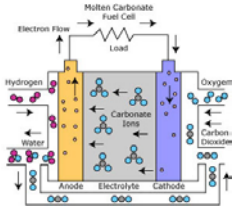
Overall Cell Reaction: $2H_{2(g)} + O_{2(g)} \rightarrow 2H_2O$ E^o_{cell} = 1.229V

22.7.2 *Molten-carbonate fuel cell* (MCFC)

Like the SOFC, this fuel cell is suited to large stationary power generators, up to 100MW. They operate at 600°C, so can generate superheated steam that can be used to generate more power, CHP. Typical cell efficiency is 55%, and 85% with high-pressure steam heat capture. Because of high operating temperatures, reforming to H_2 of natural gas fuel occurs internally and the cells are not prone to CO and CO_2 poisoning. They have a lower operating temperature than solid oxide fuel cells, which means they do not need such exotic materials (sheet stainless steel is used in the construction), which makes the design less expensive.

The electrolyte is molten alkaline carbonate salt mixture (lithium carbonate and potassium carbonate - 62 m/o Li_2CO_3 and 38 m/o K_2CO_3 (62/38 Li/K), or lithium carbonate and sodium carbonate) suspended in a porous, chemically inert ceramic matrix of beta-alumina solid electrolyte. The high temperature results in high ion mobility through the liquid electrolyte. The cathode is supplied with carbon dioxide and the fuel is normally fossil fuel or natural gas.
Key features are:

- CO_3^{2-} via liquid electrolyte,
  non-Pt uses nickel based catalyst, 600-800°C
- Fuel flexibility, longer start-up
- CO_2 and O_2 must be delivered to the cathode
- Long term electrode corrosion by CO_3^{2-} electrolyte
- Wide range of fuels, H_2, natural gas, CO, propane, diesel, etc.
- Utilities, industrial, military applications
- Overall reaction involves $n = 2$.



The reactions using CO and H_2, derived from internal CH_4 reforming conversion (see equations (22.8) to (22.9)), are:

Anode Reaction: principal reaction $2H_{2(g)} + 2CO_3^{2-} \xrightarrow{Ni} 2H_2O + 2CO_{2(g)} + 4e^-$ $E^o_{½cell} = 1.11V$

Cathode Reaction: $O_2 + 2CO_{2(g)} + 4e^- \xrightarrow{NiO} 2CO_3^{2-}$

Overall Cell Reaction: $2H_{2(g)} + O_{2(g)} + 2CO_{2(cathode)} \rightarrow 2H_2O + 2CO_{2(anode)}$ $E^o_{cell} = 1.07V$

Anode Reaction: minor reaction $2CO_3^{2-} + 2CO_{(g)} \xrightarrow{Ni} 4CO_{2(g)} + 4e^-$

Cathode Reaction: $O_{2(g)} + 2CO_{2(g)} + 4e^- \xrightarrow{NiO} 2CO_3^{2-}$

Overall Cell Reaction: $O_{2(g)} + 2CO_{(g)} + 2CO_{2(cathode)} \rightarrow 4CO_{2(anode)}$ $E^o_{cell} = 1.04V$

The cathode reaction is the same for both cases. Note that CO_2 is required at the cathode, with that CO_2 being produced at the anode, with an excess at the anode in the case of CO fuel.

22.7.3 *Solid oxide fuel cell* (SOFC)

These fuel cells are best suited for large-scale stationary power generators for providing electricity to factories or towns. The two construction types are planar and tubular, both having a cell efficiency of typically 60%. Of primary importance is the fact that SOFCs require no liquid electrolyte, with associated corrosion and electrolyte management problems.
Figure 22.7 shows the principal components of the tubular solid oxide fuel cell. Air is channelled into the cell stack and through the centre of each tube. The innermost layer of the tube is the cell cathode and is surrounded by the electrolyte (middle layer). Fuel flows over the outer layer that serves as the anode.
Bundles of tubes are joined together into modules to make a fuel cell stack. The modular nature of fuel cells is a strength, because power plants can add modules as needed, keeping the unused capacity to a minimum.

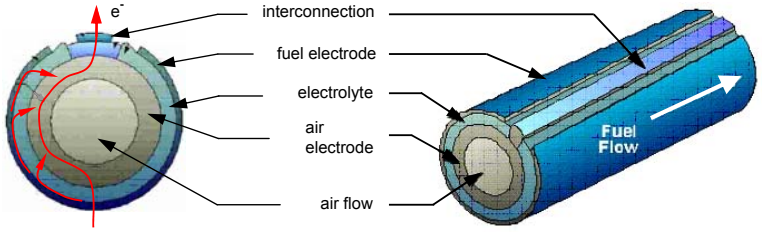


Figure 22.7. *Tubular solid oxide fuel cell construction.*

The SOFC operates at high temperatures, between 700 and 1,000°C, which introduces a reliability problem, because parts of the fuel cell can fail due to repeated on-off thermal cycling. However, solid oxide fuel cells are stable when in continuous use. The high-temperature high-quality wastes for CHP, integrated with secondary gas turbines improve the system efficiency to over 80%.

The high temperature enables them to tolerate relatively impure fuels, such as from coal gasification or gasses from industrial processes, due to internal reforming. Adversely, more expensive and exotic construction materials are required.
To operate at such high temperatures, the electrolyte is a thin, 40µm, solid ceramic material (solid oxide) that is conductive to oxygen ions, O^{2-}. The anode is made form Ni-ZrO_2 cermet, 150µm thick, while the cathode is made from lanthanum strontium manganite. High temperature cell interconnects are made from 100µm conductive ceramics, lanthanum and yttrium chromites doped with Mg, Sr and Ca. Metallic alloys can be used with operating temperatures below 800°C.

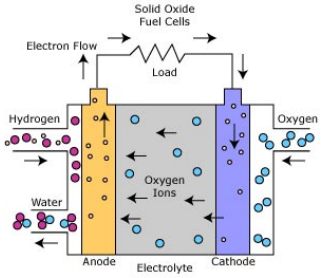
Key features are:

- Fuel flexibility
- CHP increases system efficiency
- O^{2-} via conducting non-porous hard ceramic electrolyte of ZrO_2, no catalyst, 500-1000°C
- Sulphur products must be removed by using a zinc absorbent or active carbon bed
- Thick, electrically conducting nickel cermet anode used to strengthen and support the cells
- Planar construction requires 8 hour warm-up while tubular construction requires minutes
- Small to large scale stationary; portable emerging, mobile applications - auxiliary power units
- Overall H_2 reaction ΔH = -286.0kJ/mol, with $n = 2$.
- Overall CO reaction ΔH = -283.1kJ/mol, with $n = 2$.
- Overall CH_4 reaction ΔH = -890.8kJ/mol, with $n = 8$, reformation and water-gas shift phases.

Anode Reaction: $2H_2 + 2O^{2-} \rightarrow 2H_2O + 4e^-$

Cathode Reaction: $O_2 + 4e^- \rightarrow 2O^{2-}$

Overall Cell Reaction: $2H_{2(g)} + O_{2(g)} \rightarrow 2H_2O$ $E^o_{cell} = 1.23V$



Anode Reaction: $2CO + 2O^{2-} \rightarrow 2CO_2 + 4e^-$

Cathode Reaction: $O_2 + 4e^- \rightarrow 2O^{2-}$

Overall Cell Reaction: $2CO_{(g)} + O_{2(g)} \rightarrow 2CO_2$ $E^o_{cell} = 1.066V$

Anode Reaction: $CH_4 + 4O^{2-} \rightarrow 2H_2O + CO_2 + 8e^-$

Cathode Reaction: $2O_2 + 8e^- \rightarrow 4O^{2-}$

Overall Cell Reaction: $CH_{4(g)} + 2O_{2(g)} \rightarrow 2H_2O + CO_{2(g)}$ $E^o_{cell} = 1.060V$

The disadvantage of using hydrocarbon fuels is the possible formation of coke, C, on the anode.

The cathode reaction is the same for the different anode-side fuel cases.
Reduction of the SOFC operating temperature by 200°C or more allows use of a broader set of materials, is less demanding on the seals and the balance-of-plant components, simplifies thermal management, aids in faster start up and cool down, and results in less degradation of cell and stack components. Because of these advantages, development of SOFCs capable of operating in the temperature range of 650 to 800°C is in progress. However, at lower temperatures, electrolyte conductivity and electrode kinetics decrease significantly; requiring alternative cell materials and designs.

SOFC Components

Electrolyte

Yttrium-doped zirconium oxide ($ZrO_2 + Y_2O_3$, YSZ) is the most widely used material for the electrolyte in SOFCs because of its adequate ionic conductivity, chemical stability, and mechanical strength. The only drawback of stabilized YSZ is the low ionic conductivity in the lower cell operation temperature regime, below about 750°C. This problem is resolved by decreasing the thickness of the YSZ electrolyte or other materials replace the yttrium. Scandium-doped zirconium oxide has higher conductivity than YSZ but the high cost of scandium and detrimental ageing effects in scandium doped zirconium oxide make it less attractive. Gadolinium or samarium doped cerium oxide CeO_2 materials possess higher oxide ion conductivity than zirconium based materials. However, cerium oxide (ceria) based materials, under reducing conditions at high temperatures, exhibit significant electron conductivity and dimensional change. Operation at temperatures below 600°C overcomes this cerium oxide problem. A mixture,

containing among others gallium oxide, has proved a good oxide ion conducting electrolyte. However, it has two drawbacks: uncertain cost of gallium, and oxide uncertain chemical and mechanical stability.

Cathode
The **cathode** material is important because the oxidation reaction determines the efficiency of the fuel cell. The stringent electrochemical and mechanical requirements greatly restrict the number of suitable cathode candidate materials. Lanthanum manganite $LaMnO_3$, which, when substituted with low valence elements such as calcium or strontium Sr, $La_{0.8}Sr_{0.2}MnO_3$, has good electron conductivity. Additionally, it possesses adequate electrocatalytic activity, a reasonable thermal expansion match to YSZ, and stability in the SOFC cathode electrochemical operating environment.

SOFCs cathodes operating at 650 to 800°C typically contain transition metals such as cobalt, iron, and/or nickel. These materials offer higher oxide ion diffusion rates and exhibit faster oxygen reduction kinetics at the cathode/electrolyte interface compared with high-temperature lanthanum manganite. However, the thermal expansion coefficient of cobaltites is much higher than that of the YSZ electrolyte, and the electrical conductivities of ferrites and nickelites are low. The improved cathodic performance deteriorates during the cell lifetime as a result of chemical or microstructural instability. Since these materials are reactive toward YSZ, a thin layer, generally of a cerium oxide based material, is used to reduce the chemical reaction between the cathode and YSZ. The polarization resistance depends on the microstructural grain size of the ionic conductor in the composite electrode and the volume fraction of porosity.

Anode
Nickel-YSZ composites (or Ni-ZrO_2 cermet) are the most commonly used SOFC anode materials. Nickel is an excellent catalyst for fuel oxidation; however, it possesses a high thermal expansion coefficient, and exhibits coarsening of microstructure due to metal aggregation through grain growth at cell operation temperatures. YSZ in the anode constrains nickel aggregation and prevents sintering of the nickel particles, decreases the effective thermal expansion coefficient bringing it closer to that of the electrolyte, and provides better adhesion of the anode with the electrolyte. Anode nickel is the catalyst for hydrogen oxidation and the electrical current conductor. It is also highly active for the steam reforming of methane. This catalytic property is exploited in the internal reforming SOFCs that can operate on fuels composed of mixtures of methane and water. Although nickel is an excellent hydrogen oxidation and methane-steam reforming catalyst, it also catalyzes the formation of carbon from hydrocarbons under reducing conditions. Unless sufficient amounts of steam are present along with the hydrocarbon to remove carbon from the nickel surface, the anode may be destroyed. As a result, even when using methane as the fuel, relatively high steam-to-carbon ratios are needed to suppress this detrimental reaction. Unfortunately, due to the high catalytic activity of nickel for hydrocarbon cracking, this approach does not work for higher hydrocarbons, and it is generally not possible to operate nickel-based anodes on higher hydrocarbon-containing fuels without pre-reforming with steam or oxygen. Additives to the Ni+YSZ cermet, such as 5% ceria or 1% molybdena inhibit the susceptible of nickel to become coated with a carbon layer when reacting with carbon based fuels, which could prevent further reaction. Nevertheless, nickel-YSZ composite remains the most commonly utilized anode material for SOFCs and is satisfactory for cells operating on clean, reformed fuel.
Advanced SOFC designs place additional constraints on the anode, such as tolerance of oxidizing environments and/or the ability to tolerate significant quantities of sulphur and/or hydrocarbon species in the fuel stream. Alternative anode materials, include cerium oxide or strontium titanate/cerium oxide mixtures, but the benefits obtained in terms of sulphur, hydrocarbon and/or redox tolerance are counterbalanced by other limitations, such as the difficulty of integrating such materials with existing cell and stack fabrication processes and materials. Copper based anodes can be used at intermediate temperatures (<800°C).

Interconnect
Doped lanthanum chromite is used as the interconnection for cells which operate at about 1000°C.
In cells intended for operation at lower temperatures (<800°C), it is possible to use oxidation-resistant metallic materials for the interconnection. Ferritic stainless steel alloys in this family offer:

- a protective and conductive chromium-based oxide scale,
- appropriate thermal expansion behaviour,
- ease of manufacturing, and
- low cost.

Although these alloys demonstrate improved performance over traditional compositions, several critical issues remain:

- chromium oxide scale evaporation and subsequent poisoning of cathodes;
- scale electrical resistivity in the long term;
- corrosion and spalling under interconnect exposure conditions; and
- compatibility with the adjacent components such as seals and electrical contact layers.

In mitigation, some surface coatings are applied onto the metallic interconnects to minimize scale growth, electrical resistance, and chromium volatility.

Stack design
Stack design has focused on planar and tubular design cells. The planar SOFC may be in the form of a circular disk fed with fuel from the central axis, or it may be in the form of a square plate fed from the edges. The tubular SOFC may be of a large diameter (>15mm), or of much smaller diameter (<5 mm), the so-called microtubular cells. Also, the tubes may be flat and joined together to give higher power density and easily printable surfaces for depositing the electrode layers. One of the inherent advantages of tubular cell bundles is that the air and the fuel are naturally isolated because the tubes are closed at one end.

The sealing materials and concepts for planar SOFCs is a important issue for the long-term performance stability and lifetime of planar SOFC.
The main advantage of tubular cells over planar cells is that they do not require any high temperature seals to isolate the oxidant from the fuel, and this stabilises the performance of tubular cell stacks over long periods of times. However, their areal power density is much lower (about 0.2W/cm^2) than planar cells (up to 2W/cm^2 for single cells and at least 0.5W/cm^2 for stacks) and manufacturing costs are higher. The volumetric power density is also lower for tubular cells than for planar cells.

High temperature cell operation has three mitigating advantages:

- Reduced need for precious metal, platinum, catalyst and less poisoning
- Reforming of fossil fuel to hydrogen can be performed internally
- The steam produced by the fuel cell can be channelled into turbines to generate more electricity. This process is called co-generation of heat and power (CHP) and improves the overall system efficiency.

22.8 Fuel Cell Summary

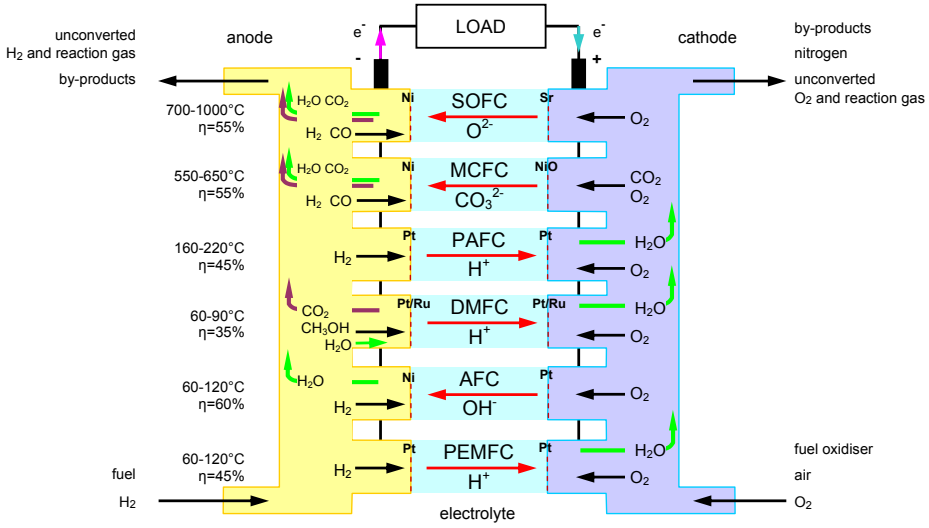


Figure 22.8. *The mechanisms of a fuel cell.*

22.9 Fuels

The choice of fuel depends on the application requirement, and the local fuel infrastructure, which influences fuel cell type, system cost, complexity, reformer requirements, etc.

Anode-side fuels:

- Hydrogen, H_2 (direct or via fuel reformer) — all FC
- Methanol, Ethanol, CH_3OH, C_2H_5OH — PEMFC, DMFC, AFC, SOFC
- Methane, C_2H_6 — SOFC, MCFC
- Carbon Monoxide, CO — SOFC, MCFC
- Hydrides, MH — AFC
- Coal — SOFC, AFC
- Metals — SOFC
- Sugars — biological-FC
- Other fuels - hydrazine, formic acid — DMFC

The chemical and consequential electrical properties of these anode-side fuel cells at 25°C are listed in Table 22.3, whilst the effects of anode fuel impurities on the different fuel cells are listed in Table 22.4.

Cathode-side fuels:
Suitable oxidants are air, pure oxygen 0_2, and hydrogen peroxide H_20_2. (see table 22.9) The use of oxygen can produce up to 30% more power than air. Undesirably, hydrogen peroxide gives a lower half-cell potential than oxygen, and is more reactive on the cell components.

H_2O_2 contains oxygen in a state of oxidation midway between molecular oxygen and water.

$$H_2O_2 \rightarrow O_2 + 2H^+ + 2e^- \qquad E_o = -0.682\ V$$
$$H_2O_2 + 2H^+ + 2e^- \rightarrow 2H_2O \qquad E_o = 1.776\ V$$

For the perhydroxyl ion (HO_2^-):

$$OH^- + HO_2^- \rightarrow O_2 + H_2O + 2e^- \qquad E_{Bo} = 0.084\ V$$
$$3OH^- \rightarrow HO_2^- + H_2O + 2e^- \qquad E_{Bo} = -0.87\ V$$

Table 22.3: Alternative anode-side fuels

Fuel Cells	Anode fuel	Cathode fuel	Enthalpy ΔH	Electrons/H_2 n	Theoretical voltage	Practical voltage	Power density
@25°C			kJ/mol		V	V	A-hr/kg
Carbon monoxide	CO	$1½0_2$	283.1	2	1.066		
Hydrogen	H_2	$½0_2$	286.0	2	1.229	0.7	26,000
Hydrazine	N_2H_4	0_2	622.4	4	1.560	0.7	2,100
Methanol	CH_3OH	$1½0_2$	726.6	6	1.214	0.8	1,400
Methane	CH_4	20_2	890.8	8	1.060		
Propane	C_3H_8	50_2	2221.1	20	1.093		

Table 22.4: Fuel constituents impact on fuel cell performance

Gas species	PEFC	AFC	DMFC	PAFC	MCFC	SOFC
H_2	fuel	fuel	fuel	fuel	fuel	fuel
CO	poison < 50ppm reversible	poison	poison	poison > ½%	fuel via shift	fuel via shift
CH_4	diluent	Poison/ diluent	fuel	diluent	diluent	fuel
CO_2, H_2O	diluent	poison	diluent	diluent	diluent	diluent
S (H_2S, COS)	poison cumulative irreversible	poison	poison	poison < 50 ppm reversible	poison < 0.5 ppm reversible	poison < 1 ppm reversible

22.10 Fuel Reformers

Low-temperature fuel cells, less than 200ºC, generally operate on pure hydrogen as the anode fuel. The transformation of fossil fuels to hydrogen is called fuel reforming. Coal, oil, and natural gas contain hydrocarbons, molecules consisting of hydrogen and carbon. A reformer splits the hydrogen from the carbon in a hydrocarbon, relatively easily. Three basic technical processing approaches can be used to recover hydrogen from hydrocarbons.

- Thermal Reforming Processes
- Electrolysis Processes
- Photolytic Processes

(i) Thermal Processes

Some thermal processes use the energy in various resources, such as natural gas, coal, or biomass, to release hydrogen, which is part of their molecular structure. In other processes, heat, in combination with closed chemical cycles, produces hydrogen from feedstocks such as water - these are known as thermo-chemical processes. Most processes involve a water-gas shift final reaction to form H_2 and CO_2 from any CO product. The various thermal reforming processes are:

- Natural Gas Reforming:
  - Steam methane reforming (see 22.12), 700 - 1000°C
  - Partial oxidation, $2CH_4 + O_2 + 2H_2O \rightarrow 2CO_2 + 6H_2$
- Coal Gasification:
  - $CH_{0.8} + O_2 + H_2O \rightarrow CO + CO_2 + H_2$ + other species
  - $CO + H_2O \rightarrow CO_2 + 2H_2$ + heat
  - $H_2S + ZnO \rightarrow ZnS + H_2O$ for sulphur removal
- Biomass Gasification:
  - $C_6H_{12}O_6 + O_2 + 2H_2O \rightarrow CO + 2CO_2 + 2H_2$
  - $CO + H_2O \rightarrow CO_2 + 2H_2$ + heat
- Renewable Liquid Fuels Reforming:
  - $C_2H_5OH + 3H_2O \rightarrow 2CO_2 + 6H_2$
  - $CO + H_2O \rightarrow CO_2 + 2H_2$ + heat
- High-Temperature Water Splitting, 500°C - 2000°C:
  - $2ZnO$ + heat $\rightarrow 2Zn + O_2$
  - $2Zn + 2H_2O \rightarrow 2ZnO + 2H_2$

(ii) Electrolytic Processes

The electrolysis process use electricity to split water into the diatomic molecules, hydrogen and oxygen. There are three basic electrolysing methods.

- Polymer Electrolyte Membrane Electrolyser, which operates at 80°C to 100°C
- Alkaline Electrolyser, which operates at 100°C to 150°C
- Solid Oxide Electrolyser, which operates at about 500°C to 800°C

Electrolytic processes use electricity to split water, H_2O into hydrogen H_2 and oxygen O_2, a process that takes place in an electrolyser or an electrolysis cell. The decomposition of water involves two partial reactions that take place at two electrodes. The electrodes are placed in an ion-conducting electrolyte (usually an aqueous alkaline solution with 30% potassium hydroxide KOH). Gaseous hydrogen is produced at the negative electrode (anode) and oxygen at the positive electrode (cathode). The necessary exchange of charge occurs through the flow of OH^- ions in the electrolyte and current (electrons) in the electric circuit. In order to prevent a mixing of the product gases, the two reaction areas are separated by a gas-tight, ion-conducting diaphragm membrane.
About 1 litre of water is required to produce 1 m^3 or 0.09kg hydrogen. Electrolysis requires 367kJ/mol of energy (which is higher than the ΔH = 286kJ/mol that can be derived from H_2) to produce one mole of hydrogen. That is, 1kW of electricity produces 1/267=0.00272 mol/s or 0.00545g/s of H_2.

Anode Reaction:	$2H_2O \rightarrow O_{2(g)} + 4H^+ + 4e^-$	
Cathode Reaction:	$4H^+ + 4e^- \rightarrow 2H_{2(g)}$	0
Overall Cell Reaction:	$2H_2O \rightarrow 2H_{2(g)} + O_{2(g)}$	4x1.23eV

(iii) Photolytic Processes

Photolytic processes use light energy to split water into hydrogen and oxygen, with low environmental impact.

- Photobiological Water Splitting
  - Chemical limit:- Glucose + $6H_2O \rightarrow 6CO_2 + 12H_2$
  - Biological limit:- Glucose + $2H_2O \rightarrow$ 2Acetate + $2CO_2 + 4H_2$
- Photoelectrochemical Water Splitting
  - Feedstock:- HCl, HBr, Biomass
    - O_2 generation :- $4Br_3^- + 6H_2O \rightarrow 12HBr + 3O_2$
    - H_2 generation:- $6HBr \rightarrow 2Br_3 + 3H_2$

22.10.1 Natural gas reforming

Steam reforming of natural gas, sometimes referred to as steam methane reforming, is the most common method of producing commercial bulk hydrogen as well as the hydrogen used in the industrial synthesis of ammonia. It is also the least expensive method and 48% of the worlds hydrogen is produced by steam methane reforming.

The process flow route is shown in figure 22.9 and follows three stages:

stage i ***Pre-Treatment of the Feed***

The hydrocarbon feedstock is de-sulphurised (chloride is also removed) using activated carbon filters, pressurised and, depending on the reformer design, either preheated and mixed with process steam or directly injected with the water into the reformer without the need of an external heat exchanger. The water is first softened and demineralised by an ion-exchange water conditioning system.

i. One option is high pressure reforming with integrated heat exchangers and a working pressure of up to 16 bar which reduces the geometric volume of the reformer vessels.

ii. The other option is to operate the reformer at low pressures, 1.5 bar, with an increased conversion ratio and compress the reformate prior to purification.

stage ii ***Steam Reforming and CO-Shift Conversion***

At high temperatures of 700 to 1100°C at 3 to 25 bar, and in the presence of a metal-based catalyst, nickel, steam reacts with methane CH_4, which is the primary component of natural gas, to yield carbon monoxide and hydrogen with minimal carbon dioxide, this synthesized gas is called syngas.

$$CH_4 + H_2O \rightarrow CO + 3H_2 \qquad \text{reformation} \qquad \Delta H = 206 \text{ kJ/mol} \qquad (22.8)$$

This reaction is endothermic ($\Delta H > 0$, consumes heat), with the heat required obtained from the combustion of fuel gas and purge/tail gas from the subsequent gas purification stage.

After reforming, the synthesis gas is fed into the CO conversion lower-temperature gas-shift reactor to produce additional hydrogen. The carbon monoxide is reacted with steam over a catalyst to form hydrogen and carbon dioxide, in two stages. The first stage is a high temperature water gas shift at 350°C and then a low temperature shift at 200°C. The reaction is exothermic ($\Delta H < 0$, produces heat) and is summarised by:

$$CO + H_2O \rightarrow CO_2 + H_2 \qquad \text{water-gas shift} \qquad \Delta H = -41 \text{ kJ/mol} \qquad (22.9)$$

The chemical reactions that take place are:

$$\begin{aligned} C_nH_m + n\,H_2O &\rightarrow n\,CO + (½m + n)\,H_2 \\ CO + H_2O &\rightarrow CO_2 + H_2 \end{aligned} \qquad (22.10)$$

stage iii ***Gas Purification***

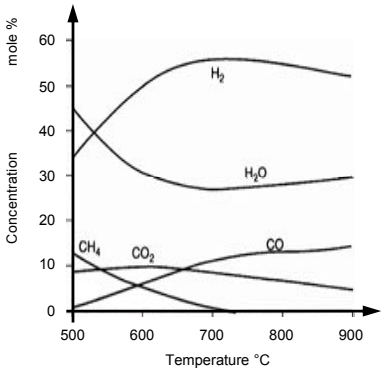
Hydrogen purification is achieved by means of pressure swing adsorption through four successive vessels filled with selected adsorbents. Purification to remove carbon dioxide and carbon monoxide can achieve hydrogen purities greater than 99.999% by volume and CO impurities of less than 1 vppm (volumetric part per million). The pure hydrogen is compressed, while the gases from recovering the adsorbents, called tail-gas, is fed to the steam reformer burner.

The complete steam reforming process is 70% efficient, with carbon sequestration.

In high-temperature fuel cells (MCFC and SOFC), CO in the fuel stream acts as a fuel. However, it is likely that the water-gas shift reaction is occurring internally and the fuel is actually hydrogen.

$$CO + H_2O \rightarrow CO_2 + H_2 \qquad (22.11)$$

Steam reforming can also be used to produce hydrogen from other fuels, such as ethanol, propane, or even petrol.



Methane:

$$CH_4 + H_2O\,(+heat) \rightarrow CO + 3H_2$$

Propane:

$$C_3H_8 + 3H_2O\,(+heat) \rightarrow 3CO + 7H_2 \qquad (22.12)$$

Ethanol:

$$C_2H_5OH + H_2O\,(+heat) \rightarrow 2CO + 4H_2$$

Petrol, using iso-octane and toluene which are only two of compounds in petrol:

$$C_8H_{18} + 8H_2O\,(+heat) \rightarrow 8CO + 17H_2$$

$$C_7H_8 + 7H_2O\,(+heat) \rightarrow 7CO + 11H_2 \qquad (22.13)$$

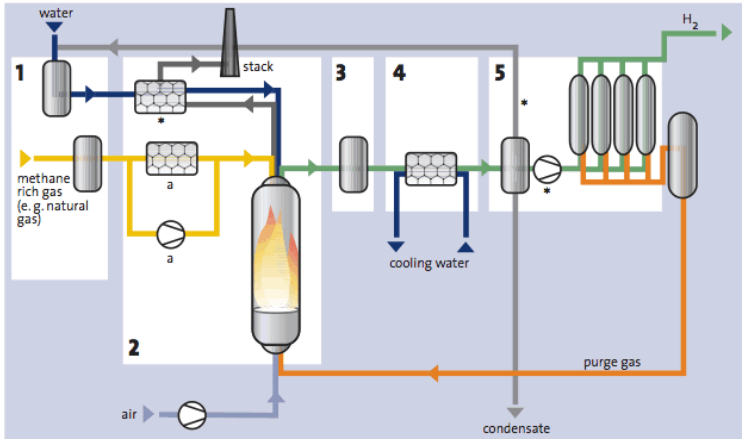


Figure 22.9. *Flow chart of a steam reformer:*
1 Feed Pre-Treatment *2 Reforming and Steam Generation*
3 High Temperature Conversion *4 Heat Exchanger Unit*
5 Purification Unit, depending on reformer design 'a' either heat exchanger for low pressure reformer or compression to 1 bar for high pressure reformer.

22.11 Hydrogen storage and generation from hydrides

Hydrogen can be stored at pressures up to 30MPa (300atm), but the energy density is low. Storage of liquid hydrogen at 22K requires a lot of energy and the energy density is still low. The viable storage possibilities are metal hydrides, glass micro-spheres and carbon nano-fibres.

Table 22.5a **Hydrogen storage**

Storage media	Density $g.cm^{-3}$ at 20°C	H_2 storage capacity g/cc
$LaNi_5$	8.3	0.11
FeTi	6.2	0.11
Mg_2Ni	4.1	0.15
Mg	1.74	0.13
MgNi eutectic	2.54	0.15
liquid H_2	0.07	0.07

Metal hydrides

Metal hydrides offer a safe, constant pressure way to store and subsequently retrieve pure hydrogen. Sodium borohydride is a white solid at room temperature, stable in dry air and decomposing only at temperatures above 400°C. Dry sodium borohydride (or alternatively $LiAlH_4$) use the following metal hydride retrieval reaction:

$$NaBH_{4(aq)} + H_2O_{(l)} \xrightarrow{cat} 4H_{2(g)} + NaBO_{4(aq)} + heat$$

When hydrogen fuel is needed, a sodium borohydride solution is pumped at a controlled rate over a catalyst within the reactor, producing pure hydrogen and sodium borate, a non-toxic compound found in detergents. Hydrogen gas is generated proportionally to the rate at which the borohydride solution is pumped into the reactor. Hydrogen gas is then separated from the liquid metaborate by-product. The hydrogen is humidified because the heat of reaction converts some liquid water to vapour. The by-product can be recycling after removal from the system. Both products and reactants are indefinitely reusable and stable to air. The system has a high effective hydrogen pressure of about 7,000 psi.

Since the highly flammable hydrogen gas is generated in a miniature reactor immediately before use, it is a safe way of storing hydrogen compared to compressed or liquefied hydrogen. $NaBH_4$'s stability in air also makes it a much safer choice than its pyrophoric reversible metal hydride counterpart. A disadvantage of the method is its high cost sine sodium borohydride, is not a common chemical reagent. Only 4% of the heavy high pressure tank mass is hydrogen.

A complex hydride system based on lithium amide has been developed based on the following reversible displacive reaction taking place at 285°C and 1 atm:

$$Li_2NH + H_2 \rightarrow LiNH_2 + LiH$$

In this reaction, 6.5 wt. % hydrogen can be reversibly stored with potential for 10 wt.%. The temperature of this reaction can be lowered to 220°C with magnesium substitution, although at higher pressures.

Alloy hydrides

AB_5 and AB_2 type alloys have advantages in hydrogen storage capabilities and operation parameters.

AB_5 alloys combine a hydride forming metal A, usually a rare earth metal (La, Ce, Nd, Pr, Y or their mixture known as Mischmetal), with a non-hydride forming element B – nickel. The nickel can be doped with other metals, such as Co, Sn or Al, to improve material stability or to adjust the equilibrium hydrogen pressure and temperature required for hydrogen discharge.

AB_2 alloys, also know as Laves phases, represent a group of alloys containing titanium, zirconium or hafnium at the A-site and a transition metal(s) at a B-site (Mn, Ni, Cr, V and others). Although reversible hydrogen storage capabilities are similar to AB_5 alloys, AB_2 alloys can store additional hydrogen at high pressures and have higher discharge rate capacity when used as battery negative electrodes.

General properties of metal hydride hydrogen storage are:

- Storage capacities up to 3.6 wt.%
- Reversible
- Stable in air and
- Operate at room temperature and 1 to 10 atmospheres

Table 22.5b Hydrogen storage in alloys

Storage media	Chemical composition	Hydrogen storage capacity wt.% @25°C	Equilibrium pressure plateau bar @ 25°C
Lanthanum-nickel alloy	$LaNi_5$	1.5 - 1.6	≈2
Lanthanum-nickel-cobalt alloy	$LaNi_4Co_{0.5}$	1.4 - 1.5	≈0.5
Mischmetal-Nickel alloy	$MnNi_5Mn$ La: 20-27%; Ce: 48-56%; Pr: 4-7%; Nd: 12-20%	1.5 - 1.6	≈10
Titanium-Manganese Alloy	$Ti_{0.98}Zr_{0.02}V_{0.43}Fe_{0.09}Cr_{0.05}Mn_{1.5}$	1.6 - 1.7	≈10

Glass micro-spheres can be encouraged to absorb H_2 gas at higher temperature, then retain it at lower temperatures before releasing it again when the temperature is raised. A technology yet to be proven. Storage is safe, is reusable, and un-contaminable, with large H_2 per unit volume. The technology is complicated, expensive, and delicate.

Carbon nano-fibres provide a large surface for H_2 to bind to and increase the amount of H_2 that can be stored for a given tank pressure. Reported capable of absorbing 67% times their own mass of hydrogen and inert storage is safe. The technology is unproven technology and there are long-term carcinogenic safety concerns with nano-fibres. Carbon nano-fibres are expensive.

22.12 Fuel Cell Emissions

Fuel cells running on hydrogen derived from a renewable source emit only water vapour. The following Table 22.6 shows a comparison of the water vapour plus carbon and nitrogen based emissions from fuel cells, running on a variety of fuels, as compared to the internal combustion engine in the first data row.

Table 22.6: Fuel cell emission properties

Engine Type	Water Vapour	Carbon Dioxide	Carbon Monoxide	Nitrous gases	Hydro-Carbons
	H_2O	CO_2	CO	NO_x	C_xH_y
	kg/km	kg/km	kg/km	kg/km	kg/km
Gasoline Combustion	0.111	0.241	0.013	0.87×10^{-3}	1.75×10^{-3}
Fuel Cell operating on Hydrogen from Gasoline	0.091	0.199	-	-	-
Fuel Cell operating on Hydrogen from Methane	0.071	0.043	0.01×10^{-3}	1.6×10^{-6}	2.1×10^{-6}
Fuel Cell operating on Renewable Hydrogen	0.071	0.0	0.0	0.0	0.0

22.13 Fuel Cell Electrical characteristics

A typical fuel cell produces a voltage from 0.6V to 0.7V at full rated load. Voltage decreases as current increases, as shown in figure 22.10, due to numerous factors:

- *Activation loss*, which is the potential difference above the equilibrium value required to produce a current flow.
- *Ohmic losses*, which are voltage drops due to resistance of the cell components and interconnects
- *Mass transport loss*, which is the depletion of reactants at catalyst sites under high loads, especially at the cathode, causing rapid loss of voltage
- *Mixed potential reactions* at the electrodes due to parasitic reactions that lower the equilibrium potential. This can be due to fuel impurities, cathode H_2O_2 intermediaries, or anode fuel crossover through the electrolyte.
- *Activation over-potential* at the cathode due to sluggish oxygen reduction kinetics, compared to faster anode reaction rate, which are most pronounced at low current densities

Energy conversion of a fuel cell can be summarized by the following equation:

Chemical energy of fuel = electrical energy + Heat energy

A single, ideal H_2/air fuel cell should provide 1.16V at zero current, that is, open-circuit, 80°C and 1 atmosphere gas pressure. A measure of energy conversion efficiency for a fuel cell is the ratio of the actual cell voltage to the theoretical maximum voltage for the H_2/air reaction. Thus a fuel cell operating at 0.7V is generating about 60% of the maximum useful energy available from the fuel as electric power. The remaining energy, 40%, is heat. The characteristic performance curve in figure 22.10 for a fuel cell represents the DC voltage delivered at the cell terminals as a function of the current density, that is, total current divided by the membrane area, being drawn from the fuel cell by the load in the external circuit. Note that the efficiency increases as the loading decreases, unlike the internal combustion engine.

The steady-state power, $P = I\times V$, expressed in watts, delivered by a cell is the product of the current I drawn and the terminal voltage V at the current I. Power is also the rate at which energy W is made available, $P = W / t$, or conversely, energy $W = P\times t = V\times I\times t = V\times Q$, expressed in units of joules, or watt-hours, is the continuous power available over a time period t.

- Specific power is the ratio of the power produced by a cell to the mass of the cell, and
- Power density is the ratio of the power produced by a cell to the volume of the cell.

High specific power and power density are important for transportation applications, to minimize the weight and volume of the fuel cell as well as to minimize cost.

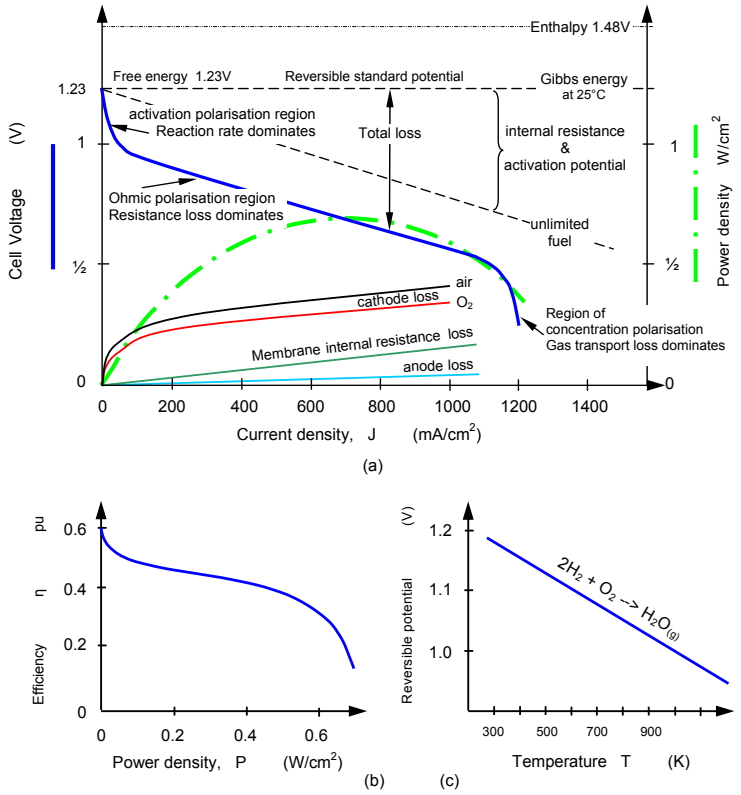


Figure 22.10. *Graphs of a hydrogen/air fuel cell: (a) voltage versus current density, (b) efficiency versus current density; and (c) open circuit voltage versus temperature.*

22.14 Thermodynamics

Prediction of the maximum possible voltage from a fuel cell process involves evaluation of energy differences between the initial state of the reactants in the process, H_2 + ½0_2, and the final state, H_20. Such evaluation relies on state thermodynamic functions in a chemical process, primarily the Gibbs free energy. The maximum cell voltage, at electrical output open-circuit, or emf, ΔE, for the hydrogen/air fuel cell reaction (H_2 + ½0 → H_20) at a specific temperature and pressure is given by the relationship, in electrical terms:

$$\Delta E = -\frac{\Delta G}{n \times e \times N_o} = -\frac{\Delta G}{n \times F} = \frac{W}{n \times F} \qquad \text{(V)} \qquad (22.14)$$

where ΔG is the Gibbs free energy change for the reaction or energy available for electrical work, W,
n is the number of moles of electrons involved in the reaction per mole of reactant, H_2, and
F is Faraday's constant, 96,487 C/mol (J/V) (26.8015Ah/kg-equiv), the charge transferred per mole of electrons.

Faradays constant is Avogadro's number $N_o = 6.02\times10^{23}$ multiplied by the charge of one electron, $e = 1.60\times10^{-19}$. The product $n\times F$ or charge Q is the amount of electricity produced. At one atmosphere of pressure, the Gibbs net free energy change ΔG in the fuel cell process (per mole of H_2), available for net useful work, is calculated from the reaction temperature T, and from changes in the reaction enthalpy, ΔH, and entropy, ΔS.

$$\Delta G = -n \times F \times \Delta E = \Delta H - T \times \Delta S \qquad \text{(J)} \qquad (22.15)$$

where T is the absolute temperature, K
ΔS is the entropy change for the reaction, ($\Delta S<0$ for a fuel cell reaction), J/K
ΔH is the enthalpy or energy (change) released by the reaction, J.

The product $T\times\Delta S$ is the heat associated with the electrochemical processes occurring, leaving $\Delta H - T\Delta S$ available for the externally diverted production of electrical energy. A change will be spontaneous when the enthalpy term (ΔH) and the entropy term ($T\Delta S$) combine to give a negative value of ΔG (when free energy decreases) and a positive emf ΔE, as illustrated in figure 22.11. At equilibrium $T\Delta S = \Delta H$.

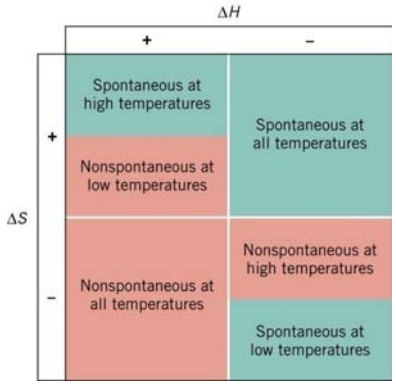


Figure 22.11. *Summary of the effects of the sign of the enthalpy and entropy change on the spontaneity of a process.*

Combining equations (22.14) and (22.15), the maximum cell voltage is therefore

$$\Delta E = -\frac{\Delta H - T \times \Delta S}{n \times F} \qquad \text{(V)} \qquad (22.16)$$

When the source is loaded, the total cell internal heat released (losses) q are

$$q = T \times \Delta S + I\left(E_{o/c} - E_I\right) \qquad \text{(W)} \qquad (22.17)$$

where $E_{o/c}$ is the cell open circuit voltage ($I = 0$), that is ΔE from equation (22.16) and
E_I is the terminal voltage at load current I.

The term $I\times(E_{o/c} - E_I)$ is the power dissipated in the fuel cell effective internal resistance, or Thevenin resistance.

The various enthalpies and entropies are characterised at 25°C (NTP) and the symbols are annotated with a zero superscript. Operation at temperatures other than NTP and reaction quotient N involves the use of Nernst equation (R is the gas constant, 8.314 J/K mol).

$$E = E_o - \frac{RT}{nF}\ell n N \qquad (22.18)$$

The overall cell efficiency comprises three components, namely the Gibbs efficiency η_G, the voltage efficiency, η_V, and the utilisation efficiency is the fraction of the input fuel used, η_u. That is

$$\eta = \eta_G\eta_V\eta_u = -\frac{nF}{\Delta H}E_I = -\frac{Q}{\Delta H}E_I = -\frac{I \times E_I}{\Delta H \times t} \qquad (22.19)$$

where the Gibbs thermodynamic efficiency is $\eta_G = \Delta G/\Delta H = 1 - T\Delta S/\Delta H = \Delta EIt/\Delta H$
the voltage efficiency is $\eta_V = E_I/E_{o/c}$ where E_I is the output voltage at current I.

Example 22.1: ***Formation of water vapour***

Based on the table below, is the following reaction spontaneous at 250°C?
$H_{2(g)} + ½O_{2(g)} \rightarrow H_2O_{(g)}$

Compound	Enthalpy ΔH_f^0	Entropy ΔS^0
	kJ/mol	J/mol.K
H_2 (g)	0.0	130.6
O_2 (g)	0.0	205.0
H_2O (g)	-241.8	188.7
H_2O (l) 1Atm & 298K	-285.14	69.91

Solution

First ΔH and ΔS are calculated for the reaction, then the value of ΔG, to see if it is less than zero.
First, look up the thermodynamic data
From the thermodynamic data in the table, ΔH and ΔS are both calculated
The standard enthalpy of formation of any element is zero, thus using Hess' law, for 1 mole of product

$$\Delta H^\circ = \Delta H^\circ_{f(products)} - \Delta H^\circ_{f(reactants)} = -285.8 \text{ kJ/mol} - 0 = -285.8 \text{ kJ/mol}.$$

$$\begin{aligned}\Delta H^\circ &= \Delta H^o_{f(products)} - \Delta H^o_{f(reactants)} \\ &= \Delta H^o_{H_2O} - \left(1 \times \Delta H^o_{H_2} + \tfrac{1}{2} \times \Delta H^o_{O_2}\right) \\ &= -241.8\text{kJ/mol} - \left(1 \times 0 + \tfrac{1}{2} \times 0\right) \\ &= -241.8\text{kJ/mol}\end{aligned}$$

The entropy is

$$\begin{aligned}\Delta S^\circ &= \Delta S^o_{f(products)} - \Delta S^o_{f(reactants)} \\ &= \Delta S^o_{H_2O} - \left(1 \times \Delta S^o_{H_2} + \tfrac{1}{2} \times \Delta S^o_{O_2}\right) \\ &= -188.7\text{J/mol.K} - \left(1 \times 130.6 + \tfrac{1}{2} \times 205.0\right) \\ &= -44.3\text{J/mol}\end{aligned}$$

Combining ΔH and ΔS to give ΔG

$$\begin{aligned}\Delta G &= \Delta H - T \times \Delta S \\ &= -241.8\text{kJ/mol} - 298\text{K} \times (-0.0443 \text{ kJ/mol.K}) \\ &= -241.8\text{kJ/mol} - 13.20\text{kJ/mol} \\ &= -228.5\text{kJ/mol}\end{aligned}$$

The reaction is spontaneous, since $\Delta G < 0$.
The chemical reaction can do 228.5kJ/mol of work and produces 13.2kJ/mol of heat to the environment.

♣

Example 22.2: ***Derivation of Ideal Fuel Cell Voltage***

The reaction $\frac{1}{2}O_{2(g)} + H_{2(g)} \rightarrow H_2O_{(l)}$ is used in fuel cells to produce an electrical current. The reaction can also be carried out by direct combustion.

Thermodynamic data, from table 22.9:
molar entropies ΔS° in J/mol.K: $H_2O_{(l)}$ 70.0; $O_{2(g)}$ 205.0; $H_{2(g)}$ 130.6;
molar enthalpies ΔH°_f in kJ/mol: $H_2O_{(l)}$ -285.8 (note, liquid H_2O).
Use this information (and table 22.9 if necessary) to find

i. The amount of heat released when the reaction takes place by direct combustion;
ii The amount of heat released by the fuel cell under the same conditions;
iii The amount of electrical work the same reaction can perform when carried out in a fuel cell at 298K under reversible conditions;
iv. The cell voltage at an operating temperature of 80°C;
v. The fuel cell has an active area of 100cm^2 and generates 0.6A/cm^2. Estimate the cell efficiency and heat generated, if the cell voltage is 0.7V at this current density; and
vi. If the water is formed as a gas, calculate the open circuit voltage and the cell thermodynamic efficiency for $H_2O_{(g)}$ ΔH°_f = -241.8 kJ/mol and ΔS° =188.7 J/mol.K.

Solution

In this reaction, 1.5 moles of gas (½ mole of O_2, 16g and 1 mole of H_2, 2g) are consumed to be replaced by 18 mL of water (1 mole of H_2O weighs 18g).

First, find ΔH° and ΔS° for the process at 298K (see example 22.1).

i. The standard enthalpy of formation of any element is zero, thus using Hess' law

$$\Delta H^\circ = \Delta H^\circ_{f(products)} - \Delta H^\circ_{f(reactants)} = -285.8 \text{ kJ/mol} - 0 = -285.8 \text{ kJ/mol}.$$

That is, when hydrogen and oxygen combine directly, the heat released is -285.6kJ/mol

Similarly, $\Delta S^\circ = S^\circ_{f(products)} - S^\circ_{f(reactants)} = (70.0) - (\frac{1}{2}\times 205.0 + 130.6) = -163.2$ J/mol.K

ii. The heat released in the fuel cell reaction is the difference between the enthalpy change (the total energy available) and the reversible work that was expended:

$$\begin{aligned}T \times \Delta S &= (298 \text{ K}) \times (-163.2 \text{ J/K}) \\ &= -48{,}600 \text{ J} = -48.6 \text{ kJ}\end{aligned}$$

iii. The Gibbs free energy change in the fuel cell process (per mole of H_2) gives the maximum electrical work the fuel cell can perform and is calculated from equation (22.15):

$$\begin{aligned}W = \Delta G &= \Delta H - T \times \Delta S \\ &= -285{,}800 \text{ J} - (298 \text{ K}) \times (-163.2 \text{ J/K}) \\ &= -237{,}200 \text{J}\end{aligned}$$

iv. From equation (22.14), for the hydrogen/air fuel cell at 1 atmosphere pressure and 25°C (298K), the cell voltage is 1.23V:

$$\Delta E = -\frac{\Delta G}{nF}$$

$$= -\frac{-237{,}200 \text{ J}}{2 \text{ x } 96{,}487 \text{ J/V}} = 1.23 \text{ V}$$

As temperature rises from room temperature to that of an operating fuel cell, 80°C or 353 K, the enthalpy values of ΔH and ΔS change slightly, but T changes by 55°C. Thus the absolute value of ΔG decreases. As an estimation, assume no change in the values of ΔH and ΔS.

$$\Delta G = \Delta H - T \times \Delta S$$

$$= -285{,}800 \text{ J/mol} - (353 \text{ K})(-163.2 \text{ J/mol K}) = -228{,}200 \text{ J/mol}$$

Thus, the maximum cell voltage decreases (for the standard case of 1 atm), from 1.23 V at 25°C to 1.18V at 80°C:

$$\Delta E = -\frac{\Delta G}{nF}$$

$$= -\frac{-228{,}200 \text{ J}}{2 \text{ x } 96{,}487 \text{ J/V}} = 1.18\text{V}$$

An additional correction for air, instead of pure oxygen, and using humidified air and hydrogen, instead of dry gases, further reduces the maximum voltage obtainable from the hydrogen/air fuel cell to 1.16V at 80°C and 1 atmosphere pressure.

v. The cell current is

$$\begin{aligned}I_{cell} &= J(\text{A/cm}^2) \times Area(\text{cm}^2) \\ &= 0.6\text{A/cm}^2 \times 100\text{cm}^2 = 60\text{A}\end{aligned}$$

The excess heat generated per fuel cell can be estimated as follows

$$\begin{aligned}P_{heat} &= P_{total} - P_{electrical} \\ &= V_{ideal} I_{cell} - V_{cell} I_{cell} \\ &= 1.16\text{V} \times 60\text{A} - 0.7\text{V} \times 60\text{A} \\ &= 69.6\text{W} - 42.0\text{W} \\ &= 27.6\text{W or } 27.6\text{J/s}\end{aligned}$$

The cell produces 69.6W of which 27.6W are internal losses with 42W delivered to the electrical load. Alternatively, the cell loses 27.6 J every second.
The effective internal resistance is

$$R = \frac{V_{ideal} - V_{cell}}{I_{cell}}$$

$$= \frac{1.16\text{V - } 0.7\text{V}}{60\text{A}} = 7.7\text{m}\Omega$$

Assuming 100% of the hydrogen reacts, for equation (22.19) the cell efficiency at 60A is

$$\eta = -\frac{nF}{\Delta H} E_1$$

$$= -\frac{2 \text{ x } 96{,}487 \text{ J/V}}{-285{,}800 \text{ J}} 1.09\text{V} = 73.6\%$$

vi. The entropy can be calculated as in part i. ΔS = [188.7] - [130.6 + ½ x 205] = -44.4 J/mol.K.

$$\Delta E = -\frac{\Delta H \; - \; T \times \Delta S}{nF}$$
$$= -\frac{-241{,}800\ \text{J} - (298\ \text{K}) \times (-44.4\ \text{J/K})}{2 \times 96{,}487\ \text{J/V}}$$
$$= -\frac{\Delta G}{nF} = -\frac{-228{,}600\ \text{J}}{2 \times 96{,}487\ \text{J/V}} = 1.185\text{V}$$

The Gibbs thermodynamic efficiency is given by

$$\eta_G = \frac{\Delta G}{\Delta H} = \frac{228{,}600\ \text{J}}{241{,}800\ \text{J}} = 94.5\%$$

which is higher than the liquid case, which from part iii is 237,200J / 285,800J = 83%.

♣

Example 22.1 along with figure 22.10c, illustrate how fuel cell open-circuit voltage ΔE (and efficiency, $\Delta G / \Delta H$) decreases with increased operating temperature. Table 22.7 shows the standard potentials for the various types of fuel cells, when functioning at their normal operating temperature which is well in excess of NTP conditions, when 1.23V is predicted if only water and heat are products. The table also shows the Carnot efficiency (η_{carnot} = 1 - T/298K) for pure hydrogen and oxygen forming water at an ambient of 25°C. This efficiency shows that the fuel cell potential efficiency (SOFC) is higher than conventional combustion energy conversion, for temperatures below 870°C.

Table 22.7: Ideal standard potential for different fuel cell types at 1 atm (1 bar) pressure

Fuel type	Cell type	Ideal $H_2 + O_2$ cell	PEMFC	AFC	PAFC	MCFC	SOFC
temperature	°C	25	80	150	200	650	800 / 1000
H_2	V	1.230	1.183	1.155	1.143	1.021	0.98 / 0.919
ΔG	kJ/mol	-237.2	-228.2	-222.7	-220.4	-196.6	-188.6 / -177.4
η Efficiency	%	83	80	78	77	69	76 / 62
η_{carnot} = 1 - T/298K	%	0	15	29	37	68	72 / 77
CH_4	V	N/A	N/A	N/A	N/A	1.036	1.034 / 1.032

Example 22.3: ***Carbon fuel cell***

A fuel cell has the following reactions:

Anode: $C + 2O^{2-} \rightarrow CO_2 + 4e^-$

Cathode: $4e^- + O_2 \rightarrow 2O^{2-}$

At STP, the changes of enthalpy and of free energy, per mole of CO_2, are:

ΔH = -393.5 kJ/mol

ΔG = -394.5 kJ/mol

i. What is the overall reaction?
ii. What is the ideal emf?
iii. What is the difference in entropy between reactants and products?

Assume that the internal resistance of the cell is 1 mΩ. Otherwise, the cell behaves as an ideal voltage source. If the cell is to deliver 1 MWh of electricity to the load in minimum possible time

iv. What is the load resistance under such conditions?
v. How much carbon is needed?

Solution

i. The overall reaction is

$$C + O_2 \rightarrow CO_2 \qquad \Delta H = -393.5\ \text{kJ/mol}$$

ii. Since for each mole of product CO_2, 4 moles of electrons pass trough the load (n_e = 4), the open-circuit voltage is

$$V_{OC} = \frac{\Delta G}{n_e q N_o}$$
$$= \frac{394.4\text{kJ/mol}}{4 \times 1.6 \times 10^{-19} \times 6.02 \times 10^{23}} = 1.024\text{V}$$

iii. Assume that the cell operates isothermally at 298 K (reactants and products at the same temperature). Then,

$$\Delta G = \Delta H - T\Delta S$$

$$\Delta S = \frac{\Delta H - \Delta S}{T}$$
$$= \frac{-393.44\text{kJ/mol} - 394.5\text{kJ/mol}}{298\text{K}} = 3\text{J/mol.K}$$

Notice that in this reaction, owing to the free energy change being (in absolute value) larger than the enthalpy change, there is an increase in entropy. If the cell is to operate reversibly, that is, with no entropy change, then it has to absorb heat - it will have to operate endothermically.

iv. Maximum power transfer occurs when the load resistance is equal to the internal resistance, $R_L = R$ = 1mΩ.

v. To transfer 1 MWh (3.6 ×10^9 J) in minimum time, maximum power is when $R_L = R$ = 1mΩ, and the current is

$$I_L = \frac{V_L}{R + R_L}$$
$$= \frac{1.024\text{V}}{2\text{m}\Omega} = 512\text{A}$$

The load voltage is

$$V_L = I_L R_L = 512\text{A} \times 1\text{m}\Omega = 0.512\text{V}$$

The power in the load is

$$P_L = V_L I_L = 0.512\text{V} \times 512\text{A} = 262.1\text{W} \ (\text{or J/s})$$

The time necessary to deliver 3.6 ×10^9 J at the rate of 262.1 J/s is

$$t = \frac{\text{energy}}{\text{power}} = \frac{3.6 \times 10^9\,\text{J}}{262.1\text{W}} = 13.7 \times 10^6\,\text{s}$$

The energy that the fuel cell must deliver is 2 ×3.6 ×10^9 J =7.2×10^9 J because the internal loss is equal to the power in the load.
If 1 mole of carbon delivers 394.5 kJ of electricity, then to deliver 7.2×10^9 J, the number of moles of carbon must be

$$N = \frac{7.2 \times 10^9\,\text{J}}{394.5\text{kJ}} = 18.25\ \text{kmoles}$$

♣

22.15 Fuel Cell features

Having considered the structure and operating mechanisms and characteristics, the following fuel cell features can be summarised.

Advantages

- Fuel cell power is a clean alternative to the internal combustion engine, fuelled with only hydrogen and producing no pollutants other than water. Direct energy conversion.
- Simple fuelling requiring only hydrogen fuel, taking the required oxygen from the air.
- Fuel flexibility in high operating temperature cells.
- No recharging is necessary, low maintenance.
- A perpetual efficient, primary cell.
- So long as fuel is provided, the cells can provide continuous power.
- Maximum efficiency at low power levels; opposite to the internal combustion engine.
- Operate with high efficiency >60%; internal combustion engine is typically about 30%.
- High energy density.
- Silent operation and safe, with no moving parts, vibration free
- Portable, modular construction.

Shortcomings

- The environmentally friendly credentials of fuel cells overlook the processes needed to generate and distribute the necessary hydrogen fuel. Fuel cells merely shift the pollution from the fuel cell to the reforming location.
- 98% of hydrogen is produced from fossil fuel sources.
- No infrastructure exists to provide and distribute the necessary hydrogen fuel.
- Electrolyte freeze-up at low temperatures.
- Electrodes and catalysts are prone to contamination and are expensive.
- Because of the exotic materials and complex design, the system is expensive.
- Not yet proven commercially viable in common usage, compared to the alternatives.
- Best as primary source.
- Limited availability.
- Low durability.
- Low power density per unit volume, not volume efficient due to associated ancillaries.
- Alternatively hydrogen can be generated in situ, as required, from hydrocarbon fuels such as ethanol, methanol, petrol or compressed natural gas from the reforming process. Reforming generates carbon dioxide as a waste product. It is also expensive with a chemical plant in situ, but this does simplify the fuel supply infrastructure problem, however the fuel could just as readily be used in an internal combustion engine.
- Despite safety precautions, there is a perception that hydrogen fuel is unsafe.
- The low cell voltage, 0.6V to 0.7V, means that many series connected cells are needed.
- Pulse demands shorten lifetime.
- The process is not reversible within the fuel cell and as with the primary cell, it cannot accept or store regenerative energy. The reactants must flow continuously.
- Fuel cells have a low dynamic range and slow transient response which causes an unacceptable lag in responding to instant power demands. At low powers levels, a power boost from a battery or supercapacitor would improve transient performance.
- Most designs operate at high temperatures so as to achieve reasonable operating efficiencies. To generate the same efficiencies at lower temperatures requires large quantities of expensive platinum catalysts.
- Complex to operate.

22.16 Fuel Cell Challenges

The shortcomings of the fuel cell represent the possible challenges of this energy source.

22.16.1 Chemical Technology Challenges

Fuel cell material development is mainly concern with electrolytic membrane materials. Many of the following challenges are orientate towards EV application using the PEM fuel cell.

- Durability: More durable PEMFC membranes are needed that can operate at temperatures greater than 100°C and still function at sub-zero ambient temperatures. A 100°C temperature target is required in order for a fuel cell to have a higher tolerance to impurities in fuel. Also, water by-product at over 100ºC has better co-generation possibilities. Because of frequent start and stop, it is important for the membrane to remain stable under cycling conditions. Membranes tend to degrade with cycling, particularly as operating temperatures rise.
- Hydration: PEMFC membranes must be hydrated in order to transfer hydrogen protons. At around 80°C, hydration is lost without a high-pressure hydration system. Membranes are needed for sub-zero temperature operation, low humidity environments, and high operating temperatures.
- The SOFC durability: Solid oxide systems have issues with material corrosion. Seal integrity is also a major concern. The cost goal for SOFC's is less restrictive than for PEMFC systems, but material costs are high. SOFC durability suffers with the cell temperature repeatedly heat cycled with start-up and shut down sequences.
- Aromatic-based membranes: An alternative to current perfluorosulphonic acid membranes are aromatic-based membranes like benzene, pyridine or indole. These membranes are more stable at higher temperatures, but still require hydration. They swell when they lose hydration, which reduces fuel cell efficiency.

22.16.2 System Technology Challenges

Cost and durability are the major challenges to widespread fuel cell commercialization. However, hurdles vary according to the application in which the technology is employed. Size, weight, and thermal and water management are barriers to the commercialization of fuel cell technology. In transportation applications, fuel cell technologies face more stringent cost and durability hurdles. In stationary power applications, where cogeneration of heat and power is desired, use of PEM fuel cells would benefit from raised operating temperatures to increase efficiency performance. The key challenges include:

- Cost: The fuel cell power system is more expensive than conventional technologies.
- Durability and Reliability: The durability of fuel cell systems has not been established. For transportation applications, fuel cell power systems are required to achieve the same level of durability and reliability of current energy sources in terms of lifespan time, performance, and ambient operating temperature.
- System Size: The size and weight of fuel cell systems must be reduced to the levels of other technologies. This reduction applies to the fuel cell stack and the ancillary components and major subsystems (for example, fuel processor, compressor/expander, and sensors) making up the balance of power system.
- Improved Heat Recovery Systems: The low operating temperature of PEM fuel cells limits the amount and type of CHP applications. Technologies need to be developed that will allow higher operating temperatures and/or more effective heat recovery systems and improved system designs to enable CHP efficiencies exceeding 80%.
- Infrastructure: unfamiliar technology to power industry, with no infrastructure in place.

22.17 Fuel cell summary

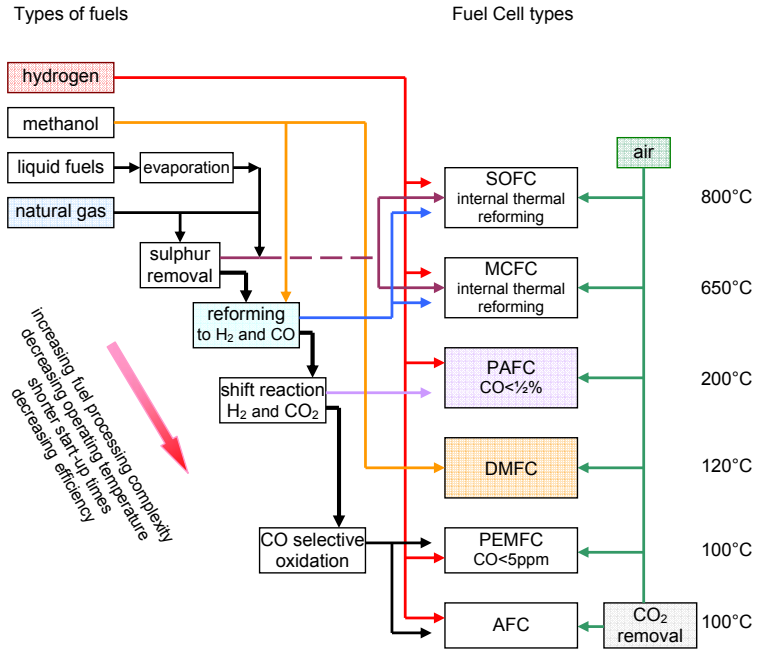


Figure 22.12. *Fuel cells operating temperatures and applicable fuels.*

The following table lists the key manufacturers by fuel cell type:

Table 22.8: Fuel cell stack manufacturers

PEM	Ballard, Plugpower, H-Power, UTC, Nuvera, Siemens, h2-interpower
AFC	Astris, UTC, Zetek
DMFC	Ballard, Manhattan Scientific, Medis, MTI Micro FuelCells, Smartfuelcell
PAFC	UTC
MCFC	Fuel Cell Energy (FCE), MTU
SOFC	Sulzer Hexis, Siemens Westinghouse, Global Thermolectric

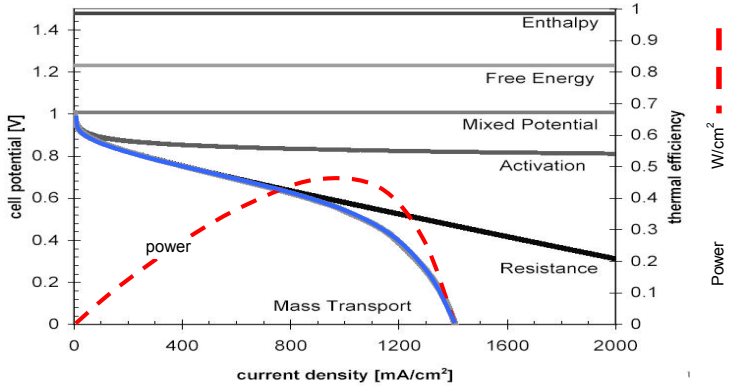


Table 22.9: Thermodynamic data for common fuel cell reactions at 1 atm pressure and 298K

Species	*Molecular weight*	ΔH_{298K}	ΔG_{298K}	ΔS_{298K}	*Freezing point*	*Boiling point*	*Enthalpy @ 25°C*	*Heat of vaporisation*	*Liquid density*
	g	kJ/mol	kJ/mol	J/mol.K	°C	°C	kJ/mol	kJ/kg	kg/l
H_2 (gas)	2.02	0	0	130.6	-259.2	-252.77	241.8	445.6	77
0_2 (gas)		0	0	205.0					
H_20 (gas)		-241.8	-228.6	188.7					
H_20 (liquid)		-285.9	-237.2	70.0					
H_20_2 (liquid)		-187.6	-120.2	109.5					
CH_4 (gas)	16.04	-74.9	-50.8	186.1	-182.5	-161.5	802.5	510	425
C (graphite)		0	0	5.7					
C_2H_6 (gas)		-84.7	32.9	229.5		-42.1			
C_3H_8 (gas)	44.1	-103.8	-23.5	269.9					
CH_3OH (gas)	32.04	-201.2	-161.9	237.7	-97.8	64.7	638.5	1100	792
CH_3OH (liquid)	32.04	-238.6	-166.2	126.8	-97.8	64.7	638.5	1100	792
C_2H_5OH (liquid)	46.07	-277.7	-174.8	160.7	-117.3	78.5	1275.9	855	789
$C_6H_{12}O_6$ (glucose)		-126.8	-91.0	212					
C_8H_{18}	114.2				-56.8	125.7	5512.0	368.1	368.1
NH_3 (gas)	17.03	-46.1	-16.45	192.5	-77.7	-33.4	316.3	1371	674
CO (gas)		-110.5	-137.2	197.5					
CO_2 (gas)		-393.5	-394.4	213.7					
KOH (solid)		-425.8	-380.2	79.3					
NO (gas)		377.8	362.3	881.6					
NO_2 (gas)		138.9	213.4	1003.7					

Table 22.10: Comparison of Fuel Cell Technologies

Ion & Fuel Cell Type	Common Electrolyte	anode cathode	Operating T °C Electo-catalyst	Density/power efficiency	Applications	Features	Shortcomings
H^+ Polymer Electrolyte Membrane (PEM)	Hydrated Solid organic polymer poly-perfluoro-sulphonic acid or Polymer membrane (ionomer) Poly-benzimidazole fibre	Carbon or metal	H_2 40 - 100°C platinum (Nafion) 50–120°C (PBI) 125–220°C	4 - 7 kW/m^3 <1kW–500kW Cell: 50–70% 53–58% (transportation) 25–35% (stationary)	•Backup power •Portable power •Small distributed generation •Transportation •1A/cm^2 @ 0.7V	•Carbon or metal electrodes •Solid electrolyte reduces corrosion and electrolyte management problems •Low temperature •Fast start-up, second/min •Solid long-life construction •non-volatile electrolyte	•Expensive platinum catalyst •CO poisoning •High sensitivity to fuel impurities •Low temperature waste heat not suitable for Cogeneration •External fuel processor •slow O_2 kinetics
OH^- Alkaline (AFC)	Aqueous (50%) alkaline solution of KOH potassium hydroxide soaked in an asbestos matrix	Anode Ni (transition) Cathode Lithiated NiO	H_2 65 - 160°C Ni, Ag, $MetalO_2$, noble metals	1 kW/m^3 1kW– 100kW Cell: 60 - 70% System: 62%	•Military •Space •1A/cm^2 @ 0.7V	•Platinum alternatives •Metal electrodes •Cathode reaction faster in alkaline electrolyte, higher performance	•Expensive removal of CO_2 from fuel and air streams required (CO_2 degrades the electrolyte) •Caustic medium, but low cost •External fuel processor, large size •No cogen heat
H^+ Direct methanol (DMFC)	Polymer membrane (ionomer) 200µm Similar to PEM	Carbon or metal 100µm stable	CH_3OH 50–100°C Pt Pt/Ru reduces anode poisoning	0.6 kW/m^3 100 kW–1 MW Cell: 30–45% System: 25%	•Portable power •Transportation •0.4A/cm^2 @ 0.7V & 60°C	•no reformer needed •Compact design •Fuel flexibility •Fast start-up, seconds/min •Liquid fuel	•×10 platinum of PEMF •Low efficiency, increases with temperature, poor cogen •External fuel processor •Low practical voltage, 0.6V
H^+ Phosphoric Acid (PAFC)	Immobilized molten liquid 100% phosphoric acid H_3PO_4 soaked in a SiC matrix	PTFE bonded Pt/ Graphite	H_2 150 - 205°C Pt or platinum alloys	1 - 2 kW/m^3 50kW –11MW (250kW/module) Cell: 55% System: 40% Co-Gen: 90%	•Distributed generation •0.8V/cell, 0.4A/cm^2@205°C	•High efficiency with Cogen •electrolyte rejects CO_2 •Increased tolerance to impurities in hydrogen, up to 1.5% CO	•Expensive platinum catalyst •Low current and power •Corrosive electrolyte •Slow start-up, hours •External fuel processor, only H_2 •large size and weight
CO_3^{2-} Molten Carbonate (MCFC)	Molten liquid lithium, sodium (62%-38%), and/or potassium carbonates $NaHCO_3$ (60%-40%) soaked in $LiAtO_2$ matrix	Anode Ni-Cr Cathode NiO-MgO	H_2 , CH_4 550 - 700°C Porous, electrode anode Ni cathode NiO	2 - 3 kW/m^3 1kW – 10MW (250kW/module) Cell: 55% System: 47% Co-Gen: 80%	•Electric utility •Large distributed generation •0.15A/cm^2 @ 0.8V & 600°C	•High efficiency with co-gen •Fuel flexibility CO/CH_4 etc. •Use a variety of catalysts •Internal fuel processor	•High temperature speeds corrosion & breakdown of cell components •Complex electrolyte management •Slow start-up, hours •Durability, limited life, S intolerant
O^{2-} Solid Oxide (SOFC)	Ceramic 40µm thick (Perovskites) Solid zirconium dioxide ZrO_2 stabilised with Yttria Y_2O_3 (>750°C)	Anode Ni-ZrO_2 Cermet (perovskite) Cathode $LaMnO_3$+ Sr	H_2 , CH_4, CO 650 - 1000°C Cermet catalyst is electrode	0.2 - 2 kW/m^3 5kW – 3MW 40–45% Co-Gen: 80%	•Auxiliary power •Electric utility •Large distributed generation •1A/cm^2 @ 0.7V	•Electrode is catalyst •High efficiency with Cogen •Fuel flexibility, S insensitive •Variety of catalysts •Solid electrolyte reduces electrolyte problems •solid state - flexible shape •0.7V/cell, 4kA/m^2 @ 990°C	•High temperature enhances corrosion and breakdown of cell components •Expensive materials •Start-up: planar type -hours tubular - minutes •High conductivity ceramic electrolyte brittleness with thermal cycling

22.18 Photovoltaic Cells: Converting Photons to Electrons

Photovoltaic, PV, as the word from the Greek implies, phos means light and voltaic is after Alessandro Volta, now taken to mean electricity, is the conversion of sunlight directly into electricity.
The photoelectric effect is the physical atomic level process by which a photovoltaic (PV) cell converts sunlight photon energy directly into electricity. When light shines on a PV cell (also called a solar cell), it may be reflected, absorbed or transmitted. Only the absorbed light photons may generate electricity.

22.19 Silicon structural physics

Single-crystal silicon is used extensively in PV cells.
The silicon atom has 14 electrons, arranged in three orbital shells. The inner two shells, those closest to the nucleus, are full. The outer shell, however, is only half full/empty, having only four electrons. A tetravalent silicon atom will attempt to fill its outer shell with eight electrons. To do this, it will covalently share electrons with four of its neighbouring silicon atoms, therein forming part of a regular lattice.
Pure silicon is a poor conductor of electricity as the covalently bonded electrons are not free to move, being tied in the crystalline structure. Atoms of elements such as boron or phosphorus used as doping agents in silicon, alter the electron conductivity.
If a silicon atom in the lattice is substituted by a group V atom, a phosphorus atom for instance, one of its five valence electrons is not involved in the tetravalent bonds; as a result of thermal agitation, it soon moves to the conduction band, thus becoming free to move randomly through the crystal, leaving behind a hole, bound to the doping atom. This is electron conduction, and the semiconductor is designated an *n-type* doped semiconductor. An *n-type* semiconductor has an abundance of electrons, which have a negative electrical charge.
If a silicon atom is substituted by an atom from group III, boron for instance, which has three valence electrons, one electron is missing from a possible tetravalent bonding state. If all bonds are to be maintained, an electron may move to fill this gap. A hole thus arises elsewhere, contributing to conduction, and the semiconductor is termed a *p-type* doped semiconductor. A *p-type* semiconductor has an abundance of 'holes', which have a positive electrical charge.
Although both p-type and n-type materials are electrically neutral, n-type silicon has excess electrons and p-type silicon has excess holes. Homogeneously melding of the two materials creates a p-n junction at their metallurgical interface. The excess electrons move from the n-type side to the p-type side. The result is a build-up of positive charge along the n-type side of the interface and a build-up of negative charge along the p-type side, thereby inducing an electric field across the two neutral charge regions, which form a space charge or charge depletion layer straddling the p-n junction charge boundary.
The silicon diode is electrically neutral at the junction, where the electric field separates and maintains the two sides at balance. The electric field acts as a diode in which electrons can only move in one direction. Placing metallic contacts on the n and p regions forms a diode.

Silicon PV cell physics
PV cells are in fact large area semiconductor p-n diodes and single-crystal silicon type PV cells are the most widely used. When light, in the form of photons, illuminates the junction region, its energy may liberate electron-hole pairs.
Photons having an energy equal to, or higher than, the width of the band gap, yield their energy to the electrons, each photon causing an electron to move from the *valence band* to the *conduction band*, leaving behind a hole. The movement of electrons effectively means holes appear to move around the material, thus giving rise to an electron-hole pair.
The bandgap, E_G, is the energy difference between the conduction band and valence band.
The relationship between frequency f, incident photon energy, and resultant electrical energy is given by:

$$W_{photon} = hf = \frac{hc}{\lambda} = \frac{p}{\lambda} = qE_G = W_{electrical} \quad \text{(J)} \qquad (22.20)$$

where: h is Planck constant, 6.626×10^{-34} Ws^2 or 4.14×10^{-15} eVs
f is frequency of the photon, Hz
p is momentum, h/λ
q is the electron charge, 1.602×10^{-19}, C
The relationship between speed c, 2.998×10^{8} m/s, frequency f, and wavelength λ is

$$c = f\lambda \quad \text{(m/s)} \qquad (22.21)$$

If the free-state electron state occurs near the junction built-in electric field, or if a free electron and free hole appear in its field of influence, the field forces the electron to the n-side and a hole appears on the p-side. This causes further disruption of electrical neutrality, and if an external current path exists, electrons flow through the external path back to their original junction side, the p-side, to recombine with holes that the electric field has maintained there, the electrons doing work along the way. The electron flow is the current, and the cell's electric field creates a voltage, thus power is produced.

Visible light is only part of the electromagnetic spectrum, as illustrated in figures 22.14 and 22.22. Electromagnetic radiation is not monochromatic - it is made up of a range of different wavelengths, and therefore energy levels. Photon energy increases with increased frequency, as shown by equation (22.20). Sunlight can be separated into different wavelengths, as in a rainbow.
The photon energy of light varies according to the different wavelengths of the light. The entire spectrum of sunlight, from infrared to ultraviolet, covers a range of about ½eV to about 2.9eV. The radiation spectrum has a peak at 0.8eV. Red light has energy of about 1.7eV, and blue light has energy of about 2.7eV. Most PV cells cannot use about 55% of the energy of sunlight, because this energy is either below the bandgap of the photoelectric material or carries excess energy.
The single p-n junction, generates only one electron-hole pair for each absorbed photon, and wastes as heat any incoming photon energy in excess of the semiconductor band gap. Each photon can produce only one electron-hole pair. The basic conversion process is shown in figure 22.13.
Different PV materials have different energy band gaps. Photons with energy equal to the band gap energy are absorbed to create free electrons. Photons with less energy than the band gap energy are re-emitted and pass through the material as if it were transparent and do not produce any electrical current or heat. Still other photons have too much energy. Only a certain amount of energy, the band gap energy, measured in electron volts (eV) and defined by the cell semiconductor material (about 1.11 eV or a wavelength of 1.13μm for crystalline silicon), is required to knock an electron loose. If a photon has more energy than the required amount, then the extra energy is lost as heat in the PV cell. These two effects alone account for the loss of around 70 percent of the radiation energy incident on a cell.
The band gap also determines the voltage of the electric field, and if this voltage is too low, then the extra current gained (by being able to absorbing more photons), is offset by having a small voltage, since power is voltage times current. The optimal band gap, balancing these two effects, is around 1.4 eV for a cell made from a single semiconducting material.
The most energy that a single junction cell can absorb is around 25 percent, with 15 percent or less being mo

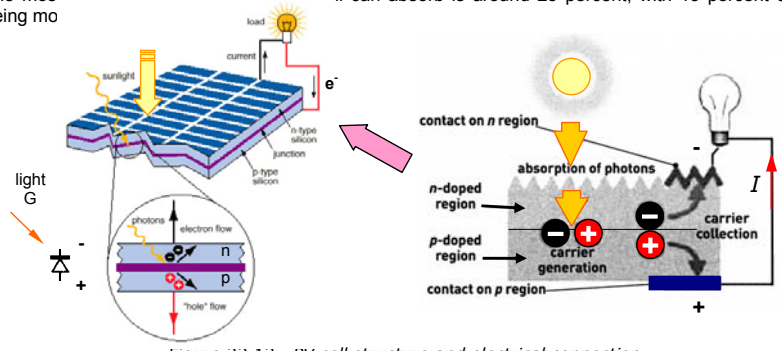


Figure 22.13. *PV cell structure and electrical connection.*

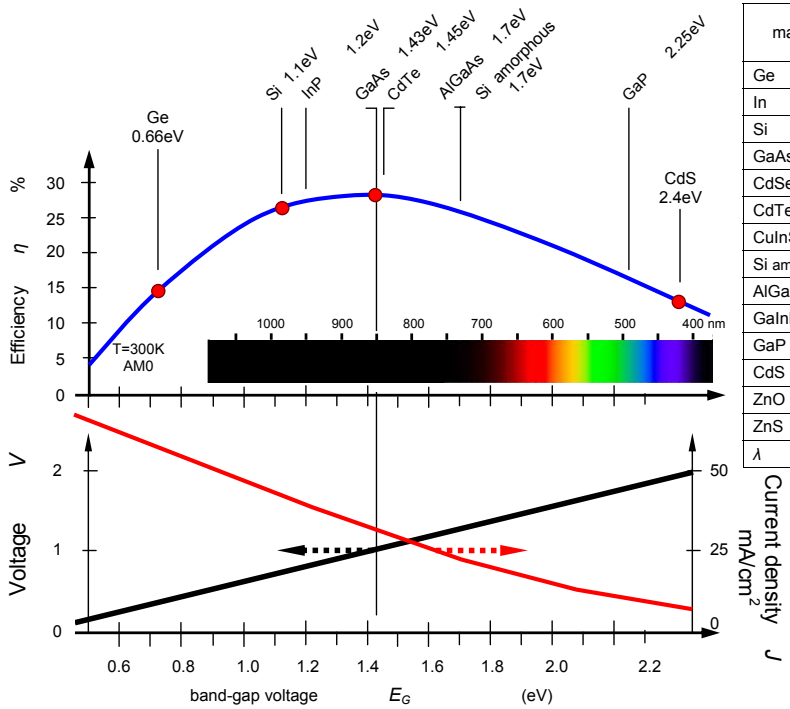
22.20 Semiconductor materials and structures

Single-crystal silicon is not the only material used in PV cells, as indicated in figure 22.15.
Polycrystalline silicon is used in an attempt to reduce manufacturing costs, although the resulting cells are not as efficient as with single crystal silicon. Amorphous silicon, which has no crystalline structure, is also used, again in an attempt to reduce costs. Other materials used include gallium arsenide, copper indium diselenide, and cadmium telluride. Since different materials have different band gaps, they are responsive to different frequencies, or photons of different energies, as given by equation (22.20).
Efficiency is improved by cascading two or more p-n junction layers of materials with different band gaps. The higher band-gap p-n material is on the surface, absorbing high-energy photons while allowing lower-energy photons to be absorbed by the lower band-gap material beneath. This technique results in much higher efficiencies. These so-called *multi-junction cells* have more than one electric field in series.

Silicon is not the only material that will respond to sunlight by generating electron-hole pairs of charge carriers. Any two dissimilar semiconductor materials will form a junction and an electric field at their interface. There are many design choices to be made in making a PV device, both in material properties and device designs. The most important design parameters to consider are:

- The material's electronic properties, especially its degree of crystallinity
- The amount of light that can be absorbed in a given thickness of material; its absorptivity
- The range of wavelengths in the sun's spectrum that can be absorbed and used; the band-gap
- The expected cost, due to the amount and type of material used and the complexity of the manufacturing process.

Single-crystal silicon has a high degree of crystallinity, which makes a high sunlight-to-electricity conversion efficiency relatively easy to attain. Its absorptivity is relatively low, however, which is why older silicon cell designs required about 300µm of semiconductor material thickness to absorb all the sunlight possible, and newer silicon cell designs still require about 100µm of material. Silicon's band gap of 1.1eV is not ideal. Single junction PV cells with the highest theoretical efficiencies employ materials with band gaps of 1.4 to 1.5eV, as shown in figure 22.14, although sunlight goes up to 4eV.



material	band-gap eV @ 300K
Ge	0.66
In	0.7
Si	1.11
GaAs	1.43
CdSe	1.74
CdTe	1.44
$CuInSe_2$	1.0
Si amorphous	1.7
AlGaAs	1.7
GaInP	1.8
GaP	2.25
CdS	2.4
ZnO	3.2
ZnS	3.6
λ nm	$1241/E_G$

Figure 22.14. *Calculated single solar cell efficiency for different band-gap semiconductors.*

PV cells can therefore be made from a wide range of semiconductor materials. The following sections discuss the various PV cell possibilities and semi-conducting material combinations:

- Silicon, Si - including
  - single-crystalline Si,
  - multicrystalline Si, and
  - amorphous Si
- Polycrystalline thin films - including copper indium diselenide (CIS), cadmium telluride (CdTe), and thin-film silicon
- Single-crystalline thin films - including high-efficiency gallium arsenide (GaAs)
- Nanocrystalline PV cells

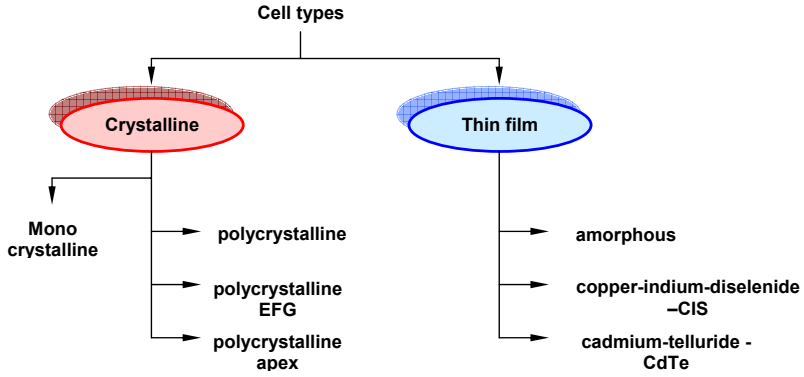


Figure 22.15. *Photovoltaic cell types.*

22.20.1 Silicon

Three types of silicon crystalline structures are introduce and considered in chapter 1.5, namely mono (single), poly (multi), and amorphous silicons.

i. Single-crystalline silicon

Mono-crystalline or single-crystalline PV cells and modules are made with the following process and production steps.

1. Casting
The raw material for making solar cells is monocrystalline silicon.
Silicon feedstock is melted in a crucible to form either monocrystalline or multicrystalline silicon, depending on the production process used.
During the melting process, a small quantity of boron is mixed to make the silicon p-type, giving it a positive electrical characteristic. It is made into ingots or bricks, which are cut to a more appropriate shape. After shaping, the silicon ingot, sawn with diamond saw into thin silicon wafers, is the foundation for PV cell production. Wafers of 1mm thickness sawn with 100µm precision are placed between two plane-parallel metal plates, which rotate into opposite directions. The procedure enables wafer thickness adjustment to 1µm precision.

2. Etching and texturing
The wafers are cleaned with industrial soaps and then etched a few micrometres to remove saw damage and crystalline-structure irregularities. Monocrystalline wafers are further etched in a hot solution of sodium hydroxide and isopropanol to form surface square-based pyramids, called texture. The texturization reduces the reflection of sunlight. This process is not particularly effective in the case of multicrystalline wafers, thus other texturing methods are used to enhance light capture.

3. Diffusion and edge isolation
Wafers that have been pre-doped with boron during the casting process are then given a negative n-type surface characteristic by diffusion with a phosphorus gaseous source at high temperature, 800°C, which creates the n-p junction. An oxide layer rich with phosphorus is formed on top of wafers due to oxygen subsequently introduced into the reaction.
Phosphorus diffuses not only into the desired wafer surface but also into the sides and the opposite surface, to some extent. This gives an edge shunt path between the cell front and rear. Removal of the shunt path at the wafer edge, called edge junction isolation or passivation, is performed by coin stacking the cells to form a cube and exposing the edges to an oxygen etch in a plasma etching chamber. The next phase removes the oxide layers from top of the wafer by wet chemical etching.

4. Anti-reflection coating
Pure silicon is a shiny grey material and can act as a mirror, reflecting more than 30% of the incident light. Photons that are reflected cannot be used by the cell. To improve the conversion efficiency of a PV cell, the amount of light reflected is minimised so that the semiconductor material can capture as much light as possible to use in liberating electrons. Thus an anti-reflective coating is applied to the top of the cell. A single layer reduces surface reflection to about 10%, and a second layer, termed a double-layer antireflection coating, lowers reflection to less than 4%.

Two techniques are commonly used to reduce incident light reflection.

- The first technique is to coat the top surface with a thin layer of silicon monoxide Si0 or titanium dioxide $Ti0_2$. The material used for coating is
  - heated until its molecules boil-off and travel to the silicon and condense, called plasma enhanced chemical vapour deposition. This process not only deposits an antireflective layer but also improves the electronic properties of the silicon by injecting hydrogen, or
  - the material undergoes sputtering. In this process, a high voltage knocks molecules off the material and deposits them onto the silicon at the opposite electrode, or
  - the silicon itself is allowed to react with oxygen or nitrogen gases to form silicon dioxide $Si0_2$ or silicon nitride Si_3N_4.
- A second technique is to texture the top surface, as performed at stage 2 and shown in figure 22.13. Chemical etching creates a pattern of cones and pyramids, which capture light rays that might otherwise be reflected away from the cell. Reflected light is redirected into the cell, where it can be absorbed. Light with an 80% probability of being absorbed by a flat surface has a 96% chance of being absorbed by a textured surface.

Commercial PV cell manufacturers tend to use silicon nitride as the anti-reflective coating.

5. Metallization and electrical contact placement

The cells are now capable of generating electricity. The front and the rear surfaces need to have contacts added, usually in the form of electrically conductive metal strips. Two metallisation techniques are common:

i. The back contact of a cell (away from the sun) is usually coated in a layer of aluminium or molybdenum metal. The front contact is a narrow grid of fingers so as not to block sunlight in to the cell. Metals such as palladium/silver, nickel, or copper are vacuum-evaporated, or sintered, through a photoresist, silk-screened, or merely deposited on the exposed portion of cells that have been partially covered with wax. In each case, the part of the cell on which a contact is not desired is protected, while the rest of the cell is exposed to the metallization.
ii. Alternatively, silver, with 1% aluminium, is a used metal for contact formation owing to its solderability. Silver in the form of a paste is screen printed onto the front and the rear. In addition, aluminium paste is applied to the rear to achieve the back surface field which improves the performance of the solar cell. These metal pastes are subsequently heated above their alloying temperature to form a good ohmic contact.

The fingers of the grid must be wide enough to conduct with low resistance, but narrow enough to block a minimal of incoming light, about 3% to 5% of the surface area.
After the contacts are in place, thin separating strips, normally tin-coated copper, are placed between cells.

6. Encapsulating the cell and solar modules

Since each solar cell produces about half a Volt when exposed to light, each PV module is made by connecting cells (usually 36) in series and parallel to achieve useful levels of voltage and current.
The PV module consists of the silicon semiconductor surrounded by protective material in a metal frame. The protection is an encapsulant of transparent silicon rubber or butyryl plastic bonded, which are then embedded in silicon rubber or ethylene vinyl acetate. The encapsulated PV cells are then placed into an aluminium frame where a polyester film such as Mylar or Tedlar makes up the backing sheet. A front glass cover is found on terrestrial arrays; a lightweight plastic cover on satellite arrays. Silicone is used as the cement to hold the structure together. The basic constructional arrangement is shown in figure 22.16.

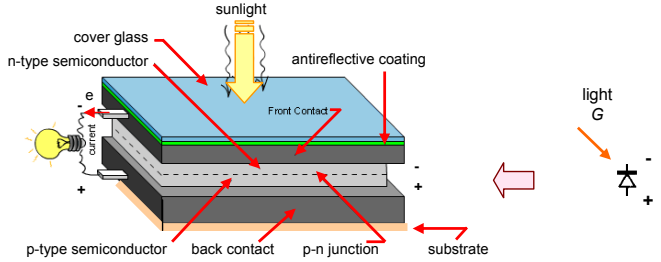


Figure 22.16. *Basic structure of a generic silicon p-n junction PV cell.*

ii. Multicrystalline silicon

Pure poly-crystalline silicon is extracted from tri-chlorine-silane. Furnaces are heated by electric current, which flows through silicon electrodes. The 2m long electrodes measure 8mm in diameter and the current flowing through the electrodes can reach up to 6000A. The furnace walls are additionally cooled preventing the formation of any unwanted reactions due to gas side products. The procedure results in pure poly-crystalline silicon which is used as a raw material for PV cell production. Poly-crystalline silicon can be formed from silicon by heating it up to 1500°C and then cooling it down to 1412°C, which is just above solidification of the material. The cooling produces an ingot of poly-crystalline silicon of dimensions 40x40x30cm, in which the silicon structure has non-uniform and irregular grain boundaries.

Free electrons and holes are much more likely to recombine at grain boundaries than within a single crystal. This recombination at the grain boundaries looks like an internal electrical resistance, hindering current flow, resulting in a loss of cell voltage. Thus the net effect of grain boundaries is to reduce the power output of a PV cell. Multicrystalline silicon cells therefore do not yield as high efficiencies as single-crystal silicon cells.

iii. Amorphous silicon

Amorphous silicon PV cells are produced with similar technological procedures as integrated circuits. Hence these cells are also known as thin-film PV cells, and thin-film modules. Amorphous PV cell production can succinctly be described as follows:

- Glass substrate is thoroughly cleaned.
- Lower contact layer is deposited.
- The surface is then structured - divided into bands.
- The amorphous silicon layer is applied in a vacuum, under a high frequency electric field.
- The surface is re-banded.
- Upper metal electrodes are deposited.

Amorphous non-crystalline silicon, like common glass, is a material in which the atoms are not arranged in any particular order. They do not form crystalline structures, and contain large numbers of structural and bonding defects.
Amorphous silicon absorbs PV radiation 40 times more efficiently than single-crystal silicon, so a film only about 1µm thick can absorb 90% of the usable light energy shining on it. For this reasons, amorphous silicon can reduce the cost of photovoltaics. Other economic advantages are that it can be produced at lower temperatures and can be deposited on low-cost substrates such as plastic, glass, and metal. This makes amorphous silicon ideal for building PV integration.
There is a short-range structure order in the sense that most silicon atoms tend to bond with four other silicon atoms at distances and angles nearly the same as in a crystal. But this uniformity does not translate any distance, resulting in deviations that disrupt the long-range order. This lack of order results in a high degree of defects such as dangling bonds, where atoms are missing a neighbour to which they can bond. These defects provide places for electrons and holes to recombine. Ordinarily, such materials are unacceptable for electronic devices because the defects limit the flow of current, in much the same way that grain boundaries interfere with the flow of current in polycrystalline material. But if amorphous silicon is deposited in such a way that it contains 5% to 10% hydrogen, then the hydrogen atoms combine chemically with some of the dangling bonds. This removes the dangling bonds and enhances the freedom of movement, or mobility, of electrons and holes in amorphous silicon (see chapter 1.5).

Because of amorphous silicon's unique properties, PV cells are designed to have an ultra-thin (0.008µm) p-type top layer, a thicker (0.5 to 1µm) intrinsic central layer, and a thin (0.02µm) n-type bottom layer. This design is called a *p-i-n* structure, because of the three layers, as shown in figure 22.17. The top layer is made so thin and relatively transparent that most light passes through it, to generate free electrons in the intrinsic layer. The p and n layers produced by doping the amorphous silicon create an electric field across the entire intrinsic region, thus inducing electron movement in the *i-layer*. Since the charge carriers are generated in the intrinsic layer, the poor charge mobility in the doped p+ and n+ layers is less important.
Amorphous silicon has a band-gap energy of about 1.7eV, which is greater than crystalline silicon's band-gap energy of 1.1eV. A PV cell's output voltage is directly related to its band gap magnitude, so PV cells made of amorphous silicon have higher output voltages than cells made of crystalline silicon. The higher output voltage compensates for the fact that lower-energy photons (with energies below 1.7eV) are not absorbed by amorphous silicon.

Cell construction

Typically, the construction of a hydrogenated amorphous silicon p-i-n cell is shown in figure 22.17, and begins with the deposition of a p+ layer onto a textured transparent conducting tin-oxide film - the top electrode. This is followed by the intrinsic i layer, and a thin n+ layer. The cell is then given a reflective,

conductive coating on the bottom, usually of aluminium or silver. A textured transparent conducting oxide substrate and the reflector helps trap the maximum possible amount of light in the PV cell.
After amorphous silicon modules are first exposed to light, their conversion efficiency decreases by 10% to 20%. The problem is eliminated by low temperature annealing. Self-annealing occurs in devices operated at about 50°C to 80°C.

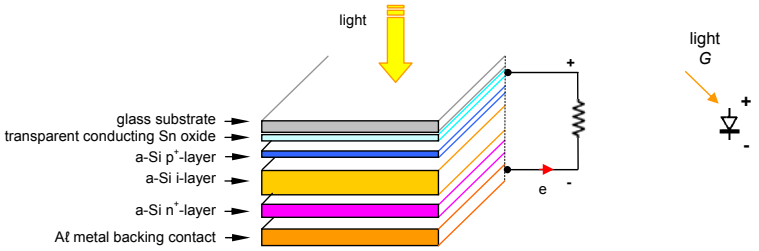


Figure 22.17. *A typical amorphous silicon p-i-n cell, with an intrinsic layer (i-layer) between a p-layer and an n-layer.*

Glow-discharge deposition is the method used to make the most efficient amorphous silicon PV cells. A stream of silane (SiH_4) and hydrogen gas is passed between a pair of electrodes whose polarity is reversed at high frequency. This voltage reversal induces an oscillation of energetic electrons between the electrodes. The electrons collide with the silane, breaking it down into SiH_3 and hydrogen. Since SiH_3 is chemically a radical, hence unstable, it adheres to a substrate on one of the electrodes to gain stability. Hydrogen is then released from the substrate, leaving a film of amorphous silicon with about 10% hydrogen. Doping is accomplished by adding diborane (B_2H_6) or phosphine (PH_3) gas to the silane.

Other PV cells

Among less frequently used PV cell types are cells produced by the EFG (Edge defined Film fed Growth) method and Apex PV cells from silicon, cadmium telluride PV cells and copper-indium selenide (CIS) PV cells. EFG monocrystalline PV cells are produced directly from silicon melt eliminating wafer sawing, which results in lower production costs and material saving for there is no waste due to sawing. Using the EFG procedure, a silicon ribbon tubular shaped with eight flat sides is drawn from silicon melt. The tube length extends several metres. The tube flat sides are laser sawn into separate PV cells. Most PV cells are square shaped, about 100x100mm.

The module power is greater with a smaller surface than crystal modules with truncated sides. The window surface contact strips are of copper.
Apex cells are poly-crystalline. Cadmium telluride and copper-indium selenide (CIS) cells are rarely used.

22.20.2 Polycrystalline thin-films

The *thin-film* term comes from the method used to deposit the film, not from the thinness of the film: thin-film cells are deposited in very thin, consecutive layers of atoms, molecules, or ions. They have many advantages over 'thick-film' counterparts. For example,

- they use much less material - the cell's active volume is usually only 1 to 10μm thick, whereas thick films are typically 100 to 300μm thick.
- thin-film cells can be manufactured in a large-area process, which can be an automated, continuous production process.
- they can be deposited on flexible substrate materials, such as flexible plastics, metal, and glass.

Unlike most single-crystal cells, the typical thin-film cell does not use a metal grid for the window-surface electrical contact. Instead, a thin layer of a transparent conducting oxide is used. Suitable oxides, such as tin oxide, indium tin oxide, and zinc oxide, are transparent and electrically conductive. They collect the current effectively from the top of the cell, and losses due to lateral resistance are minimal. A separate antireflection coating may be used on the top surface, or the transparent conducting oxide may also serve this function.
This thin-film flexibility offers potential for building applications.

Thin-Film Deposition
Several different deposition techniques can be used, and are less expensive than the ingot-growth techniques required for crystalline silicon. Deposition techniques can be broadly classify (of typically hydrogenated amorphous silicon) into physical vapour deposition, chemical vapour deposition, electrochemical deposition, or a combination. Individual 'cells' are formed by vertical laser scribing. Production is faster, cheaper, uses less energy, and is applicable to mass production. Like amorphous silicon, the layers can be deposited on various low-cost substrates (or 'superstrates') such as glass, stainless steel or plastic, in virtually any shape.
In addition, the deposition processes can be scaled up, which means that the same technique used to make a 5cm x 5cm cell can be used to make a 60cm x 160cm PV single-cell module. Thin-film devices can be made as a single monolithic unit with layer upon layer being deposited sequentially on some substrate, including deposition of an antireflection coating and the transparent conducting oxide.

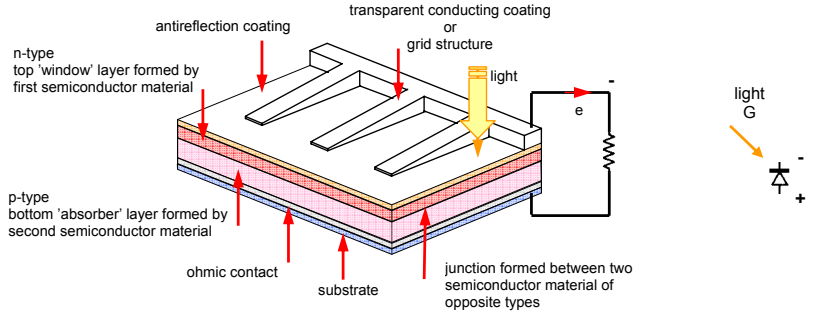


Figure 22.18. *A polycrystalline thin-film cell showing an n-type window and p-type absorber layer.*

Thin-Film Cell Structure
Polycrystalline thin-film cells are made of many tiny crystalline grains of semiconductor materials. An electric field is created at the interface between two different semiconductor materials. This interface is called a *hetero-junction* (hetero because the junction is formed from two different semiconductor materials, as opposed to the *homo-junction,* meaning formed by two doped layers of the same material).
The typical polycrystalline thin film, as shown in figure 22.18, has a thin (less than 0.1μm) n-type layer on top called the 'window' layer. Its role is to absorb light energy from only the high-energy end of the spectrum. It is thin enough and has a wide enough bandgap (2.8eV or more) to let all the available light through the interface (hetero-junction) to the p-type absorbing layer. The absorbing layer under the window is typically only 1 to 2μm thick and has a high absorptivity (ability to absorb photons) for high current and a suitable band gap to provide an adequate voltage.
An ohmic contact is often used to provide a good electrical connection to the substrate. But the top contact uses the thin transparent conducting oxide layer.

Three types of polycrystalline thin-film PV cells are:

- Copper indium diselenide (CIS)
- Cadmium telluride (CdTe)
- Thin-film silicon

i. Copper Indium Diselenide (CIS)
Copper indium diselenide ($CuInSe_2$ or 'CIS') has an extremely high absorptivity, which means that 99% of the light shining on CIS will be absorbed in the first micrometre of the material. It is stable, with no degradation due to environmental exposure. Cells made from CIS are usually heterojunction structures - structures in which the junction is formed between semiconductors having different bandgaps. The most common material for the window layer is cadmium sulphide (CdS), although zinc is can be added to improve transparency. Adding small amounts of gallium to the lower absorbing CIS layer boosts its bandgap from a normal 1.0eV, which improves the voltage and therefore the efficiency of the cell. This variation is called a copper indium gallium diselenide or 'CIGS' PV cell.
CIS is an efficient, about 17%, but complex thin film material that is difficult to manufacture, a process that involves the toxic gas hydrogen selenide, as indicated in figure 22.19.

CIS element combinations		
$(I)(III)(VI)_2$		
Group I	Group III	Group VI
Cu	Al	S
Ag	Ga	Se
Au	In	Te

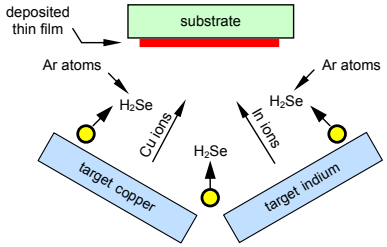


Figure 22.19. *Typical evaporated formation of a CIS PV cell.*

The CIS cell layers are formed by several different processes. The CIS layer consists of three different elements, specifically copper, indium, and selenium. Preparation methods for the CIS layer include:

- *Evaporation*. Small amounts of each of the elements are electrically heated to a point where the atoms vaporize. They then condense on a cooled substrate to form a CIS layer.
- *Sputtering*. High-energy ions bombard the surface, driving off atoms of the target material. These then condense on a substrate to form a thin layer.
- *Pyrolysis*. Copper indium diselenide is deposited on a substrate by spray pyrolysis. Solutions of the salts of the necessary elements are sprayed onto a hot substrate. They react under elevated temperatures to form the required CIS layer, while the solvent evaporates.
- *Electrodeposition*. The CIS layers are deposited in the same way precious metals are plated. Passing electricity through a solution containing ions of the required elements causes them to be deposited out of solution onto an electrode, which acts as the substrate.

CIS layers can be made using any one of these methods to deposit only the copper and the indium. This is followed by a treatment with hydrogen selenide gas, called selenization, to add the selenium.

ii. Cadmium Telluride (CdTe)

Cadmium telluride is a polycrystalline thin-film material produced by deposition or sputtering. A disadvantage is that poisonous production materials are used. PV cells efficiency is up to 10%.

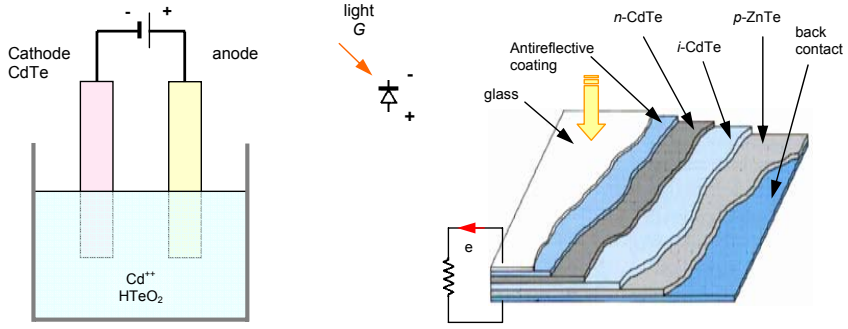


Figure 22.20. *A CdTe PV cell:*
(a) formation by electrolysing deposition and (b) the basic layer structure.

With a near ideal bandgap of 1.44eV, CdTe also has a high light absorptivity. Although CdTe is most often used in PV devices without being alloyed, it is easily alloyed with zinc, mercury, and a few other elements to vary its properties. Like CIS, films of CdTe can be manufactured using low-cost techniques. Also like CIS, the best CdTe cells employ a heterojunction interface, with cadmium sulphide (CdS) acting as the thin window layer. Tin oxide is used as a transparent conducting oxide and antireflection coating. The p-type CdTe films are highly resistive electrically, which leads to significant internal resistance losses. The back electrical contact tends to suffer rust-like deterioration. This high resistivity

problem is circumvented by the CdTe layer being intrinsic (that is, neither p-type nor n-type, but undoped), and a layer of p-type zinc telluride (ZnTe) is added between the CdTe and the back electrical contact. Although the n-type CdS and the p-type ZnTe are separated, as shown in figure 22.20b, they still form an electrical field that extends through the intrinsic CdTe. A wide variety of CdTe production methods are possible, including closed-space sublimation, electro-deposition (figure 22.20a), and chemical vapour deposition. Such cells designs show no degradation after 3,000 hours of service.
CdTe contacts are more stable if properly protected from water vapour and oxygen by hermetically sealing. Cadmium is an accumulative toxic heavy metal.

iii. Thin-Film Silicon

The term 'thin-film silicon' typically refers to silicon-based PV devices other than amorphous silicon cells and single-crystalline silicon cells, where the silicon layer is thicker than 200µm. These thin films have a high light absorptivity and require cell thicknesses of only a few micrometers. Nanocrystalline silicon and small-grained polycrystalline silicon - considered thin-film silicon - can replace amorphous silicon alloys as the bottom cell in multi-junction devices. As with other thin films, advantages include material savings, monolithic device design, use of inexpensive substrates, and manufacturing processes that are low temperature and possible over large areas.

22.20.3 Single-Crystalline Thin Film

Single-crystal thin-film GaAs

To be cost-effective, gallium arsenide GaAs high-efficiency cells are best suited for concentrator systems. GaAs offers high light absorptivity and high energy conversion efficiency, up to 25%, and up to 28% with concentrated solar radiation.
Gallium arsenide (GaAs) is a compound semiconductor: a mixture of two elements, gallium and arsenic. Gallium is a by-product of the smelting of metals, notably aluminium and zinc, and is rarer than gold. Arsenic is not rare, but is poisonous.
GaAs is suitable for use in multi-junction and high-efficiency PV cells, for several reasons:

- The GaAs bandgap is 1.43 eV, ideal for single-junction PV cells.
- GaAs has a high absorptivity and requires a cell only a few microns thick to absorb sunlight. Crystalline silicon requires a layer of at least 100µm.
- Unlike silicon cells, GaAs cells are relatively insensitive to heat. Cell temperatures can often be high in concentrator applications.
- Alloys made from GaAs and aluminium, phosphorus, antimony, or indium have characteristics that are complementary to those of gallium arsenide, allowing cell design flexibility.
- GaAs is resistant to radiation and along with its high efficiency, GaAs is desirable for space applications.

A GaAs base cell can have several layers of different compositions; allowing precise control of the generation and collection of electrons and holes.
In silicon cells, the p-n junction is formed by treating the top of a p-type silicon wafer with an n-type dopant. This diffusion process does not work with gallium arsenide. Instead, all the active parts of a GaAs PV cell are made by growing sequences of thin, single-crystal layers on a single-crystal substrate. As each layer is grown, it is doped in different ways to form the p-n junction and to control other aspects of cell performance. This control facilitates cell efficiencies closer to theoretical levels. Common GaAs cell structures use a thin window layer of aluminium gallium arsenide. The thin layer results in the charge carriers being created close to the electric field across the junction.

The GaAs layers are grown via one of two techniques:

- In molecular beam epitaxy, a heated substrate wafer is exposed to gas-phase atoms of gallium and arsenic that condense on the wafer growing a thin GaAs film, or
- In metal-organic chemical vapour deposition, a heated substrate is exposed to gas-phase organic molecules containing gallium and arsenic, which react under the high temperatures, freeing gallium and arsenic atoms to adhere to the substrate.

In each case, single-crystal GaAs layers grow epitaxially with the new atoms deposited on the substrate continuing the same crystal lattice structure as the substrate, resulting in a high degree of crystallinity and in high cell efficiency.

GaAs Design Challenges

GaAs cells use a costly single-crystal GaAs substrate. GaAs cells are normally used in concentrator systems, each concentrator cell measuring about ¼cm^2 in area. In this configuration, the cost is sufficiently low to make GaAs cells competitive, assuming that module efficiencies are better than 25% and that the PV ancillary system is cost-effective.

Cheaper substrates result by growing GaAs cells on a removable, reusable GaAs substrate or making GaAs thin films, similar to those made of copper indium diselenide and cadmium telluride.
Growing quality GaAs crystals for PV cells requires a substrate with a crystal structure matching that of gallium arsenide and with similar thermal expansion properties. Silicon and germanium substrates can be used which are cheaper and more durable than gallium arsenide. But the slight mismatch in crystal structures between Si or Ge and GaAs causes imperfections in the GaAs crystal growth. So a buffer layer of gallium arsenide is deposited between the silicon and the active GaAs cell. The buffer layer progressively rectifies the imperfection effects in the new crystal growth structure. Another method is to treat cells with hydrogen, as with semi-crystalline silicon, which lessens the effects of imperfections and dangling bonds, but does not address the thermal mismatch aspect.

22.20.4 Nanocrystalline

These structures use thin-film light absorbing materials but are deposited as an extremely thin absorber on a supporting matrix of conductive polymer or mesoporous metal oxide, resulting in a high surface area with increased internal reflections, hence increased light absorption.
PV cells based on a silicon substrate with a nanocrystal coating are called quantum dots (electron confined nanoparticles), and increase the efficiency and reduce the cost of silicon photovoltaic cells. Quantum dots of lead selenide PbSe can produce as many as seven excitons from one high energy photon of sunlight (7×1.11eV = 7.8 times the bandgap energy). Conventional photovoltaic cells produce one exciton per high-energy photon, with high kinetic energy carriers losing their energy as heat. This multiple excitation does not result in a 7-fold increase in output, but boosts the maximum theoretical efficiency to over 40%. Quantum dot photovoltaics are cheap to manufacture, as they are made using simple chemical reactions.

Table 22.11: PV cell general characteristics

Characteristics of photovoltaic cells					
structure	material	Electronic material properties	absorptivity	band gap	cost
single junction	single crystal silicon	high	low	medium	high
	polysilicon silicon	medium	low	medium	medium
	amorphous silicon	low	high	medium	low
	single crystal thin film	high	high	high	high
	polycrystalline thin film	medium	high	medium-high	low
multi-junction	silicon	low-high	high	high	low-high

Table 22.12: PV cell general properties

	Efficiency	Durability	Comments
Monocrystalline	13% - 15%	> 30 years	• Highest efficiency = least surface area. • Expensive due to complicated manufacturing process. • Highest achieved efficiency is 23%. • Generate approximately 35mA/cm^2 at a voltage of 550mV at full illumination.
Polycrystalline	10% - 13%	> 25 years	• Most commonly used type of cell as it offers good efficiency at reasonable cost.
Thin Film / Amorphous	5% - 7%	> 20 years	• Low efficiency hence requires large surface area. • Made from flexible material. • Better in diffused light than mono and polycrystalline. • Current density of up to 15mA/cm^2, and the cell open-circuit voltage of 0.8V, is more that a crystalline cell. • Spectral response reaches maximum at the wavelengths of blue light therefore, ideal with fluorescent light sources.

The properties and features of the different PV technologies are summarised in tables 22.11 and 22.12.

22.21 PV Cell Structures

Most PV devices use a single junction to create an electric field. In a single-junction PV cell, only photons with energy equal to or greater than the band gap of the cell material can elevate an electron into the conduction band. In other words, the photovoltaic response of single junction cells is limited to the portion of the sun's spectrum with energy above the band gap of the absorbing material. Lower-energy photons are not used.
One way around this limitation is to use two or more cells, with more than one band gap and more than one junction, to generate a series cell voltage. These structures are referred to as multi-junction cells. Multi-junction cells can achieve a higher total conversion efficiency because they can convert more of the energy spectrum of light to electricity.
The actual structural design of a photovoltaic cell depends on the limitations of the material used in the PV cell. There are four basic device designs commonly used with the semiconducting materials.

- Homojunction
- Heterojunction
- p-i-n/n-i-p
- Multi-junction

22.21.1 Homojunction Device

A single semiconductor material, crystalline silicon, is altered so that one side is p-type, dominated by positive holes, and the other side is n-type, dominated by negative electrons. The p-n junction is located so that the maximum amount of light is absorbed near it. The free electrons and holes generated by the light in the silicon diffuse to the p-n junction, then separate to produce an external current.
In the homojunction design, several cell aspects are varied to increase the conversion efficiency:

- Depth of the p-n junction below the cell's surface
- Amount and distribution of dopant atoms on either side of the p-n junction
- Crystallinity and purity of the silicon

Numerous homojunction Si cell examples have been presented in previous sections of this chapter.

22.21.2 Heterojunction Device

An example of this type of device structure is a CIS cell (figure 22.19), where the junction is formed by contacting two different semiconductors, CdS and $CuInSe_2$. This structure is often used for thin-film cells, which absorb light much better than silicon. The top and bottom layers in a heterojunction device have different functions. The top layer, or 'window' layer, is a material with a high bandgap selected for its light transparency. The window allows almost all the incident light to reach the bottom layer, which is a material with a low bandgap that readily absorbs light. This light then generates free electrons and holes near the junction, which separates the electrons and holes before they can recombine.
A high band-gap window layer reduces the cell's series resistance. The window material can be made highly conductive, and the thickness can be increased without reducing the transmittance of light. Light-generated electrons can therefore readily flow laterally in the window layer to the electrical contact.

22.21.3 p-i-n and n-i-p Devices

Typically, amorphous silicon thin-film cells use a p-i-n structure, whereas CdTe cells use an *n-i-p* structure. A three-layer sandwich is created, with a middle intrinsic (i-type or undoped) layer between n-type and p-type layers. This geometry sets up an electric field between the p and n type regions that breaches the middle intrinsic resistive region. Light generates free electrons and holes in the intrinsic region, which are then separated by the electric field.

In the p-i-n amorphous silicon cell shown in figure 22.17, the top layer is p-type amorphous silicon, the middle layer is intrinsic silicon, and the bottom layer is n-type amorphous silicon. Amorphous silicon has many atomic-level electrical defects when it is highly conductive. Little current flows due to diffusion. However, in a p-i-n cell, current flows because the free electrons and holes are generated within the influence of an electric field, rather than having to move toward the field. Carrier lifetimes are long in the intrinsic layer.

In a CdTe cell shown in figure 22.20, the device structure is similar to the amorphous silicon i cell, except the order of layers is inverted. Specifically, in a typical CdTe cell, the top layer is p-type cadmium sulphide (CdS), the middle layer is intrinsic CdTe, and the bottom layer is n-type zinc telluride (ZnTe).

22.21.4 Multi-junction Devices

A multi-junction structure can achieve a higher total conversion efficiency by capturing a larger portion of the solar spectrum. In the typical multijunction cell shown in figure 22.21, individual cells with different bandgaps are stacked in descending bandgap order beneath one-another. The sunlight falls first on the material having the largest bandgap E_{G1}. Photons not absorbed in the first cell are transmitted to the second cell E_{G2}, which then absorbs the higher-energy portion of the remaining solar radiation while

remaining transparent to the lower-energy photons. These selective absorption processes continue through to the final cell, which has the smallest bandgap E_{G3}.

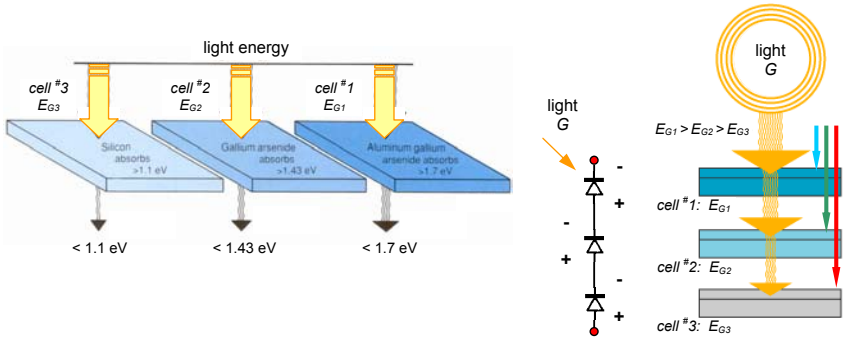


Figure 22.21. *A typical multi-junction PV cell showing three progressively decreasing band gap cells.*

A multi-junction cell can be made in two different ways.

- In the mechanical stack approach, two individual PV cells are made independently, one with a high bandgap and one with a lower bandgap. Then the two cells are mechanically stacked, the one with the highest bandgap on top of the other.
- In the monolithic approach, one complete PV cell is made first, and then the layers for the second cell are grown or deposited directly on the first cell.

i. The multi-junction device in figure 22.22 has a top cell of gallium indium phosphide, then a *tunnel junction* to allow the flow of electrons between the cells through a thin insulating layer, and a bottom cell of gallium arsenide.

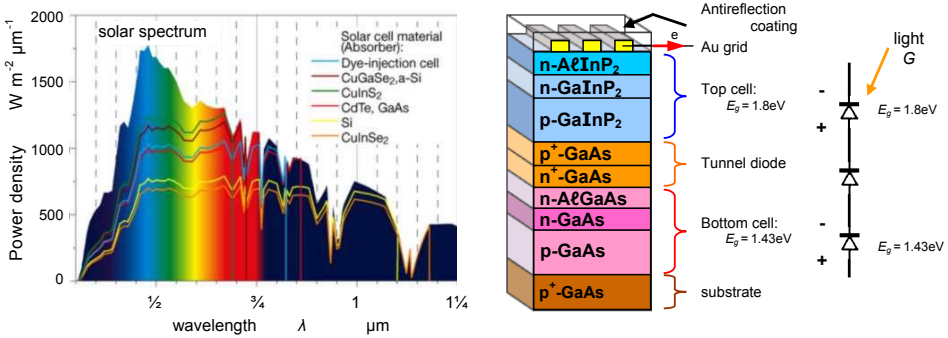


Figure 22.22. *A monolithic multi-junction PV cell based on GaAs and GaInP.*

ii. Another monolithic cell, a device that uses indium phosphide for the top cell and indium gallium arsenide for the bottom cell (InP/InGaAs), reaches 31.8% efficiency under 50 suns concentration. There are several aspects about this device.

- Traditionally, for highest efficiency in a two-junction device, the top cell should have a high band gap of approximately 1.9eV while the bottom cell should have a band gap of about 1.4eV. For this device, the band gap of the top cell is 1.35eV and that of the bottom cell is about 0.75eV.
- The cell is ideally suited for space applications because the resistance of indium phosphide to radiation is 50% better than that of silicon and 15% better than that of gallium arsenide, which are the two materials used for PV power in space. This means that arrays using InP/InGaAs cells are lighter, cheaper, more reliable, have longer lifetime, and are more powerful.

- The band gap of indium gallium arsenide can be altered by varying the ratio of the constituent materials. This offers possibilities for spectrally tuning the cell.

High efficiency multi-junction cells focus on gallium arsenide as one, or all, of the component cells which gives efficiencies of more than 35% under concentrated sunlight. Other multi-junction cells include the use of amorphous silicon and copper indium diselenide.

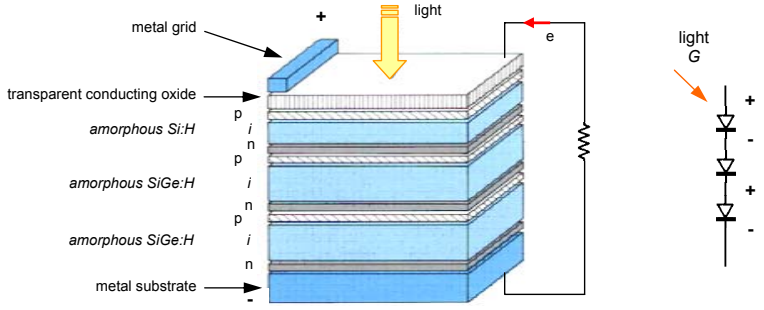


Figure 22.23. *An amorphous silicon multi-junction p-i-n type PV cell.*

i. Amorphous silicon multi-junctions

Amorphous silicon cells, include silicon carbon alloyed p-i-n cells, n-i-p cells, and stacked cells. Multi-junction devices not only achieve higher efficiencies than single-junction cells, but they also experience less light induced degradation. Because the intrinsic layers are so thin, the electric field sweeps charge carriers from these layers with minimal recombination.

Multi-junction devices use two, or three, individual amorphous silicon cells stacked as shown in figure 22.23. To capture a broader portion of the sun's spectrum, the cells are made of materials with different band gaps. Amorphous silicon alloys with carbon, germanium, nitrogen, and tin can be used to vary the band gap and material properties to improve multi-junction devices.

Since the depositions needed to make thin-film multi-junction devices do not use much energy, such devices are potentially inexpensive to fabricate. Making a multi-layered cell is similar to making one cell; just adding one thin-film after another.

ii. Copper Indium Diselenide multi-junction

Copper indium diselenide cells have a band gap of 1.0eV and are used as a single junction cell or with a higher band-gap material in a multi-junction device. Amorphous silicon, with a band gap of about 1.7eV, on top of CIS is an example of a multi-junction cell, as shown in figure 22.24.

Copper indium diselenide, a versatile material, is used as a bottom cell in conjunction with top cells of materials such as CdTe and GaAs, which have band gaps of approximately 1.43 and 1.44eV.

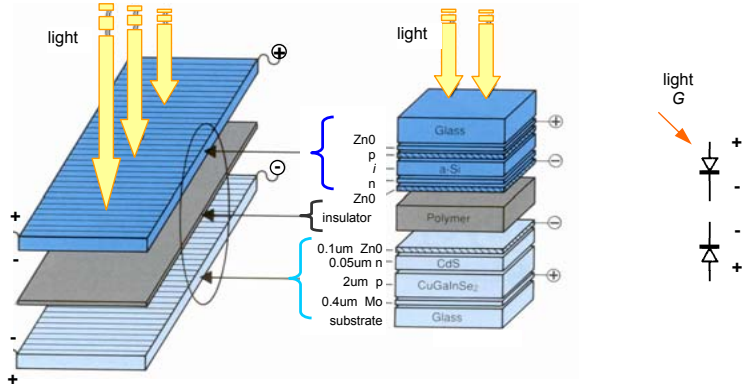


Figure 22.24. *A CID multi-junction PV cell.*

22.22 Equivalent circuit of a PV cell

To understand the electronic behaviour of a PV cell, it is useful to create a circuit model which is electrically equivalent, and is based on discrete electrical components whose behaviour is well known. An ideal single PV cell may be modelled by a current source in parallel with a diode; in practice no PV cell is ideal, so shunt resistance and series resistance components are added to the model.

22.22.1 Ideal PV cell model

During darkness the PV cell is not active and behaves as a diode, that is, a p-n junction diode, not producing current.

The simplest PV cell model consists of diode and current source parallel connected as shown in figure 22.25a. The current source current I_{ph} is directly proportional to the solar radiation G (with incident optical power P_{in}). The diode represents the pn junction of a PV cell. The equations of an ideal PV cell, which represents the ideal PV cell model, are:

$$I_D(V) = I_o\left\{e^{\frac{qV}{\gamma kT}} - 1\right\} = I_o\left\{e^{\frac{V}{\gamma V_{th}}} - 1\right\} \tag{22.22}$$

$$I_{ph} = \eta_g \times G \times A_c = \frac{q}{h \times v} P_{in} \tag{22.23}$$

where G is the ambient irradiance, W/m^2,
η_g is the generation efficiency, and
A_c is the cell effective or active area, m^2.

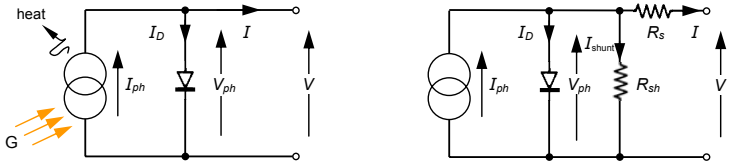


Figure 22.25. *PV cell electrical circuit model:*
(a) basic model and (b) extended model incorporating series and shunt resistances.

The net output current I is the difference between the photocurrent I_{ph} and the normal diode current I_D

$$I(V) = I_{ph} - I_D = I_{ph} - I_o\left\{e^{\frac{V}{\gamma V_{th}}} - 1\right\} \tag{22.24}$$

where V is the diode voltage (V)
I_{ph} is the photo-current from the current source (proportional to the incident light intensity), A
I_o is the diode dark saturation current or reverse saturation current, A (approximately 10^{-8}/m^2)
k is Boltzmann's constant, 1.38 x 10^{-23} J/K, h is Planck's constant, 6.626x10^{-34} J.s
q is the charge on an electron, 1.6 x 10^{-19} J/V, C or As, v is the photon frequency, Hz, and
T is the working temperature of the cell in degrees Kelvin, K.

$$V_{th} = \frac{kT}{q} \tag{22.25}$$

where V_{th} - thermal voltage, V_{th} = 25.7mV at 25°C,
γ - diode non-ideal factor = 1 to 2 (γ = 1 for ideal diode)

Both I_o and I_{ph} are strongly dependant on temperature.

$$I_o(T) = I_{SC}(T_1)\left(\frac{T}{T_1}\right)^{\frac{3}{\eta}}\left(\frac{e^{\frac{-qE_G}{\gamma k}\left(\frac{1}{T}-\frac{1}{T_1}\right)}}{e^{\frac{V_{OC,T_1}}{\gamma V_{th,T1}}} - 1}\right) \tag{22.26}$$

$$I_{ph}(T) = G\frac{I_{SC,nom,T_1}}{G_{nom}}\left\{1 + \frac{I_{SC,T_2} - I_{SC,T_1}}{T_2 - T_1}(T - T_1)\right\} \tag{22.27}$$

G_{nom} = 1kW/m^2, termed 1 sun,
E_G is the band gap voltage, eV,
V_{OC} is the open circuit output voltage, which has a temperature coefficient of 2.3mV/°C for silicon, and
I_{SC} is the short circuit output current, the largest output current, when V = I_D = 0 in equation (22.24), whence I_{SC} = I_{ph}.

The cell open circuit voltage V_{OC} (the cell voltage at night, G = 0) is when the output current is zero, I = 0, such that the model diode current I_D equals to photo generated current I_{ph}. That is, equating equation (22.24) to zero

$$I(V) = I_{ph} - I_o\left\{e^{\frac{V}{\gamma V_{th}}} - 1\right\} = 0 \tag{22.28}$$

gives

$$V_{oc} = \gamma V_{th}\,\ell n\left(\frac{I_{ph}}{I_o} + 1\right) \tag{22.29}$$

22.22.2 Practical PV cell model

The equivalent circuit of a practical PV cell is a constant current source in parallel with an ideal diode and a shunt resistor R_{sh}, plus that all in series with a series resistor R_s.

The equations which describes the relationship between the output current and output voltage characteristics, I-V, of the PV cell are

$$I(V) = I_{ph} - I_D - I_{shunt} = I_L - I_o\left\{e^{\frac{V-IR_s}{\gamma V_{th}}} - 1\right\} - \frac{V - IR_s}{R_{sh}} \tag{22.30}$$

22.22.3 Maximum-power point

A PV cell may operate over a wide range of output voltages V and currents I. The product of I and V is the power output of the cell P, and the solution of the equation that maximizes that product yields the voltage V_{mpp} and current I_{mpp} at the maximum power point P_m, as shown in equation (22.31). By increasing the load resistance on an irradiated cell, continuously from zero (a short circuit) to a very high value (an open circuit) the maximum-power point can be determine, the point that maximizes $V \times I$, that is, the load for which the cell can deliver maximum electrical power at that level of irradiation. The output power is zero in both the short circuit and open circuit extremes.

$$P_m = V_{mpp} \times I_{mpp} \tag{22.31}$$

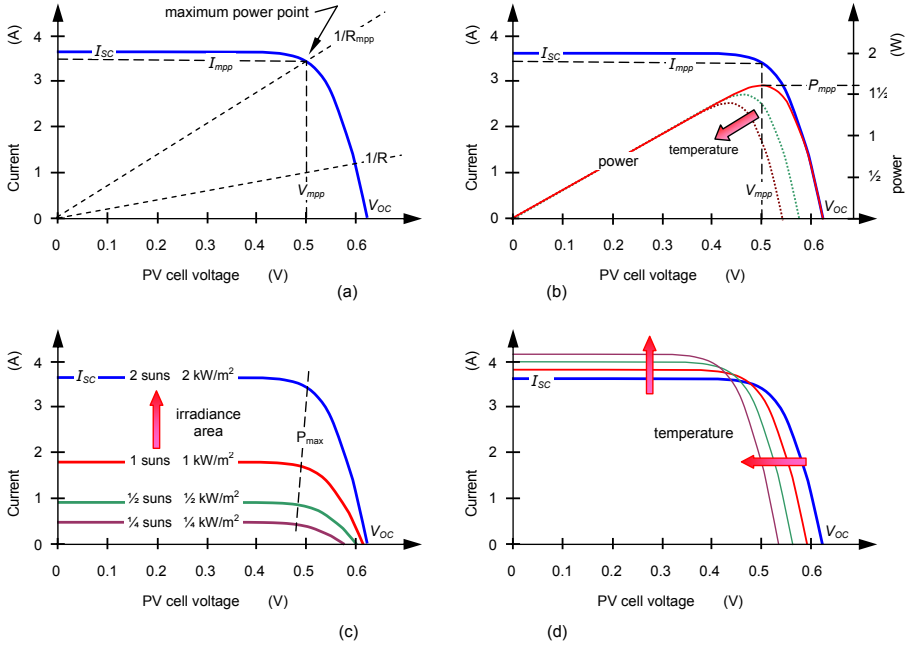


Figure 22.26. *A typical PV cell characteristics:*
(a) voltage-current I-V curve; (b) I-V and power characteristics; (c) effects of increased incident irradiance and cell area; and (d) effects of increased temperature on I-V characteristics.

A high quality, monocrystalline silicon PV cell, at 25°C cell temperature, may produce an open circuit output voltage of V_{OC} = 0.60V. The cell temperature in full sunlight, even with a 25°C ambient air temperature, will probably be close to 45°C, reducing the open circuit voltage to 0.55 Volts per cell, as indicated in figure 22.26d. The voltage drops modestly, with this type of cell, until the short circuit current I_{SC} is approached. Maximum power (with a 45°C cell temperature) is typically produced with 75% to 85% of the open circuit voltage (0.43V in this case) and 80% to 90% of the short circuit current. This output can be up to 70% of the V_{OC} x I_{SC} product. The power output of the cell is almost directly proportional to the intensity of the sunlight. For example, if the intensity of the sunlight is halved the power will also be halved, as shown in figure 22.26c. The short circuit current I_{SC} from a cell is nearly proportional to the illumination, while the open circuit voltage V_{OC} may drop only 10% with a 80% drop in illumination. Lower quality cells have a more rapid drop in voltage with increasing current and may produce only ½V_{OC} at ½I_{SC}. The usable power output could thus drop from 70% of the V_{OC} x I_{sc} product to 50% or even as little as 25%.
[Typical temperature coefficients: V_{OC}=-2 x10^{-3}/°C, V_{mpp}=-2.5 x10^{-3}/°C, I_{sc}=0.6x10^{-3}/°C, I_{mpp}=0.4x10^{-3}/°C]

22.23 Photovoltaic cell efficiency factors

Energy conversion efficiency

PV cell energy conversion efficiency η, is the percentage of power converted (from absorbed light to electrical energy) and collected, when a PV cell is connected to an electrical circuit. This efficiency is calculated using the ratio of the maximum power point, P_m, divided by the input light irradiance (G, in W/m²) under standard test conditions (STC) and the surface area of the PV cell, A_c in m².

$$\eta = \frac{P_m}{G \times A_c} \tag{22.32}$$

STC specifies a temperature of 25°C, an irradiance of 1000W/m², and a 0 m/s wind speed with an air mass 1.5 (AM = 1/cosθ, where angle θ is relative to the zenith, the vertical,) spectrum. These correspond to the irradiance and spectrum of sunlight incident on a clear day upon a sun-facing 37° - tilted surface with the sun at an angle of 41.81° above the horizon.
The losses of a PV cell are divided into:

- reflectance losses,
- thermodynamic efficiency,
- recombination losses, and
- resistive electrical loss.

The overall efficiency is related to the product of each of these individual losses, but these losses are not directly measurable. Other measurable parameters are used instead to specify efficiency:

- Thermodynamic Efficiency,
- Quantum Efficiency,
- V_{oc} ratio, and
- Fill Factor, *FF*.

Reflectance losses are a portion of the Quantum Efficiency. Recombination losses make up a portion of the Quantum Efficiency, V_{oc} ratio, and Fill Factor. Resistive losses are predominantly categorized under Fill Factor, but also make up minor portions of the Quantum Efficiency and V_{oc} ratio.

Thermodynamic efficiency limit

PV cell efficiency is limited because it operates as a quantum energy conversion device. Photons with energy below the band gap of the absorber material can not generate a electron-hole pair, and as such the photon energy is not converted to useful output. When a photon of greater energy than the band gap is absorbed, the excess energy above the band gap is converted to heat as the excess kinetic energy is released as the electron slows to equilibrium velocity. This loss is termed the *Thermodynamic Efficiency Limit*.

Quantum efficiency

An elevated electron-hole pair may travel to the surface of the PV cell and contribute to the current produced by the cell; such a carrier is said to be collected. Alternatively, the electron may give up its energy and once again become bound to an atom within the PV cell without reaching the surface; this is called recombination, and carriers that recombine do not contribute to the production of electrical current.
Quantum efficiency refers to the percentage of photons that are converted to electric current (that is, collected carriers) when the cell is operated under short circuit conditions. External quantum efficiency is the fraction of incident photons that are converted into electrical current, while internal quantum efficiency is the fraction of absorbed photons that are converted into electrical current. Mathematically, internal quantum efficiency is related to external quantum efficiency by the reflectance of the PV cell; given a perfect anti-reflection coating, they are the same.

V_{oc} ratio

Due to electron recombination, the open circuit voltage V_{oc} of the cell will be below the band gap voltage of the cell semiconductor. Since the energy of the photons must be at or above the band gap to generate a carrier pair, a cell voltage below the band gap voltage represents a loss. This loss is represented by

$$V_{OC}\ ratio = \frac{V_{OC}}{E_G} \tag{22.33}$$

Fill factor

Another term in the overall efficiency behaviour of a PV cell is the fill factor, *FF*, which is the ratio of the maximum power divided by the product of the open circuit voltage V_{oc} and the short circuit current I_{sc}:

$$FF = \frac{P_m}{V_{OC} I_{SC}} = \frac{I_{mpp} V_{mpp}}{V_{OC} I_{SC}} = \frac{\eta A_c G}{V_{OC} I_{SC}} \tag{22.34}$$

The cell fill-factor can be expressed as a function of the open circuit voltage by:

$$FF \approx \frac{\frac{V_{OC}}{\gamma V_{th}} - \ell n\left(\frac{V_{OC}}{\gamma V_{th}} + 0.72\right)}{\frac{V_{OC}}{\gamma V_{th}} + 1} \tag{22.35}$$

Concentrators

A concentrator is a PV cell designed to operate under illumination greater than 1 sun. The incident sunlight is focused or guided by optical elements such that a high intensity light beam shines on a small area PV cell. Concentrators have several potential advantages, including a higher efficiency potential than a one-sun PV cell and the possibility of lower cost. The short-circuit current from a PV cell depends linearly on light intensity, such that a device operating under 10 suns would have 10 times the short-circuit current as the same device under one sun operation. However, this effect does not provide an efficiency increase, since the incident power also increases linearly with concentration. Instead, the efficiency benefits arise from the logarithmic dependence of the open-circuit voltage on short circuit. Therefore, under concentration, V_{OC} increases logarithmically with light intensity, specifically:

$$V_{OC}^{X} = \gamma V_{TH}\, \ell n\left(\frac{X \times I_{SC}}{I_o}\right) = \gamma V_{TH}\left[\ell n\left(\frac{I_{SC}}{I_o}\right) + \ell n X\right] = V_{OC} + \gamma V_{TH}\, \ell n X \tag{22.36}$$

where X is the concentration of sunlight.

From equation (22.36), a doubling of the light intensity (X=2) causes an 18 mV rise in V_{OC}.
The cost of a concentrating PV system may be lower than a corresponding flat-plate PV system since only a small area of PV cells is needed.
The efficiency benefits of concentration may be reduced by increased losses in the series resistance as the short-circuit current increases and also by the increased operational temperature of the PV cell. As losses due to short-circuit current depend on the current squared, power loss due to series resistance increases as the square of the concentration.

Example 22.4: Solar cell characteristics

A 1cm² silicon solar cell has a saturation current I_o of 10^{-12}A and is illuminated with sunlight yielding a short-circuit photocurrent I_{SC} of 25mA. Calculate the solar cell efficiency and fill factor, assuming a power density of 1kW/m².

Solution

The maximum power is generated when:

$$\frac{dP}{dV} = \frac{d}{dV}\left(V\left[I_{ph} - I_o\left(e^{\frac{V}{V_{th}}} - 1\right)\right]\right) = 0 = I_{ph} - I_o\left(e^{\frac{V_{mpp}}{V_{th}}} - 1\right) - V_{mpp} I_o \frac{1}{V_{th}} e^{\frac{V_{mpp}}{V_{th}}}$$

where the voltage, V_{mpp}, is the voltage corresponding to the maximum power point. This voltage is obtained by solving the following transcendental equation, with V_{th} = 25.7mV and I_{ph} =25mA (since the short circuit current equals the photocurrent, I_{SC} = I_{ph} at V_{SC}) at 25°C:

$$V_{mpp} = V_{th} \ln\frac{1 + \frac{I_{ph}}{I_o}}{1 + \frac{V_{mpp}}{V_{th}}} = 0.0257\text{V} \times \ln\frac{1 + \frac{25 \times 10^{-3}\text{A}}{10^{-12}\text{A}}}{1 + \frac{V_{mpp}}{0.0257\text{V}}}$$

Iteration gives V_{mpp} =0.536V. The corresponding maximum power point current I_{mpp} is

$$I_{mpp}\left(V_{mpp}=0.54\text{V}\right)=I_{ph}-I_o\left(e^{\frac{V}{V_{th}}}-1\right)$$

$$=25\text{mA}-10^{-12}\left\{e^{\frac{0.536\text{V}}{0.0257\text{V}}}-1\right\}=25\text{mA}-1.14\times10^{-3}=23.86\text{mA}$$

The efficiency, assuming the irradiance power G of the sun is 1kW/m^2, equals:

$$\eta=\frac{V_{mpp}I_{mpp}}{P_{in}}=\frac{0.536\text{V}\times0.0239\text{A}}{1\times10^{-4}\text{m}^2\times1\text{kW/m}^2}=12.8\%$$

The open circuit voltage is calculated from

Error! Objects cannot be created from editing field codes.

which rearranged gives

$$V_{OC}=V_{th}\ln\left(\frac{I_{ph}}{I_o}+1\right)$$

$$=0.0257\text{V}\times\ln\left(\frac{25\text{mA}}{10^{-12}}+1\right)=0.615\text{V}$$

The fill factor is:

$$\textit{fill factor}=FF=\frac{V_{mpp}I_{mpp}}{V_{OC}I_{SC}}=\frac{0.536\text{V}\times0.0239\text{A}}{0.615\text{V}\times0.025\text{A}}=83.3\%$$

♣

22.24 Module (or array) series and parallel PV cell connection

Identical PV cells are series and parallel connection to form high-voltage, high-current, modules. For high current applications, modules are connected in parallel. For high voltage applications, modules are connected in series.

If all the PV cells in a module have identical electrical characteristics, and all experience the same insolation and temperature, then all the cells will operate at the same current and voltage. In this case, the I-V curve of the PV module has the same shape as that of the individual cells, except that the voltage and current are increased.

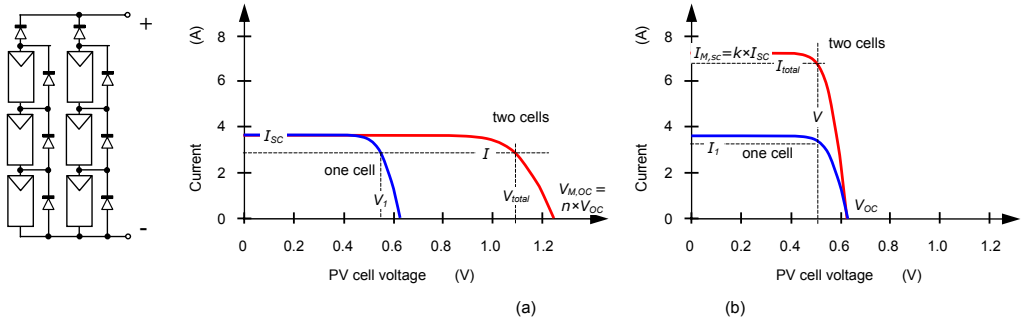


Figure 22.27. *I-V characteristics of identical PV cells connected in (a) series and (b) parallel.*

Series connection of n identical cells:

- When series connected, voltage adds $V_{total}=V_1+V_2+\ldots+V_n$
- The current remains constant $I_{total}=I_1=I_2$

With a series connection of voltage sources, the voltage of each source adds incrementally.

For n series connected PV cells, each with open circuit voltage V_{OC}, the module open circuit voltage is

$$V_{M,OC}=n\times V_{OC} \tag{22.37}$$

Parallel connection of k identical cells:

- When parallel connected, the currents add $I_{total}=I_1+I_2+..+I_k$
- The voltage remains constant $V_{total}=V_1=V_2.....$

With the parallel connection of voltage sources, the currents add. The voltage corresponds to that of a single source.

For k parallel connected PV cells, each with short circuit current I_{SC}, the module short-circuit current is

$$I_{M,sc}=k\times I_{SC} \tag{22.38}$$

If each cell has a series resistance R_s, the Thevenin equivalent resistance of the module is

$$R_{M,s}=\frac{n}{k}R_s \tag{22.39}$$

The same formula applies to the calculation of the equivalent module shunt resistance:

$$R_{M,sh}=\frac{n}{k}R_{sh} \tag{22.40}$$

The module maximum power P_M deliverable to a resistive load is

$$P_M=n\times k\times P_m \tag{22.41}$$

The module terminal voltage V_M and current I_M characteristic equation can be expressed in terms of each cell or the characteristics of the module, namely

$$I_M\left(V_M\right)=I_{M,SC}\left(1-\text{e}^{\frac{V_M-V_{M,OC}+R_{M,s}\times I_M}{n\gamma V_{th}}}\right) \tag{22.42}$$

or, in terms of the individual cell characteristics:

$$I_M\left(V_M\right)=k\times I_{SC}\left(1-\text{e}^{\frac{V_M-n\times V_{OC}+\frac{n}{k}R_s\times I_M}{n\gamma V_{th}}}\right) \tag{22.43}$$

In terms of the simple model in figure 22.25a

$$I_M\left(V_M\right)=k\times I_{ph}-k\times I_o\left(\text{e}^{\frac{V_M}{n\gamma V_{th}}}-1\right) \tag{22.44}$$

22.13:Typical and maximum module and cell conversion efficiencies at Standard Test Conditions

Type	Typical module efficiency	Maximum recorded module efficiency	Maximum recorded laboratory efficiency
	%	%	%
Single crystalline silicon	12-15	22.7	24.7
Multi-crystalline silicon	11-14	15.3	19.8
Amorphous silicon	5-7	-	12.7
Cadmium telluride, CdTe	-	10.5	16.0
$CuInGaSe_2$, CIGS	-	12.1	18.2

Example 22.5: PV cell and module characteristics

A 100cm^2 PV cell has I-V characteristics as shown in figure 22.25 parts a and b, at 25°C, achieved using a concentrator which gives 2kW/m^2 on each cell.

The PV cell is characterised by $V_{OC}=0.625\text{V}$ $V_{mpp}=0.50\text{V}$

$I_{SC}=3.7\text{A}$ $I_{mpp}=3.5\text{A}$

With diode properties $\gamma=1.5$ and $V_{th}=25.7\text{mV}$ at 25°C.

Use this information (and figures 22.26 and 22.27, if necessary) to find

i. The maximum cell power and optimal load resistance for delivering this power;
ii The fill factor FF;
iii The V_{OC} efficiency ratio, if the band gap of the silicon cell is 1.11eV at 25°C;
iv. The open-circuit voltage and short-circuit current values if a module uses 36 of such cells series connection, with three cells parallel connected at each voltage level; and
v. The module maximum power output and optimal load resistance for the given conditions.

Solution

i. The maximum cell power P_m is given by equation (22.31), that is

$$P_m = V_{mpp} \times I_{mpp}$$
$$= 0.5\text{V} \times 3.5\text{A} = 1.75\text{W}$$

ii. The efficiency fill factor *FF* is

$$FF = \frac{P_m}{V_{OC} I_{SC}}$$
$$= \frac{1.75\text{W}}{0.625\text{V} \times 3.7\text{A}} = 75.7\%$$

Alternatively, using equation (22.35)

$$FF \approx \frac{\frac{V_{OC}}{\gamma V_{th}} - \ell\text{n}\left(\frac{V_{OC}}{\gamma V_{th}} + 0.72\right)}{\frac{V_{OC}}{\gamma V_{th}} + 1} = \frac{16.21 - \ell\text{n}(16.21 + 0.72)}{16.21 + 1} = 77\%$$

iii. The V_{OC} efficiency ratio is

$$V_{OC}\ ratio = \frac{V_{OC}}{E_G} = \frac{0.625\text{V}}{1.11\text{eV}} = 56.3\%$$

iv. With 36 series connected cells the open circuit output voltage is

$$V_{M,OC} = n \times V_{OC}$$
$$= 36 \times 0.625\text{V} = 22.5\text{V}$$

The short circuit output current for three parallel connected cells at each voltage level is

$$I_{M,sc} = k \times I_{SC}$$
$$= 3 \times 3.7\text{A} = 11.1\text{A}$$

v. The module maximum power P_M is

$$P_M = n \times k \times P_m$$
$$= 36 \times 3 \times 1.75\text{W} = 189\text{W}$$

where the output voltage is 36×0.50V=18V and the output current is 3×3.5A=10.5A (18V×10.5A=189W)

♣

22.25 Battery storage

PV cells can only produce electricity during the day, and then only on clear days. To be independent of the grid, energy storage is needed but batteries add cost and maintenance to the PV system. This problem is avoided with connection to the utility grid, buying power when needed and selling when producing more than needed, as shown in figure 22.28. This way, the utility acts as a practically infinite bidirectional storage system. Suitable inverter equipment is needed to ensure that the power sold to the utility is synchronous with the grid, specifically the same sinusoidal waveform and frequency. The utility has to be assured that if there is a power outage, the PV system does not try to feed electricity into lines. This disconnection is called *islanding*.

Batteries need regular maintenance, and replacement after a few years, although the PV modules last for 20 years or more. Batteries in PV systems can also be dangerous because of the voltage level at which the energy is stored and the acidic electrolytes they contain, so the protective environment must be well-ventilated and non-metallic.

A lead-acid car type battery is a shallow-cycle battery and PV storage requires deep-cycle batteries which can discharge a significant amount of their stored energy while still maintaining long life. PV batteries generally have to discharge a small current for a long period (such as all night), while being charged during the day. The nickel-cadmium battery is the most commonly used deep-cycle battery, which is more expensive than the lead-acid battery, but last longer and can be discharged more completely without harm. Lead-acid batteries cannot be deep-cycled or discharged 100 percent without seriously shortening battery life. Generally, if a PV system uses lead-acid batteries, the system is designed to discharge the batteries by no more than 40 percent or 50 percent.

For long battery life, a charge controller is needed to prevent overcharging and complete discharge.

The other problem besides energy storage is that the electricity generated by PV modules, and extracted from the storage batteries if used, is in dc form while household appliances and the utility grid need ac voltage. Most distributed generation inverters automatically control the system. Some PV

modules, called AC modules, have a built-in system inverter, eliminating the need for a large, central inverter, and simplifying wiring and safety issues.

A PV system requires little maintenance (especially if no batteries are used), and can provide electricity cleanly and quietly for 20 years or more.

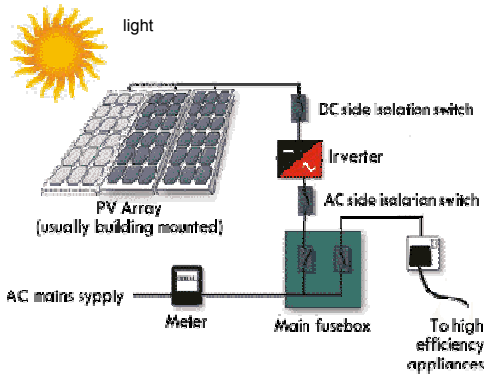


Figure 22.28. *Schematic of a typical residential PV system with battery storage.*

Weather and temperature

Weather naturally affects the performance of PV modules but not as may be expected. The amount of sunlight is most important in determining the output a PV electric system will produce at a given location, but temperature is also important. PV modules actually generate more power at lower temperatures with other factors being equal. This is because PV cells generate electricity from light, not heat. Like most electronic devices, PV cells operate more efficiently at cooler temperatures. In temperate climates, PV panels will generate less energy in the winter than in the summer but this is due to the shorter days, lower sun angles and greater cloud cover, not the cooler temperatures.

PV module output is proportional to the sun's intensity, so cloud cover will reduce the system output. Typically, the output of any industrial PV module is reduced to 5% to 20% of its full sun output when operated under cloudy conditions.

During a typical sunny day, a PV array one metre square exposed to the sun at noon will receive approximately 1kW of power. Multi-crystalline cells convert roughly 15% of this irradiation energy into electricity, hence 1m² of cells generates 150W in full sunshine. The warmer the cells become, the higher the losses. At 60ºC the capacity decreases by about 18% to 123W, and at 70ºC output drops by 24% to 114W. As shown in figure 22.26d, this reduced power output is because:

- The voltage decreases, at an increasing rate, with increased temperature.
- The current increases minimally with increasing temperature.

This means that the maximum power output of a PV cell decreases with increased temperature, especially under higher irradiation levels.

22.26 The organic photovoltaic cell

Following absorption of photons by the polymer, bound electron-hole pairs (excitons) are generated, subsequently undergoing dissociation, as shown in figure 22.29. Due to inherent limitations in organic materials (exciton lifetime and low charge mobility), only a small fraction of photon-generated electron-hole pairs effectively contribute to the photocurrent. The organic cell facilitates volume distribution of the photogeneration sites, thereby enhancing exciton dissociation. This is achieved by increasing the junction surface area, with an interpenetrating network of the donor-acceptor type, effecting transport of holes to the indium-tin oxide anode, and of electrons to the aluminium metallic cathode. While quantum separation efficiency, for photo-induced charges in systems associating a semiconducting polymer, polythiophene, with a fullerene derivative, is thus close to unity, the objective is to restrict recombination and trapping processes, limiting charge transport and collection at the electrodes, to improve overall device efficiency, this currently still being low, less than 5%. The rise of the pathway is also dependent on the cell aging mechanisms and the thin-film technologies, to protect the device against atmospheric oxygen and water vapour ingress.

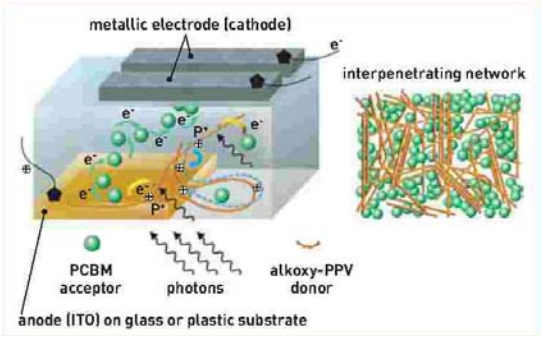


Figure 22.29. *Organic PV cell.*

Organic PV cells
Operation of organic PV cells is mechanistically more complex. First, a molecule of an organic compound absorbs a photon and forms an excited state (exciton). Further, the exciton diffuses to a junction border between n- and p-types of semiconductor where it dissociates to form free charge carriers. Organic p- and n-transporters are also known as donors and acceptors respectively.
If there is no junction border nearby, the exciton may recombine (decay) via photoluminescence, or thermally, back into the ground state of the molecule. This is the main reason, why mono- and bilayered (the scheme in figure 22.29) organic PV cells were poorly performing devices.

The bulk heterojunction shown in figure 22.30 is a compact blend of a p-type conductor (donor), and n-type conductor (acceptor) in the photoactive layer of a device, where the concentration of each component often gradually increases when approaching the corresponding electrode. This magnifies the p-n-junction's total surface and strongly facilitates exciton dissociation. The implication of this concept in practice allows increased power conversion efficiencies of up to 5% for all-organic PV cells.

The realization of the bulk-heterojunction concept, the organic cells and materials need further development in order to foster commercial application. Advantages of organic PV cells are: lightweight, environmentally friendly, no requirements for rare metals and minerals, no high temperatures and purity demand at the production stage, potentially inexpensive, virtually unlimited possibilities for further material improvements.

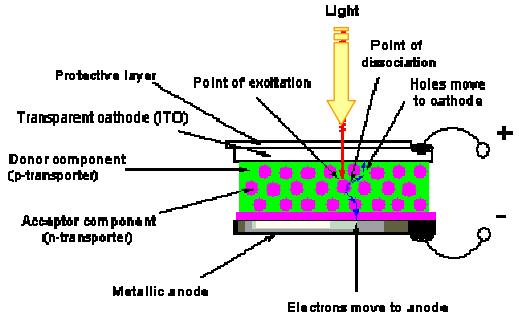


Figure 22.30. *Bulk heterojunction organic solar cell.*

22.27 Summary of PV cell technology

PV advantages

- Infinite source of input energy. The 89 petawatts of sunlight reaching the earth's surface is plentiful - almost 6,000 times more than the 15 terawatts of average power consumed by humans.
- The silicon cells manufactured from one ton of sand produce as much electricity as burning 500,000 tons of coal.
- Solar electric generation has the highest power density (global mean of 170W/m²) among renewable energies.
- Low maintenance and long service lifetime, durable with minimal degradation over 10 years.
- No emissions of $C0_2$, $S0_2$, $N0_2$, radiation.
- Peak output matches daily peak demand.
- Can be installed near point of energy need.
- Grid-connected PV electricity can be used locally thus reducing transmission/ distribution losses or used decentralized.
- Grid system integration possible, or independent of the grid thereby avoiding lack any of transmission and distribution infrastructure.
- Quick installation.
- Modular and expandable, for increased power.
- Low operating costs, no fuel, transport and storage costs.
- No moving parts to wear, operate silently with minimal system movement.
- Environmental benign, no waste, no noise, free energy, no fuel transport costs.
- High public acceptance.
- Excellent safety record, with no combustible fuels.
- Suitable for harsh environments.
- Can be portable.
- Suitable for remote locations, including high-temperature, high-altitude environments, with improved performance, where hydrocarbon based systems are derated.

PV disadvantages

- Sunlight is a low intensity energy source, hence limited power density, 2 to 5 kW·h/m^2.
- Produces dc, which must be converted to ac with at least 4% losses.
- Hi tech skills to create the technology, but not necessary for installation, operation and maintenance.
- Some PV materials are toxic, for example the cadmium in cadmium telluride PV cells, therefore requiring careful end of life treatment.
- Energy not available at night or during overcast weather.
- Expensive initial cell costs, plus expensive component replacement costs.
- High installation costs.
- Poor ancillary equipment reliability of grid tie inverters and storage methods, e.g., batteries, although PVs are highly reliable if maintained (cleaned and protected).
- Lack of commercial storage facilities.
- Accelerated sunlight ageing of associated synthetic materials.
- Local weather patterns and sun conditions directly affect the potential of photovoltaic systems. Some locations will not be able to use solar power.

There are two disadvantages often used by environmentalist concerning high-tech PVs:

- Production Pollution:- Fossil fuels are extensively utilized to extract, produce and transport PV panels. These processes also entail corresponding sources of pollution. The life cycle analysis of a PV system is a net positive for the environment because it can offset fossil fuel energy production over its approximately 25-year lifetime.
- High energy cost:- Require much energy to produce. The three types of photovoltaic (PV) materials, which make up the majority of the active PV market: single crystal, polycrystalline, and amorphous silicon PV cells pay for themselves in terms of energy in a few years (1 to 5 years). They thus generate enough energy over their lifetimes to reproduce themselves many times (6 to 31 reproductions) depending on what type of material, balance of system, and the geographic location of the system.

Table 22.14: Summary of PV cell technology (AR: - antireflective coating)

Material	Thickness	Efficiency	Colour	Features
Mono-crystalline Si	300µm	15-18%	Dark blue, black with AR coating, grey without AR coating	Lengthy production procedure - wafer sawing necessary. Highly researched PV cell material. Highest power/area ratio.
Polycrystalline Si	300µm	13-15%	Blue, with AR coating, silver-grey without AR coating	Wafer sawing necessary. Currently the most important production procedure.
Polycrystalline transparent Si	300µm	10%	Blue, with AR coating, silver-grey without AR coating	Lower efficiency than monocrystalline PV cells.
Edge defined Film fed Growth EFG	280µm	14%	Blue, with AR coating	Limited use of this production procedure. Very fast crystal growth, no wafer sawing necessary .
Polycrystalline ribbon Si	300µm	12%	Blue, with AR coating, silver-grey without AR coating	Limited use of this production procedure, no wafer sawing necessary. Decrease in production costs expected.
Apex (polycrystalline Si)	30 to 100µm + ceramic substrate	9.5%	Blue, with AR coating, silver-grey without AR coating	Single source wafer production procedure, no wafer sawing, production in form of band possible. Significant decrease in production costs.
Monocrystalline dendritic web Si	130µm incl contacts	13%	Blue, with AR coating	Limited use of this production procedure, no wafer sawing, production in form of band possible.
Amorphous silicon	0.1µm + 1 to 3mm substrate	5-8%	Red-blue, Black	Lower efficiency, shorter life span. No sawing necessary, production in the form of band.
Cadmium Telluride (CdTe)	8µm + 3mm glass substrate	6-9% (module)	Dark green, Black	Poisonous raw materials. High production costs.
Copper-Indium-Diselenide (CIS)	3µm + 3mm glass substrate	8-10% (module)	Black	Limited indium supply in nature. High production costs.
Hybrid silicon (HIT)	20µm	18%	Dark blue, black	Limited use of this production procedure, higher efficiency, better temperature coefficient, and lower thickness.

Reading list

http://www.mpoweruk.com/

http://science.howstuffworks.com/solar-cell3.htm

http://photovoltaics.sandia.gov/

http://americanhistory.si.edu/fuelcells/index.htm

http://www1.eere.energy.gov

http://www.fctec.com/fctec_basics.asp

http://en.wikipedia.org/wiki/Solar_cell

http://www.pvresources.com/en/technologies.php

http://www.doitpoms.ac.uk/tlplib/index.php

http://www.fuelcells.org/info/

Problems

22.1 Consider the decomposition of H_2O_2 (hydrogen peroxide) at 298 K and 1 atm pressure according to:

$$2\ H_2O_2(l) \leftrightarrow 2\ H_2O(l) + O_2(g)$$

Substance	ΔG_f°	ΔH_f°	S°
	kJ/mol	kJ/mol	J/ K.mol
$H_2O_2(l)$	-120.2	-187.6	109.5
$H_2O(l)$	-237.0	-285.8	69.4
$O_2(g)$	-	-	205

Find the
a) standard enthalpy of reaction.
b) standard entropy of reaction.
[-196.4 kJ; 125 J/K]

22.2. For the reaction in the previous question, find the
a) standard (Gibbs) free energy of reaction
b) the value of the (thermodynamic) equilibrium constant at 298 K, 1 atm
[-233.6 kJ; 8.85 x 10^{40}]

22.3. Carbon monoxide in the atmosphere slowly converts to carbon dioxide at normal atmospheric temperatures according to:

$$CO(g) + ½\ O_2(g) \leftrightarrow CO_2(g)$$

The standard enthalpy of reaction is -284 kJ and the standard entropy of reaction is -87 J/K. Estimate the temperature at which the equilibrium begins to favour the decomposition of CO_2. Assume that the enthalpy and the entropy of reaction are not affected by temperature.
[3260 K or 3300 K]

22.4. Indicate if TRUE or FALSE:

The entropy of a gas increases with increasing temperature.
The energy of a perfect crystal is zero at 0 K.
Spontaneous processes always increase the entropy of the reacting system.
All spontaneous processes release heat to the surroundings.
An endothermic reaction is more likely to be spontaneous at high temperatures than at low temperatures.
The entropy of sugar decreases as it precipitates from an aqueous solution.

[T F F F T T]

22.5. For each type of fuel cell below, what are the charge carrying species, i.e., electrolyte type?

(a) Polymeric exchange membrane
(b) Immobilised alkaline solution
(c) Direct methanol
(d) Phosphoric acid
(e) Molten carbonate
(f) Ceramic solid oxide

[H+, OH–, H+, H+, CO_3^{2-}, O^{2-}]

22.6. Which cells in question 22.5 are able to internally reform?

[High temperature, molten carbonate and solid oxide cells.]

22.7. The efficiency limit (voltage) of H_2 cells reduces as operating temperature is increased. Why are high efficiency systems often run at high temperatures?

[The heat produced can be more useful than the efficiency gained at lower temperatures. Running a system at a higher temperature allows the hot exhaust gasses to be utilised in a turbine-hybrid system, raising the overall efficiency of the system can be in excess of 80%.]

22.8. Give two reasons why fuel cells are not currently used in cars.

[(i) High temperature fuel cells are not easily portable and low temperature ones require hydrogen as the fuel in order to produce any useful power. Hydrogen is not yet a readily available fuel. It is also difficult to store.
(ii) Fuel cells also suffer from low reaction rates and therefore low power output.]

22.9. Why are nickel anodes used in SOFCs? How is it treated in order to aid bonding to YSZ?

[It is cheap and although it functions effectively, it flakes off unless mixed with zirconia.]

22.10. Give three ways to provide a mobile fuel cell with pure hydrogen.

[(i) Pure H_2 may be stored as used either under pressure, cryogenically, in a metal hydride, in a high surface area carbon nano-tube arrangement or even in glass nano-spheres.
(ii) Many fuels may be carried and reformed before use e.g. alcohols, alkali metal hydrides, ammonia, sodium borohydride, methane and even hydrazine.
(iii) Alternately the cell could be run directly on methanol.]

Blank

CHAPTER 23

Energy Sources and Storage -

Secondary Sources

The progressive proliferation of embedded and distributed generation with renewable energy sources has spurred research into alternative energy sources and storage methods. This chapter is concerned with secondary energy sources, viz., so called super or double-layer capacitors, electro-chemical batteries, and thermoelectric modules. Their energy and power density capabilities (and those of primary electrical sources) have been put into context by considering conventional energy sources in Chapter 22.1, specifically the hydrocarbons and hydrogen gas.

In electrical terms, primary and secondary energy sources are defined as follows.

Primary source is not a reversible energy source. During energy discharge, the original states are permanently changed as electrical energy is released until the original energy reactant sources, or any one of them, are depleted. A primary cell can be used only once.

Secondary source is reversible and the original states can be reconstituted by the application of an electrical potential that injects conversion energy into the source. A secondary cell can source and sink energy many times.

23.1 Batteries

An electrochemical battery cell is an 'electron pump' that stores energy in chemical form in its active materials and can convert this stored chemical energy to electrical energy on demand, typically by means of an electrochemical oxidation-reduction reaction or a physical reaction.
This energy conversion is achieved by a chemical reaction in the battery that releases electrons. If a load is placed across the battery terminals, the chemical reaction produces electrical power. The process can be reversible. If electrical energy is directed into the battery (charging the battery), the chemical reaction reverses and restores the battery to a fully charged condition.

Cells are classified as either primary cells or secondary cells, depending on whether or not the electrochemical cell is rechargeable.

Primary electrochemical cell: The electrochemical reaction is not reversible. During discharge, the chemical compounds are permanently changed and electrical energy and heat are released until the original compounds, or any one of them, are depleted. A primary cell is assemble in the charged state, discharged during utilisation, used only once until discharged, then discarded, possibly involving recycling of components.

Secondary electrochemical cell: The electrochemical reaction is reversible and the original chemical compounds can be reconstituted by the application of an electrical potential between the electrodes injecting energy into the cell. The cell discharge and charge current directions are opposite. A secondary cell can be discharged and recharged many times. It is usually assembled in a discharged state and has to be charged before under-going discharged.

Batteries come in a wide range of types, sizes and shapes, from wafer-thin to button-size devices to large industrial battery systems. All can be categorised as either primary storage or secondary storage batteries. Batteries function best at room temperature (expect lithium polymer batteries) and elevated temperature operation dramatically increases the internal resistance and shortens battery lifetime.

Primary Batteries

Primary batteries have the following properties:

- Designed as a single use battery, used until exhausted then discarded or recycled after depletion
- High impedance which translates to long-life energy storage, and for low current loads
- Available in carbon-zinc, alkaline, silver oxide, lithium, zinc air and some lithium metal batteries (like lithium-thionyl-chloride)
- Lithium-thionyl-chloride batteries, for example, come in cylindrical form factors of AAA to D. Larger C and D size lithium-thionyl-chloride batteries are a chemical hazard and are restricted when transported by air
- Operating temperature range is typically -40ºC to +85ºC

Although primary cells may consist of the same active materials as secondary battery types, they are constructed so that only one continuous or intermittent discharge can be obtained.

Secondary Batteries

Secondary batteries are recharged by a flow of direct current through them in a direction opposite to the current flow on discharge. By recharging after discharge, a higher state of oxidation is created at the positive electrode plate (cathode) and a lower state at the negative electrode plate (anode), returning the plates to approximately their original charged condition.

Secondary batteries have the following properties:

- Designed to be recharged
- Can be recharged up to 1,000 times depending on the usage and battery type
- Very deep discharges result in a shorter cycle life, whereas shorter discharges result in long cycle life for most secondary batteries
- Charge time varies from one to twelve hours, depending upon battery condition, temperature, depth of discharge, and other factors
- Include NiCd, lead-acid, NiMH, some lithium metal and Li-ion batteries
- Lead-Acid and NiCd batteries are toxic and are subject to disposal regulations

Some of the general limitations suffered by secondary batteries are limited life, limited power capability, low energy-efficiency, and disposal concerns.

The secondary cell is the subject of this section. The primary cell is not specifically considered further.

23.2 The secondary electro-chemical cell

Each electro-chemical energy cell consists of at least three, sometimes four, components

- The ***cathode*** or positive electrode (the oxidising electrode) accepts electrons from the external circuit and is chemically reduced during the electrochemical (discharge) reaction. It is usually a metallic oxide or a sulphide. The cathodic process is the reduction of the oxide ion - anion - to leave the metal. (To gain or to accept electrons is termed reduction). The cathode is the half-cell with the higher electrode potential.

- The ***anode*** or negative electrode (the reducing or fuel electrode) gives up electrons to the external circuit and is oxidised during the electrochemical (discharge) reaction. It is generally a metal or an alloy. The anodic process is the oxidation of the metal to form metal ions - cations. (To lose or to supply electrons is termed oxidation). The anode is the half-cell with the lower (least positive) electrode potential.
- The ***electrolyte*** (the ionic conductor) provides the medium for transfer of charge in the form of ions inside the cell in either direction between the cathode and anode. The electrolyte is typically a solvent containing dissolved chemicals providing ionic conductivity. It should be a non-conductor of electrons to avoid internal self-discharge of the cell.
- The ***separator*** electrically isolates the positive and negative electrodes but allows ions to travel back and forth in the electrolyte, between the electrodes.

The operation of lead-acid and nickel-cadmium batteries are based on oxidation and reduction chemistry, called REDOX. The nickel-metal-hydride and lithium-ion batteries are not based on REDOX reactions, but rather involve an ion transfer mechanism called *intercalation*, which is the insertion and extraction of ions into and out of the crystalline lattice of an electrode, without chemically altering its crystal structure.

23.2.1 REDOX Galvanic Action

Galvanic action is when chemical reactions produce electricity, while electrolysis is the reverse process where electricity is used to produce chemicals.
Different metals have different affinities for electrons. When two dissimilar metals (or metal compounds) are put in contact or connected through a conducting medium there is a tendency for electrons to pass from the metal with the smaller affinity for electrons, which becomes positively charged, to the metal with the greater electron affinity which becomes negatively charged. A potential difference between the metals will therefore build up until it just balances the tendency of the electron transfer between the metals. At this point, the 'equilibrium potential' is that which balances the difference between the propensity of the two metals to gain or lose electrons.
A battery can be considered as an electron pump. The internal chemical reaction within the cell between the electrolyte and the negative metal electrode produces a build-up of free electrons, each with a negative charge, at the cell's negative terminal - the anode. The chemical reaction between the electrolyte and the positive electrode (the cathode) inside the cell produces an excess of positive ions (atoms that are missing electrons, thus with a net positive charge) at the cell cathode. The electrical (pump) pressure or potential difference between the positive and negative electrodes is called voltage or electromotive force, EMF.

i. The Discharge Process

When the cell is fully charged there is a surplus of electrons on the anode giving it a negative charge and an electron deficiency on the cathode giving it a positive charge, resulting in a potential difference between the cell electrodes.
When the external electrical circuit is completed the surplus electrons flow in the external circuit from the negatively charged anode which loses its charge to the positively charged cathode which accepts it, neutralising its positive charge. This action reduces the potential difference across the cell to zero. The external circuit electron flow is balanced by the internal flow of positive ions through the electrolyte from the anode to the cathode.
Since the electrons are negatively charged, the electrical current they conventionally represent flows in the opposite direction, from the cathode (positive terminal) to the anode (negative terminal).
The battery's chemical reaction continues to generate electrical current until at least one of the active materials involved in the reaction is depleted or the external load connection is removed.

ii. The Charging Process

The charging process strips electrons from the cathode leaving it with a net positive charge and forces them onto the anode giving it a negative charge. The electrical energy pumped into the cell transforms the active chemicals back to their original state.

iii. Half-cell reaction

Chemical reactions occur on both the anode electrode and the cathode electrode. Half cell reaction refers to the chemical processes occurring at each electrode (half-reactions), namely the negative anode and the positive cathode.
The cell is modelled as two half-cells. Gibbs free energy equation ($\Delta G = -nFE^o$, see Chapter 22.14, equation (22.14) and Chapter 23.8, equation (23.9)) is used to calculate electrode potentials and characterise the chemical reactions within the cell. The cell voltage or EMF to force the external current from a cell is the difference in the standard electrode potentials of the two half-cell reactions under standard conditions. The actual voltage of a chemical cell is dependant on the temperature, pressure,

and concentrations of the reactants and products. The Nernst equation ($E = E^o - RT/nF \times \ell nN$, see Chapter 22.14, equation (22.18) and Chapter 23.8, equation (23.10)) can be used to calculate the EMF for non-standard cell conditions. The cell EMF decreases as the concentration of the active chemicals diminishes as they are consumed, until one of the reactants is exhausted. The half-cell zero reference potential is defined to be zero for the hydrogen electrode. All the equations are written as reductions. The two half-reaction potentials add to give the overall cell potential. The alkaline half-cell reactions, during discharge, for example, are shown in Table 23.1.

Table 23.1: Half-cell electro-chemical equations for the alkaline cell

Location		**Reaction** (discharge) 185A-hr/kg	**Potential**
Anode - oxidation	Negative terminal	$Zn_{(s)} + 2OH^-_{(aq)} \rightarrow ZnO_{(s)} + H_2O + 2e^-$	$E^-_{½cell} = +0.76V$
Cathode - reduction	Positive terminal	$2MnO_{2(s)} + H_2O + 2e^- \rightarrow Mn_2O_{3(s)} + 2OH^-_{(aq)}$	$E^+_{½cell} = +2.31V$
Net REDOX reaction		$Zn_{(s)} + 2MnO_{2(s)} \rightarrow ZnO_{(s)} + Mn_2O_{3(s)}$	$E^o_{cell} = 1.55V$

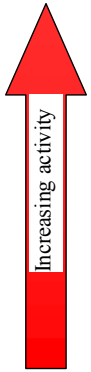
The cell potential under standard conditions (25°C, 1 mole, 1 Atmosphere) is

$$E^o_{cell} = E^+_{½cell} - E^-_{½cell} = E^o_{cathode} - E^o_{anode} \qquad (23.1)$$

The half-cell equations are reversed during cell charging, although the standard alkaline cell may not be rechargeable, that is primary and secondary variations exist.

Table 23.2: Strength of Oxidizing and Reducing Agents (25°C, 1 mol, 1 Atmosphere)

Anode Materials		Cathode Materials	
(Negative Terminal)		(Positive Terminal)	
BEST - Most Negative		BEST Most Positive	
Cathode (Reduction) Half-cell Reaction	Standard Potential $E^o_{½}$ (V)	Cathode (Reduction) Half-cell Reaction	Standard Potential $E^o_{½}$ (V)
$Li^+_{(aq)} + e^- \rightarrow Li_{(s)}$	-3.045	$F_{2(g)} + 2e^- \rightarrow 2F^-_{(aq)}$	+2.87
$K^+_{(aq)} + e^- \rightarrow K_{(s)}$	-2.925	$MnO_4^{2-} + 4H^+ + 2e^- \rightarrow MnO_{2(s)} + 2H_2O$	+2.257
$Ca^{2+}_{(aq)} + 2e^- \rightarrow Ca_{(s)}$	-2.766	$Co_3O_{4(s)} + 8H^+ + 2e^- \rightarrow Co^{2+} + 4H_2O$	+2.11
$Mg^{2+}_{(aq)} + 2e^- \rightarrow Mg_{(s)}$	-2.353	$PbO_{2(s)} + SO_4^{2-}{}_{(aq)} + 4H^+_{(aq)} + 2e^- \rightarrow PbSO_{4(s)} + 2H_2O_{(l)}$	+1.685
$Al^{3+}_{(aq)} + 3e^- \rightarrow Al_{(s)}$	-1.66	$FeO_4^{2-} + 6H^+ + 3e^- \rightarrow Fe(OH)_2^+ + 2H_2O$	+1.56
$Zn^{2+}_{(aq)} + 2e^- \rightarrow Zn_{(s)}$	-0.763	$MnO_2 + 4H^+ + 2e^- \rightarrow Mn^{2+} + 2H_2O$	+1.23
$Fe^{2+}_{(aq)} + 2e^- \rightarrow Fe_{(s)}$	-0.440	$O_{2(g)} + 4H^+ + 4e^- \rightarrow 2H_2O$	+1.229
$Ni^{2+}_{(aq)} + 2e^- \rightarrow Ni_{(s)}$	-0.250	$Ag_2O_{(s)} + 2H^+ + 2e^- \rightarrow 2Ag_{(s)} + H_2O$	+1.173
$Pb^{2+}_{(aq)} + 2e^- \rightarrow Pb_{(s)}$	-0.126	$HgO_{(s)} + 2H^+ + 2e^- \rightarrow Hg_{(l)} + H_2O$	+0.93
$2H^+ + 2e^- \rightarrow H_2(g)$	0.00	$CuO_{(s)} + 2H^+ + e^- \rightarrow Cu^+ + H_2O$	+0.62
$Cu^{2+}_{(aq)} + 2e^- \rightarrow Cu_{(s)}$	+0.337	$NiOOH + H_2O + e^- \rightarrow Ni(OH)_2 + OH^-$	+0.49
$Ag^+_{(aq)} + e^- \rightarrow Ag_{(s)}$	+0.799	$NiO_{2(s)} + 2H_2O + 2e^- \rightarrow Ni(OH)_{2(s)} + 2OH^-$	+0.49



iv. Choice of Active Cell Chemicals

The voltage and current generated by a galvanic cell is directly related to the types of chemical materials used for the electrodes and electrolyte.
The propensity of an individual metal or metal compound to gain or lose electrons in relation to another material is known as its electrode potential. Thus the strengths of oxidizing and reducing agents are indicated by their standard electrode potentials. Compounds with a positive electrode potential are used for anodes and those with a negative electrode potential for cathodes. The larger the difference between the electrode potentials of the anode and cathode, the greater the EMF of the cell and generally the greater the amount of energy that can be produced by the cell.

The Electrochemical Series is a table of metallic elements or ions arranged according to their electrode potentials, as shown in Table 23.2. The order shows the tendency of one metal to reduce the ions of any other metal below it in the series, at 25°C, standard pressure and 1mol of reactant.
Table 23.2 shows some common chemicals used for battery electrodes arranged in order of their relative electrode potentials. Lithium has the most negative standard potential, -3.045V, indicating that it is the strongest reducing agent. The strongest practical oxidizing agent is fluorine with the largest positive standard electrode potential of +2.87V. Although halogen acids have high electrode potentials, they are usually too aggressive with the electrodes to be used in normal cells. Metals are commonly used in cells as anode materials, but the cathode is commonly an oxide. That is, in Table 23.2, metals appear in the left columns while oxides appear in the right columns.

v. Gassing
Cells using aqueous (containing water) electrolytes are limited in voltage to less than 2.4V because the oxygen and hydrogen in water dissociate producing H_2 and O_2 in the presence of voltages above this voltage. Lithium batteries use non-aqueous electrolytes hence do not have dissociation problems and are available in voltages between 2.7V and 3.7V. However, the use of non-aqueous electrolytes results in such cells having relatively high internal impedance.

23.2.2 Intercalation Action
From intercalation process studies, it is known that small ions (such as ions of lithium, sodium, and the other alkali metals) can fit in the interstitial spaces of a carbon graphite lattice crystal. These metallic molecules force the graphitic planes apart to fit numerous layers of metallic molecules between the carbon sheets. This is an efficient way to store the metal-ion in a battery.
The anode of a conventional alkali metal-ion cell is made from graphite (carbon), the cathode is a metal oxide, and the electrolyte is a metal salt in an organic solvent.
Both the anode and cathode are materials into which metal-ions insert and extract, termed intercalation. The process of metal-ions moving into the anode or cathode lattice is referred to as insertion, and the reverse process, in which metal-ions move out of the anode or cathode is referred to as extraction.
Using lithium as an example, the underlying chemical reaction that allows Li-ion cells to provide electricity (equations to the right) are:

anode	$C_6Li + xLi^+ + xe^- \rightleftarrows C_6Li_{1+x}$	negative electrode	0.1V wrt Li
cathode	$LiCoO_{2(s)} \rightleftarrows Li_{1-x}CoO_{2(s)} + xLi^+ + xe^-$	positive electrode	3.8V wrt Li
	$C_6LI + LiCoO_2 \rightleftarrows Li_{1-x}CoO_2 + C_6Li_{1+x}$		3.7V net

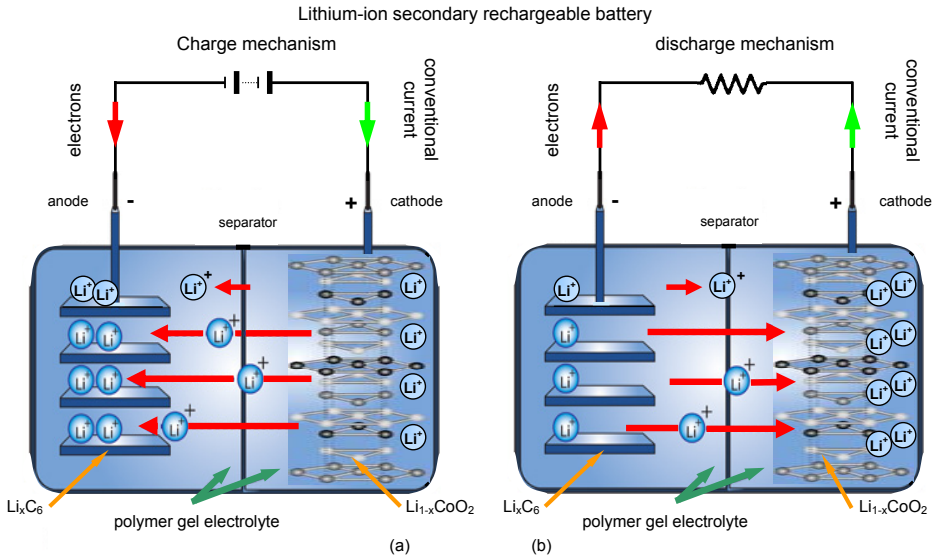


Figure 23.1. *Cells charge mechanism and discharge mechanism.*

The lithium-ions are not oxidized; rather, they are transported to and from the cathode or anode, with the transition metal, cobalt, in $LiCoO_2$ being oxidized from Co^{3+} to Co^{4+} during charging, and reduced from Co^{4+} to Co^{3+} during discharge. At no stage is any alkali lithium metal present or involved.
The intercalation host electrodes have two key properties

- Open crystal structures which allow the insertion and extraction of alkali metal-ions
- The ability to simultaneously accept compensating electrons - conductive

When **discharging** a cell, the metal-ions are extracted from the anode and inserted into the cathode, via the electrolyte, as shown in figure 23.1b. That is, alkali metal ions move through the electrolyte from the negative electrode to the positive electrode and attach to the carbon. The electrolyte is non-conducting to electrons. At the same time compensating electrons, which form the external circuit current, transfer from the positive to the negative electrodes, and are accepted by the internally arriving metal-ions, thereby balancing the equation.
When **charging** the cell, the reverse process occurs: metal-ions are extracted from the cathode and inserted into the anode, as shown in figure 23.1a. The metal ions move back to the anode from the carbon cathode, while external current electrons flow from the negative to positive electrodes. The anode graphite is a two-dimensional crystal structure, which under charging, is forced to laterally shift and simultaneously strain to 10% greater separation to accommodate the Li-ions. The lattice deformation is relieved when Li-ions are removed from the anode under cell discharge.

23.3 Characteristics of Secondary Batteries

A wide range of secondary batteries exists, each offering different attributes, limitations, properties, etc. The four secondary batteries to be considered are

- Lead-acid
- Nickel-cadmium
- Nickel-metal-hydride
- Lithium-ion

Electrochemical lead-acid and nickel-cadmium battery technologies are mature. The lead-acid battery is economical for high power applications where weight is of little concern. It is an inexpensive, robust technology, found extensively in automotive applications and UPS equipment. The nickel-cadmium battery has a higher energy density than the lead acid battery, and offers longer lifetimes, higher discharge rates and a wider operating temperature range than the lead-acid battery. The nickel-cadmium battery contains toxic metals, and is used in power tool applications.
The nickel-metal-hydride battery trades a higher energy density than nickel-cadmium for reduced cycle life. Importantly it uses non-toxic metals, and is used in satellites, mobile phones and laptops. Lithium-ion, a newer developing technology, offers higher again energy densities, but cell series connection poses sharing and balancing problems. Because of its high energy density properties, lithium-ion is increasingly being used in notebook computers, mobile phones, and power tools.
Key technology properties and features of different battery technologies are summarised in tables 23.3a and 23.3b.

Table 23.3a. Anode capabilities of different metals

Anode	Symbol	Atomic Mass	Standard Potential	Density	Melting Point	Electrochemical Equivalence
		g	V	g/cm^3	°C	Ah/g
Lithium	Li	6.94	-3.05	0.54	180	3.86
Sodium	Na	23.0	-2.7	0.97	97.8	1.16
Magnesium	Mg	24.3	-2.4	1.74	650	2.20
Aluminium	Aℓ	26.9	-1.7	2.7	659	2.98
Calcium	Ca	40.1	-2.87	1.54	851	1.34
Iron	Fe	55.8	-0.44	7.85	1528	0.96
Zinc	Zn	65.4	-0.76	7.1	419	0.82
Cadmium	Cd	112	-0.40	8.65	321	0.48
Lead	Pb	207	-0.13	11.3	327	0.26

In Table 23.3a, lithium is the lightest metal, and has the highest standard potential. Lithium metal is volatile (with water and nitrogen in air) and has a low reactivity with many cathode materials and non-aqueous electrolytes. Because of the energy density potential offered by lithium-ion technologies, different cathode material types have emerged.
Three key performance aspects should be defined, namely power density, energy density, and capacity.

i. Power density (specific power): *Volumetric Power density* is the ratio of the power available from a battery to its volume (W/litre). *Specific power* (or gravimetric power density) refers to the ratio of power to mass (W/kg). Comparison of power to cell mass is more common.
Power Density (W/kg) indicates how much power a battery can deliver on demand. Manganese and phosphate-based lithium-ion, as well as nickel-based chemistries, give the best performance. High power density cell uses are power tools, medical devices and transportation systems. The focus is on power bursts, such as drilling through heavy steel, rather than runtime.

ii. Energy density (specific energy): *Volumetric Energy density* refers to the ratio of a battery's available energy to its volume (Wh/litre). *Specific energy* (or gravimetric energy density) refers to the ratio of energy to mass (Wh/kg). The energy W is determined by the charge q that can be stored and the cell voltage E, that is, $W = q \times E$.
Energy Density (Wh/kg) is a measure of how much energy a battery can retain or store. The higher the energy density, the longer the possible runtime. Cell size and its chemistry determine energy storage density. Lithium-ion cells with cobalt cathodes offer the highest energy densities, 190Wh/kg. Typical applications are cell phones, laptops and digital cameras.
An Ah/kg rating can be determined from the cell standard voltage, E and energy density rating Wh/kg, while power and energy densities are related by time, $W = P \times t$, as shown in figure 23.2.
An analogy between energy and power densities can be made with a water bottle. The size of the bottle is the energy density, while the opening area denotes the power density. A large volume bottle can carry a lot of water (energy), while a large opening can pour it quickly (power). A large container (energy) with a wide area mouth (power) is the best combination.
Table 23.3 shows some typical relative energy and power per unit weight examples of common secondary cell chemistries. In general, higher energy densities are obtained by using more reactive chemicals. But reactive chemicals tend to be unstable and require safety precautions. The energy density is also dependent on the quality of the active materials used in the cell construction, with impurities limiting cell capacity.

iii. Capacity and Battery Rating, C-rate
In general terms, the electrical energy capacity of a cell or battery is the amount of charge available expressed in ampere-hours, Ahr. An ampere, the unit of measurement used for electrical current, is defined as a coulomb of charge passing through an electrical conductor in one second. Therefore, the unit ampere-hours, Ahr, equates directly to the quantity of electrical charge stored: 1Ahr is the same as 3600 coulombs of charge. The capacity of a cell or battery is related to the quantity of active materials in it, the amount of electrolyte, and the surface area of the plates. The capacity of a battery/cell is measured by discharging at a constant current until its terminal voltage is reached (typically about 1.75V for lead acid cells). This test is usually done at a constant temperature, under standard conditions of 25°C. The capacity is calculated by multiplying the discharge current value by the time required to reach the terminal voltage.
The term used to describe a battery's ability to deliver current is its rated capacity, specified in ampere-hours at a specific discharge rate. For example, a lead-acid battery rated for 200Ahr (for a 10-hour rate) will deliver 20A of current for 10 hours under standard temperature conditions, 25°C. Alternatively, a discharge rate may be specified by its charge-rate or C-rate, which is expressed as a multiple of the rated capacity of the cell or battery. For example, a battery may have a rating of 200Ahr at a C/10 discharge rate. The discharge rate is determined by the equation below:

$$\frac{C}{10}\,rate = \frac{200\text{Ahr}}{10\text{hr}} = 20\text{A}$$

Battery capacity varies with the discharge rate. The higher the discharge rate, the lower the cell capacity. Fast charging or discharging can generate heat inside the battery that lowers its chemical efficiency, hence reduces its effective capacity. Lower discharge rates result in higher capacity, but very slow charging or discharging can also give lower effective capacity since all batteries suffer from internal self-discharging effects. Batteries are normally specified at several discharge rates (in amperes) along with the associated discharge time (in hours). The capacity of the battery for each of the discharge rates can be calculated as discussed above.
The lead-acid battery does not perform well at a 1C discharge rate (discharged in one hour). The rated capacity for lead-acid batteries is usually specified at 10 or 20-hour rates (C/10, C/20). UPS batteries are rated at 8-hour capacities and telecommunications batteries are rated at 10-hour capacities.

Table 23.3b: Basic comparison of different battery characteristics and technologies

		Lead-acid	Nickel-cadmium	Re-chargeable Alkaline	Nickel-metal-hydride	Lithium-ion		
						$LiCoO_2$	$LiMn_2O_4$	$LiFePO_4$
		sealed	sealed			cobalt	manganese	phosphate
gravimetric energy density	Wh/kg	30-50	45-70	80	60-120 250kJ/kg	150-190	100-135	90-120
volumetric energy density	Wh/L	60-90	60-150	220	140-300 360MJ/m^3	220-350	270	170
power density	W/kg	180	150	50	250-1000	760	1800	1400
cell voltage (theoretical)	V	2 (2.1)	1.25 (1.35)	1.5	1.25	3.6	3.7	3.3
internal resistance	mΩ/cell	20	30		40	100	50	35
load current	Peak, C Optimum Depth of discharge	5 0.2 80%	20 1 100%	½ < 0.2	5 0.5 80%	< 3 < 1 80%	> 40 < 10 80%	> 35 < 8 80%
operating temperature range	°C	-20 to 60	-40 to 60	0 to 65	-20 to 60	-20 to 60		
cycle lifetime	80% capacity	250	1500	<25	400	300-500	400	>1000
fast charge	hr	8	1	2-3	3	2	> 1	> 1
Charge V limit	V	2.4	constant I		constant I	4.2	4.2	3.7
overcharge tolerance		high	moderate	moderate	low	low – no trickle charge tolerated		
self-discharge	%/month @ 20°C	2-5	15-20	0.3%	18-30	6-10		
maintenance	discharge to months	6	1V/cell 2	Not required	1V/cell 3	Not required		
cool storage	% charge top-up yr	100 ½	40 ½		40 ½	40 Not required		
safety		Thermally stable	Fuse Thermally stable		Fuse Thermally stable	Mandatory protection Stable to 150°C	Protection needed Stable to 150°C	Protection needed Stable to 150°C
electrolyte	liquid	H_2SO_4 4M	KOH 7M	KOH	KOH 6M	$LiPF_6$ + solvent		
toxicity		Toxic lead and acid	Highly toxic	Low toxicity	Low toxicity, recyclable	Low toxicity		
Cost	pu	low 0.6	low 1	low 0.5	medium 1.6	high 2		
commercialisation		1970	1950	1992	1990	1991	1996	2006
other features		Heavy Inexpensive Rugged	Long-life Durable Memory		Bulky High pressure	Low weight Low maintenance Needs temperature monitoring Needs over/under V and I protection		

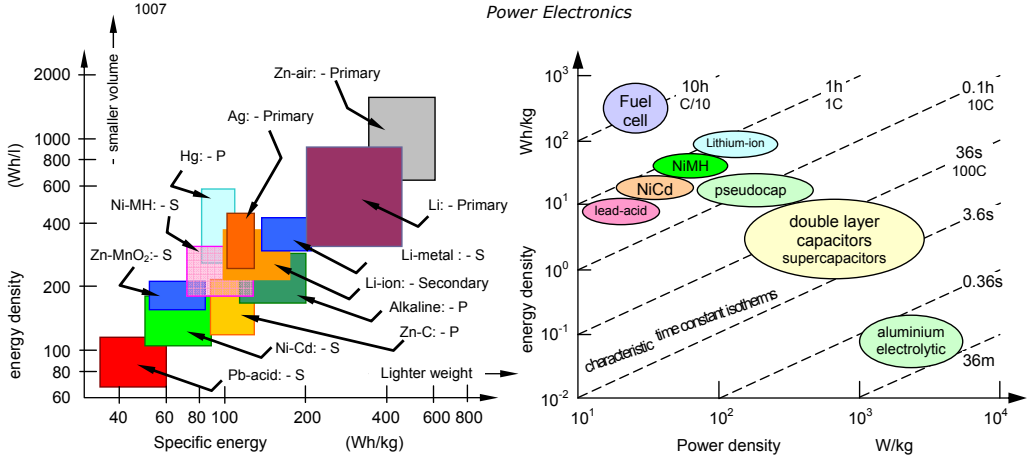


Figure 23.2. *Gravimetric energy and power densities of different cell technologies.*

23.4 The lead-acid battery

Anode: Sponge metallic lead + Antimony, tin or calcium
Cathode: Lead dioxide, PbO_2
Electrolyte: Dilute aqueous sulphuric acid, H_2SO_4 (4.5 Mol)
Applications: Motive power in cars, trucks, forklifts, construction equipment, recreational water craft, standby/backup systems
Typical ratings: Specific energy density: 35 to 50 Wh/kg
Volumetric energy density: 60 to 70 Wh/l

23.4.1 Basic lead-acid cell theory

The traditional lead-acid battery is made up of flat plates, a lead anode, and a lead oxide cathode (various other elements are used to change density, hardness, porosity, etc.) suspended in a 35% sulphuric acid and 65% water solution. This solution is called electrolyte and causes a chemical reaction that produces electrons. A fibreglass separator between the plates prevents them from touching and short-circuiting.

The lead-acid battery discharge half-cell chemistry reactions shown in Table 23.4 (the reactions are reversed during charging).

Table 23.4: Half-cell electro-chemical (double sulphate) equations for the lead-acid cell

Location		**Reaction** (discharge) 55A-hr/kg	**Potential**
Anode	Negative terminal	$Pb_{(s)} + HSO_4^-{}_{(aq)} \rightarrow PbSO_{4(s)} + H+_{(aq)} + 2e^-$	$E^-_{½cell}$ = - 0.356V
Cathode	Positive terminal	$PbO_{2(s)} + HSO_4^-{}_{(aq)} + 3H^+ + 2e^- \rightarrow PbSO_{4(s)} + 2H_2O$	$E^+_{½cell}$ = 1.685V
Net REDOX reaction		$PbO_{2(s)} + Pb_{(s)} + 2H_2SO_{4\,(aq)} \rightarrow 2PbSO_{4(s)} + 2H_2O$ positive active material; negative active material; positive & negative electrolyte; active material; electrolyte	E^o_{cell} = 2.041 V

23.4.2 Cell/battery construction

The four key battery internal components shown in figure 23.3 are:

- Plates (grid and paste)
- Separator
- Electrolyte
- Vent (valve regulated)

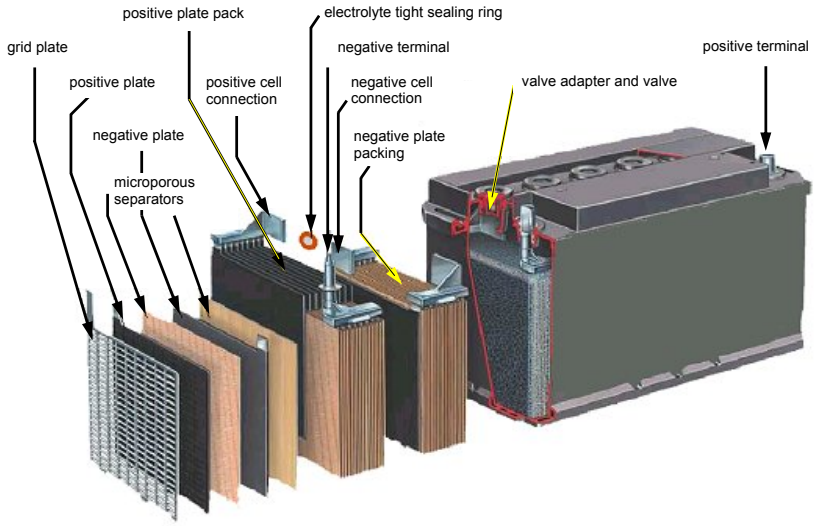


Figure 23.3. *Cells used in the Lead Acid Battery. [Source Eurobat]*

i. Plates

The plates and their grid structure that holds the active material have two functions:

- To provide mechanical support for the active material, which itself does not have the rigidity to be self-supporting.
- To act as a conductor to transmit current from all parts of the active material to the plate terminal. The ideal grid design should maintain a uniform current distribution throughout the active material. An uneven current distribution results in plate buckling during charge and heavy discharge cycles.

The ideal grid plate structure has the following characteristics:

- High conductivity
- High strength
- Corrosion resistance
- Compatibility with the active materials
- Manufacturability

Plate grid materials and characteristics

The plate grid structure in both pasted and tubular plate batteries is made from a lead alloy. A pure lead grid structure is not strong enough to stand vertically when supporting the active material. Other metals in small quantities are alloyed with lead for added strength and improved electrical properties.
The alloy used for plate grid frames varies:

- lead-antimony (<2% for hardening and resistance to corrosion);
- lead-calcium (for hardening) - along with Sn and Aℓ additives, but free of antimony, to improve cyclability; and
- pure lead. Other alloys are also used for cyclability, such as tin, cadmium, and rare earths - selenium.

Lead-antimony cells are recommended for applications requiring long life under regular deep cycling regimes, discharging to depths greater than 20% of rated capacity, plus infrequent operation with a discharge depth as high as 50-60% of the 6-hour to 8-hour rated capacity.
Grids of lead calcium alloy or pure lead can also be cycled but repetitive cycles are restricted to a depth less than 20% of capacity plus infrequent operation with a discharge depth up to 50 to 60% of the 6-hour to 8-hour rated capacity.
Pure lead alloy plate types are used when low charged standing loss is required and occasional deep cycles are expected.
Lead-calcium and pure lead positive grids are used in standby applications when long periods of float are expected between power outages and discharge cycles, and where the average discharge depth is

less than 20%. Pure lead designs offer better cycle life than lead-calcium, but poorer plate mechanical strength. During charging the cells are maintained at a constant charge voltage, and the charge current required to maintain lead-calcium and pure lead positive grid cells at voltage is much lower than for the lead-antimony grid. A reduction in float charge current reduces water loss and the maintenance cost of adding water to battery cells.

- Lead-antimony batteries can be deep cycled more times than lead-calcium batteries.
- Flooded lead-antimony batteries require more frequent maintenance as they near end-of-life since they use an increasing amount of water and require periodic equalization charges.
- Lead-calcium batteries have lower self-discharge rates as shown in figure 23.4 and therefore, draws less current while on float charge than lead-antimony batteries.
- Lead-calcium positive plates grow in length and width because of grid oxidation at the grain boundaries. This oxidation is usually caused by long-term overcharging, which is common to UPS and other batteries on constant-float changing. Grids may grow in size sufficiently to cause buckling or rupture of their containers.

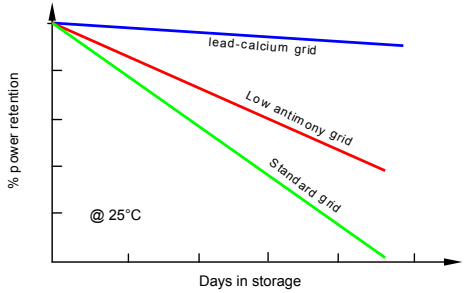


Figure 23.4. *Self-discharge rates of different plate materials.*

Lead-selenium (<2%) is the least common grid material. Lead-calcium (0.1% Ca plus ½% Sn) is a common **anode** grid material, but a lower cost alternative, lead-antimony (2 to 4% Sb), offers:

- lower internal heat and water losses,
- better mechanical strength, manufacturability,
- longer service life,
- easier to charge,
- lower cost,
- higher maintenance (water consumption increases dramatically at near end of life),
- higher discharge depths, but
- a high self-discharge rate of up to 5% per week, since antimony is very reactive.

Lead acid batteries with electrodes modified by the addition of calcium providing the following advantages:

- More resistant to corrosion, overcharging, gassing, water usage, and self-discharge, all of which shorten battery life,
- Larger electrolyte reserve area above the plates,
- Higher Cold Cranking Amp ratings,
- Little or no maintenance, and
- Tin increases grid strength, improves manufacturability, and inhibits positive plate passivation.

Plate structures

Lead acid batteries have positive and negative electrode 'plates' and are classified by the type of positive plate (the negative plate is always the pasted lead flat plate type):

- Planté positive plate
- Manchex positive plate
- Tubular positive plate
- Pasted flat (positive, always negative) plate
- Rod (positive, always negative) plate

Negative plates in all lead-acid cells are formed from flat grid structures, pasted with pure lead as the electrical active material. A grid structure containing active material is termed a plate. The positive grid plate is pasted with porous lead dioxide. Both of these pastes are in a porous or spongy form to optimize surface area, thereby maximizing capacity.

(a) Planté plate:- The simplest lead-acid battery electrode is the *Planté plate*, which is a flat casted plate composed of pure lead that is attached to an antimony alloy lead connecting strap, as shown in figure 23.5. Since the capacity of a lead-acid battery is proportional to the electrode surface area exposed to the electrolyte, various methods are employed to increase electrode surface area per unit volume or weight. Planté plates are grooved, etched, striated, or perforated to increase the effective surface area. A typical Planté plate, which is only used as the positive electrode, is shown in figure 23.5. It offers good energy density and 100% capacity for its lifetime, which is determined by when less than 80% capacity is available.

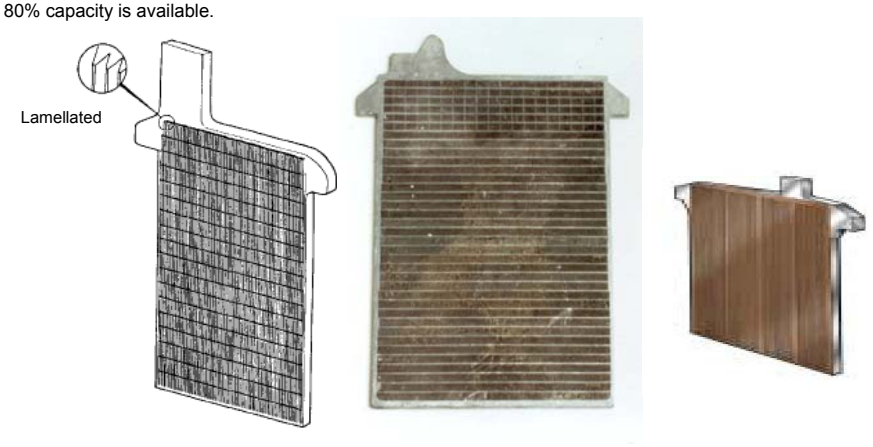


Figure 23.5. *Planté grid structure and active paste filling.*

(b) Manchex (Manchester) plate:- The *Manchex* type positive grid (or modified Planté plate) is casted with a low antimony, lead alloy. A button or rosette is a pure lead ribbon, which is serrated and rolled into a spiral form. These in turn are pressed or wedged into the holes of the grid. The surface of the buttons is oxidized to PbO_2 forming the positive active material. The grid is heavy and therefore gives long life, particularly in standby type service with moderate cycling such as railroad signal and utility application.

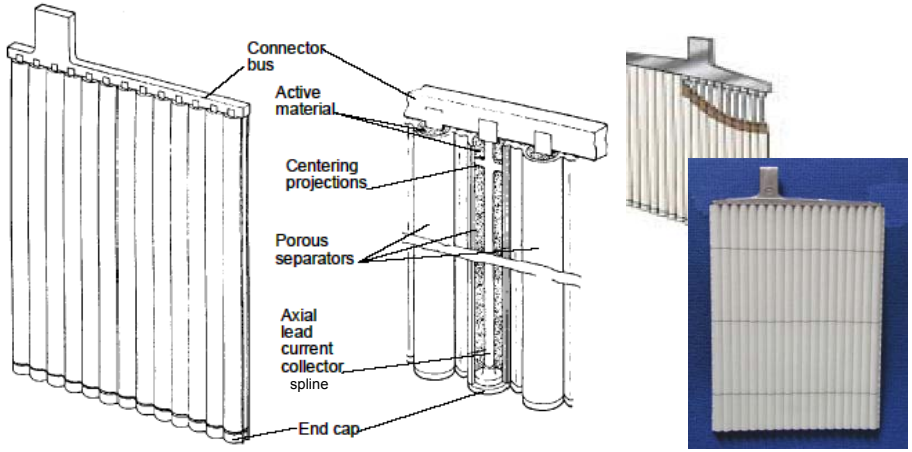


Figure 23.6. *Tubular plate structure and paste filling.*

(c) Tubular Positive plate:- The tubular plate construction is shown in figure 23.6. The grid, which is the current conducting member, is a series of low antimony lead spines. Woven or porous plastic or glass material is used for the tubing, which is centred on each individual vertical spine. Then the active material is added and the ends are sealed. The tubular plate is only used as a positive electrode.
This type of battery is generally used in lift truck and material handling applications where deep cycling routines require an average discharge depth of 70 to 80 percent of the 6-hour rated capacity and recharge within an 8-hour period. Tubular positive plate batteries are also used for on-the-road diesel starting (motive power applications) and utility switchgear. In 20% depth, shallow cycling regimes, over 4,000 cycles can be obtained when multi-cycles per day are performed. It has excellent high charging characteristics, good standby life, and form a versatile cell type. Float life lasting up to 15 years is achieved. Tubular batteries are normally produced in one plate thickness. Variations in capacity are obtained by increasing the number of tubes per plate and/or by varying the tube (or plate) height.

(d) Flat Pasted plate:- The pasted plate is the most common electrode for vented cells and can be used for both positive and negative plate structures. A typical pasted plate construction is shown in figure 23.7. The lattice grid is punched or cast with pure lead, lead-calcium, pure lead-tin or lead-antimony depending on the size of the plate and the application. Grid life is dependant on the grid alloy and plate thickness.
Active material, lead oxide PbO, is applied as a wet paste and the plate is then cured, dried and formed. The paste, or active material, is mounted into a frame or grid structure that mechanically supports it and serves as the electrical conductor carrying the current during both charge and discharge cycles. After the lead oxide paste has dried, the plates are immersed in a dilute sulphuric acid solution and current is passed through them, with opposite polarities for the positive and negative plates. The lead oxide (PbO) is converted to lead dioxide (PbO_2) in the positive plates and to lead (Pb) in the negative plates.
Glass mats and a perforated plastic retainer are wrapped around the positive plate to minimize the loss of positive active material and to obtain good cycle life. This wrap performs the same function as the retainer tube of tubular positive plates.
During the last stages of charge, oxygen gas is formed at the surface of positive plates. The agitation of gas bubbles streaming from the surface of exposed grids and active material tends to erode the active material, which is shed through the glass retainer and settles into the sediment space at the bottom of each cell. In light cycle or in float service, such positive active material shedding is not the major failure mechanism. In these applications, the glass mat retainer is lighter, thinner, and the perforated outer wrap is omitted.

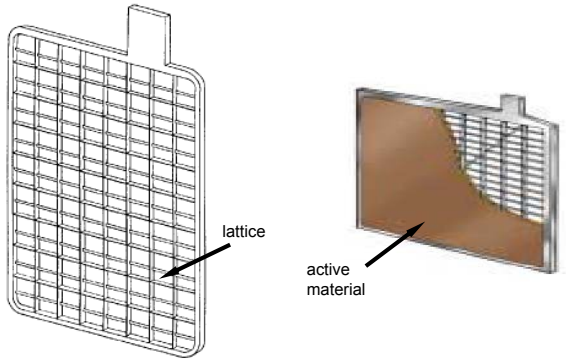


Figure 23.7. *Pasted plate structure with paste filling (left: negative grid, right: positive grid).*

Such designs depend upon a ribbed microporous separator adjacent the negative plate to achieve longest life. Pasted plates have good energy density and are made with thin (1mm) or thick (7mm) grids depending on the application, which are mainly automotive and standby. When the application demands a high ampere rate for a short time, many thin plates are used. Thicker plates with fewer plates per cell are used for applications with a low ampere drain for long periods of time. In general, for a giving service condition, thin plates result in a shorter life than thick plates.

(e) Rod plate:- Rod plate electrode shown in figure 23.8 can form either positive or negative electrodes, Robust rod plate electrodes use a lead antimony or lead-calcium cast grid, resulting in a moderate energy density. They are mainly used for standby applications and yield a 12 to 20 year life.

Figure 23.8. *Rod plate structure with paste filling.*

ii. Expanders
The chemical reaction of the lead-acid battery depends, in part, on how well the sulphuric acid in the electrolyte can react with the lead in the positive and negative plates. An expander is added to the active material in each plate to facilitate the diffusion of electrolyte into the plate. It is generally a mixture of an organic compound, typically lignin or a lignin derivative such as barium sulphate and carbon black.
As the ratio of expander-to-active material is increased, the electrolyte's ability to make contact with all of the active material improves. Longer-life cells typically have thick plates to better withstand the unavoidable corrosion process; expanders help the electrolyte reach the interior of thicker plates.
With too much expander added to the plate, the active material tends to dissolve into the electrolyte, destroying the cell. Too little expander has the effect of reducing the cell's capacity because the electrolyte is restricted in its ability to enter the interior active material in the plate. The addition of expander to the active material is another variable that has to be controlled during the manufacturing.

iii. Separators and Retainers
Separators are installed between the plates to prevent them from touching and shorting. A separator is constructed of a highly porous, nonconductive, inert material (non-woven glass fibre cloth, sintered PVC, or PVC/polyethylene) that allows conduction of electrolyte ions between the plates. The ideal separator has the following characteristics:

- As thin as possible so that it does not add volume to the cell
- As stiff as possible to hold its position throughout the cell's life
- Highly porous to allow electrolyte diffusion throughout the cell and to reduce the internal resistance
- Small pore size to prevent electrical shorts between the plates
- Puncture resistant so that dendrites and plate imperfections do not damage it
- Nonconductive, so as not to contribute to plate self-discharge
- Electrochemically compatible with the acid and lead environment (heat and oxidation resistance)
- Electrical low resistance to minimize effect on cell voltage
- Manufacturability and inexpensive

A retainer is a porous mat of inert material, such as fibreglass, that is either pressed between the plates or wrapped around the positive plate. Repeated charge and discharge cycles tend to cause shedding of active material from the plate surface. The retainer helps keep the active material in position on the plates. The separator might also have ridges to ensure compression of the retainer against the plates.

iv. Electrolyte
Dilute sulphuric acid H_2SO_4 is used as the medium for conducting ions in the electrochemical reaction in the battery. Sulphuric acid is highly reactive and ionizes almost completely in water. The ions are in constant motion, attracted or repelled by one another. This constant random motion tends to cause the ions to diffuse throughout the electrolyte. The diffusion process is not immediate and can take a relatively long time to reach equilibrium throughout the electrolyte.
Sulphuric acid electrolyte has a specific gravity of 1.320 at 25°C (water has a specific gravity of 1.0).

Absorbent Glass Mat *(AGM) Separator*
Micro porous non-woven fibreglass (silica) cloth acts as both a separator prevent shorting between positive and negative plates and an electrolyte absorber (by capillary action). The cloth has a high heat and oxidation resistance and high porosity offers excellent electrolyte absorption and retaining ability, as well as excellent ion conductivity, compressibility, puncture resistance and electrical resistance, with high heat and oxidation resistance. The blend of glass micro fibres having an optimum ratio of fine and extra fine fibre sizes which is chemically stable, inert, to lead oxide and in the dilute sulphuric acid electrolyte, which is immobilized. This blend features superior wicking characteristics and promotes maximum retention of the electrolyte. The flexible and resilient AGM layer is squeezed to an optimum

compression level during assembly to provide sufficient long-term contact with the surface of the plate, thus prolonging the battery cyclic life. This compression also promotes retention of the active material if the battery is exposed to shock or vibration conditions. Since the plates are completely wrapped by the separator and the electrolyte is completely absorbed in the separator and plates, shedding of active material, which can cause shorting and reduced battery safety with flooded batteries, is avoided.
The AGM also allows oxygen diffusion from the positive plate for the oxygen recombination cycle at the negative plate.
The AGM electrolyte contains high purity sulphuric acid and demineralised, deionised water to increase battery performance. Since the designs are 'acid-starved', between 90 and 95 percent - but not fully, saturated, to protect the plates from deep discharge, the acid concentration can drop to nearly zero during an extremely deep discharge. Substances that will not dissolve in acid may become soluble when the concentration drops this low. Upon recharge, these dissolved substances crystallize out of the electrolyte, potentially destroying the battery. The electrolyte prevents these possibilities.
The AGM separator can be enveloped with a thin layer of microporous polyethylene which is wrapped around the glass-matted plate and then sealed along the sides to eliminate the possibility of shorts at the edges of the plate (a common failure mode). The microporous polyethylene is more durable and puncture resistant than the AGM material alone and significantly reduces the occurrence of plate-to-plate shorts.

***Gel** electrolytes*
Gelled electrolyte contains sulphuric acid, fumed silica, pure demineralised, deionised water, and a phosphoric acid additive. The phosphoric acid is key to batteries delivering a long cycle life.
A gel glass mat, with double insulating separators is also used.

- The fibreglass mats embed into the surface of the plates, reinforcing the plates and lock the active material onto the plate for longer life and extended performance.
- The clean separators have no oil contamination or other impurities, therefore, resistance is low and battery performance is high.
- Excellent porosity allows maximum charge flow, which means better power-per-kg.
- Superior resistance to oxidation reduces separator failure, which extends life.

The AGM and gelled electrolyte battery designs have common components such as containers, plates and pressure relief valves, but they have different separator and electrolyte immobilization systems. Consequently, the gelled design can have a greater electrolyte reserve than the AGM cell and is usually better suited to long duration applications. A gelled electrolyte cell is typically heavier and larger than an absorbed electrolyte cell for a given capacity.

v. Pressure Relief Safety Valve
Each cell in the VRLA battery has a pressure relief safety valve, to prevent case bursting, where the sealing material is neoprene rubber. The normally closed valve is designed to release excess gasses that may abnormally build up over time having not recombined inside of the battery, resulting from overcharging, high operating temperatures or misusage. Once the pressure is vented, the valve instantly automatically re-seals. This valve function is repeatable, whenever venting is required. The oxygen and some hydrogen gasses that escape, rapidly disperse into the atmosphere. Hydrogen can ignite at concentrations as low as 4% in air. The one-way valve also ensures that no air gets into the battery since the oxygen would react with the negative plates causing internal discharge. The pressure relief valve design accomplishes the following functions:

- Limits the maximum internal pressure of the cell
- Maintains a minimum internal pressure to promote recombination and minimize water loss
- Prevents atmospheric oxygen from entering the cell and discharging the negative plates

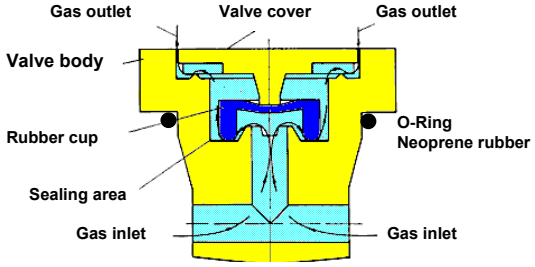


Figure 23.9. *Valve system for sealed batteries.*

Due to an oxygen recombination efficiency of about 99%, the battery barely loses electrolyte and requires no addition of distilled water throughout its service life.
Both AGM and Gel cells employ a self-resealing, spark arresting, venting valve. The labyrinth system construction of the valve avoids any outside spark or fluid from penetrating into the battery via the valve.
The principle of the valve used for VRLA-batteries is shown in figure 23.9, along with its adapter, which depends on the battery type.
The valves operate with the following pressures, p:
Opening: $0.180 \geq p \geq 0.06$ bar: 1 to 6 psi (7 to 44kPa, 0.07 to 0.44 bar)
Closing: $0.155 \geq p \geq 0.03$ bar

vi. Cases
The moulded battery case, with cell dividers is ideally formed from flame retardant, non-conductive, translucent ABS (or PVC, polycarbonate, polystyrene, styrene-acrylonitrile) plastic resin and epoxy resin (or ultrasonic or heat) sealed, both materials being acid resistant.

23.4.3 Characteristics of the flooded lead acid cell

i. The chemistry of discharge
During discharge, both the PbO_2 (lead dioxide) of the positive plate and the Pb (spongy lead) of the negative plate are converted into $PbSO_4$ (lead sulphate), whilst sulphuric acid H_2SO_4 in the electrolyte is consumed. Specifically, lead dioxide (positive plate) and lead (negative plate) react with sulphuric acid to create lead sulphate, water and energy. This electrochemical reaction causes a reduction of the specific weight of the electrolyte, as the sulphuric acid contained in the electrolyte passes into the plates during discharge. These two plate processes, during discharging, are shown in figures 23.10 and 23.11a.
These processes are reversed during the charging phase.
The lead-acid battery performs best at a slow 24-hour discharge. Pulsed discharge is also effective because the intervening non-discharge periods allow dispersion of the depleted acid back into the electrode plate. A discharge at 1C of the rated capacity (Ah for 1 hour) yields the poorest efficiency. The lower level of conversion, or increased polarization, manifests itself in a momentary higher internal resistance due to the depletion of active material in the reaction.
The lead-acid battery has a relative low energy density, making it unsuitable for portable devices. In addition, the performance (both charging and discharging) at low temperatures is marginal.

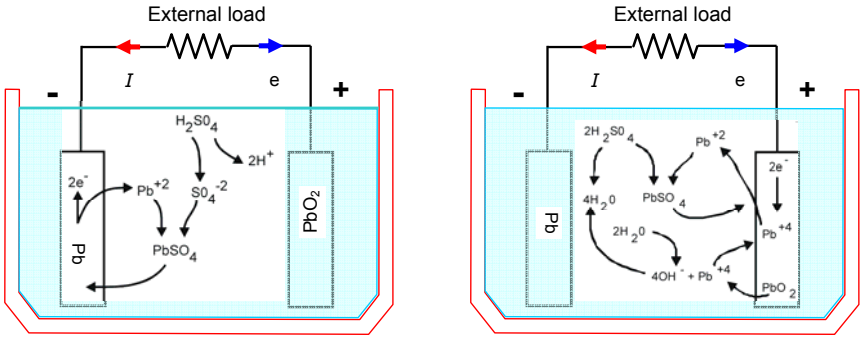


Figure 23.10. *Lead acid battery plate discharge reaction: (a) negative plate and (b) positive plate.*

ii. The chemistry of charge
During the charging phase, shown in figure 23.11b, which is the reverse of the discharge phase, the $PbSO_4$ (lead sulphate) of the positive plate oxidizes and reforms as PbO_2, while in the negative plate, the $PbSO_4$ (lead sulphate) re-forms as Pb (spongy lead). That is, the lead sulphate and water are electro-chemically converted to lead, lead oxide and sulphuric acid by the external charging source.
The general formula in Table 23.4, concerning the transformations occurring during the charge/discharge phases, correspond to an electric quantity of 2F (Farads) or 53.6 Ahr (Ampere-hour).
For a discharge reaction to occur, active materials are required in the ratio of 239.2 grams of PbO_2, 207.2 grams of Pb, and 196.2 grams of H_2SO_4. The same weight ratio holds for the charge reaction.

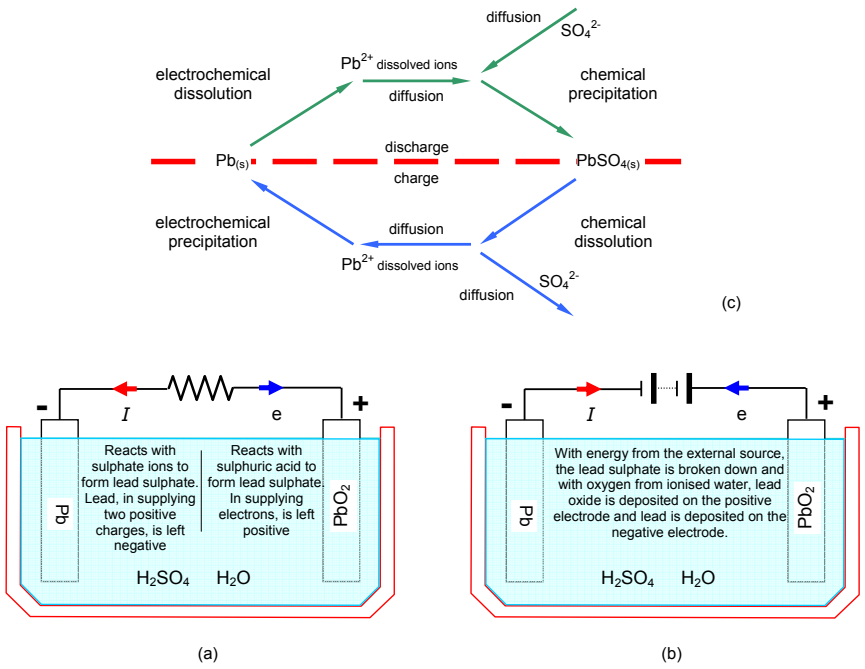


Figure 23.11. *Lead acid battery plate reaction when being:*
(a) discharged, (b) charged, and (c) charge/discharge processes in the lead negative anode.

The lead-acid cell does not lend itself to fast charging, with a typical charge time of over 8 hours. A periodic fully saturated charge is essential to prevent sulphation and the battery must always be stored in a fully charged state. Leaving the battery in a discharged condition causes sulphation and a recharge may not be possible.
Finding the ideal charge voltage limit is critical. A voltage above 2.40V/cell produces good battery performance but shortens the service life due to grid corrosion of the positive lead plate, as shown by the cathode chemical equation in Table 23.6. A low voltage limit is subject to sulphation on the negative plate. Leaving the battery on float charge for a prolonged time does not cause damage.

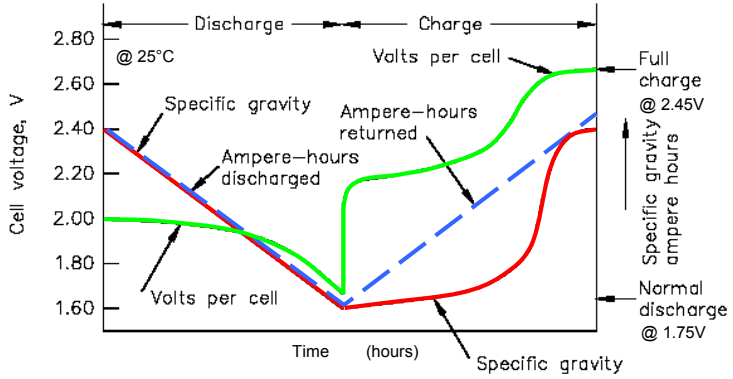


Figure 23.12. *Electrolyte specific gravity of a lead acid cell during discharge, then charging.*

iii. Specific Gravity of a battery
When testing a non-sealed (flooded) battery with a hydrometer, it is the amount of sulphuric acid in the electrolyte that is being measured. A low specific gravity reading means that the chemistry that makes electrons is deficient. A sulphate (from the acid electrolyte) is deposited on the battery plates when discharging but recharging the battery returns the sulphate to the electrolyte. During discharge, the specific gravity decreases linearly with the ampere-hours discharged as indicated in figure 23.12. It increases during recharging. Specific gravity, which decreases with temperature, is thus used as an indicator of the state of charge of a cell, not its energy capacity.
The negatively sloping specific gravity line during discharge is approximated by [acid% = 90 x (SG - 1)]:

Specific gravity = cell open-circuit voltage - 0.845

or

Cell open circuit voltage = specific gravity + 0.845 = 1.11 x acid% + 1.845.

A specific gravity range of 1.210 (19% acid) to 1.240 (21% acid) is usual for vented cells (at 25°C). VRLA batteries commonly use electrolyte with a lower specific gravity of 1.250 (22% H_2SO_4 acid to 78% water) to 1.300 (27% acid) to compensate for the smaller quantity of electrolyte.

When selecting a battery for a specific application, the effects of operating with a high or low specific gravity to be considered are:

Higher specific gravity	Lower specific gravity
Higher capacity	Lower capacity
Shorter life	Longer life
Less space required	More space required
Higher pulsed discharge rates	Lower pulsed discharge rates
Less adaptable to 'floating' operation	More adaptable to 'floating' operation
Higher standing loss	Lower standing loss

Specific gravity measurements should be taken 72 hours after an equalizing charge and the subsequent float charge.

iv. Cycling*:* Lead-acid does not like deep cycling. A full discharge decreases battery service life. This deterioration characteristic also applies to other battery chemistries in varying degrees. To prevent the battery from being stressed through repetitive deep discharge, a heavier battery (higher Ahr rating) should be used. Lead-acid technology is inexpensive but the operational costs can be higher than a nickel-based system if repetitive full cycles are required.
Depending on the depth of discharge and operating temperature, the valve regulated lead-acid cell provides 200 to 300 discharge/charge cycles. The primary reasons for this relatively short cycle life are:

- positive electrode grid corrosion;
- active material depletion; and
- positive plate expansion.

These deteriorating changes are more prevalent at higher operating temperatures. The optimum operating temperature for a lead-acid battery is 25°C with higher temperatures reducing longevity. As a guideline, every 8°C temperature rise halves remaining battery life.
Cycling does not prevent or reverse the deteriorating trend.

v. Self-discharge*:* Self-discharge of a flooded cell is about 40%/year, which is low for rechargeable batteries. In comparison, nickel-cadmium self-discharges the same amount in three months.

vi. Flooded cell plates: Sulphation occurs on the negative **anode** lead plate if the battery is left in a partially or fully discharged state. Due to self-discharge, large, non-conducting sulphate crystals with a low surface area, build-up and block effective recharging conduction paths. With a density of 6.287gm/cm^3, the sulphate occupies a larger volume than the original paste, hence the plates deform under the stress associated with the increased volume.
The **cathode** is lead oxidised to 80% lead oxide, with red lead, Pb_3O_4, for better conductivity. The oxide is mixed with H_2SO_4, grid pressed, and cured to form a cohesive porous solid. The service life of a lead-acid battery can generally be measured by the thickness of the positive cathode lead plates. The thicker the plates, the longer the remaining life. Thus the weight of a battery is an indication of the lead content

hence life expectancy. During charging and discharging, lead on the plates is gradually consumed and sediment falls to the bottom of the cell. Higher temperature, typically over 60°C and/or an uncharged state, accelerate corrosion of the lead oxide positive plate.
The plates of automotive starter batteries are about 1mm thick, while the typical golf cart battery will have plates 1.8 to 2.8mm thick. Forklift and traction batteries have plates that exceed 6mm. Most industrial flooded deep-cycle batteries use lead-antimony plates that improve plate life at the expense of increased gassing and water loss.

vii. Disposal*:* The high lead content makes the lead-acid battery environmentally unfriendly, although the materials are recyclable by furnace heat treatment, which recovers the metals.

23.4.4 Different lead-acid cell and battery arrangements

Three lead-acid systems, valve-regulated based, have emerged in an attempt to immobilise or stabilise the electrolyte:

- the larger Valve-Regulated-Lead-Acid (VRLA) cell, based on the flooded cell,
- the smaller ***GEL cell*** (mistakenly referred to as Sealed Lead-Acid (SLA) cell), and
- the newer absorbed glass matt (AGM) cell.

Batteries based on these cells are similar and are designed with a low over-voltage potential to prevent the battery from reaching its gas-generating potential during charge since excess charging causes gassing and water depletion. Consequently, they are never charged to their full capacity. To reduce dry-out, sealed lead-acid batteries use lead-calcium instead of the lead-antimony anode plates.
The sealed lead-acid battery is typically rated at a 5-hour @ 0.2C and 20-hour @ 0.05C discharges. Longer discharge times realise higher capacity because of lower losses. The lead-acid cell performs well on high load currents but not deep discharge.

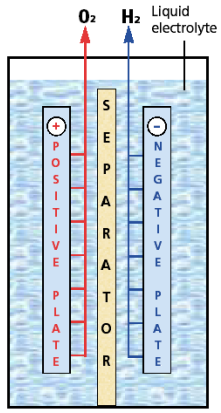
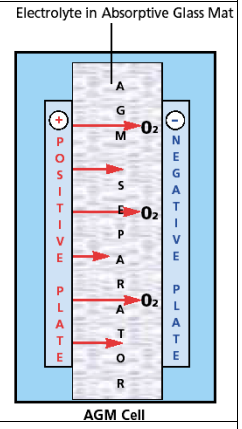
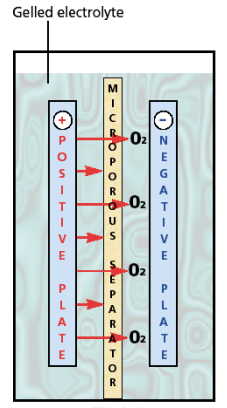
The ***gel cell lead-acid battery*** is a 'sealed' maintenance-free battery that operates in any position. The liquid electrolyte is gelled into moistened separators and the enclosure is sealed. A safety valve allows venting during charge and atmospheric pressure changes.

The ***Absorbed Glass Mat Battery*** (AGM) is a newer type of 'sealed' lead-acid battery that uses absorbed glass mats between the plates. It is sealed, durable, maintenance-free, and the sturdy plates are compressed and rigidly mounted to withstand shock and vibration. AGM batteries recombine 99% of the oxygen and hydrogen, so there is virtually no water loss.
The charging voltages are the same as for other lead-acid batteries. Even under severe overcharge conditions, hydrogen emission is below 4%. The low self-discharge of 1 to 3% per month allows long storage before recharging. The AGM costs twice that of flooded batteries of the same capacity.
The high-density packing and other cell features give the AGM battery a low internal resistance, therein allowing fast recharging. This low resistance means AGM batteries tend to stay cooler during heavy charging and discharging and maintain a higher terminal voltage under heavy loads.

The three main valve-regulated lead acid battery types are summarized and characterised in Table 23.5.

In the composite ***carbon-graphite foam grid battery*** the lead metal negative grids found in the conventional lead acid battery are replaced with lightweight carbon-graphite foam, which offers a higher energy density. The high chemical interface surface area of the carbon-graphite foam allows for greater electron flow from the battery's chemistry, and is highly resistant to sulphation, which is a common lead-acid battery failure mode. The foam is not reactive in the lead-acid environment, so does not corrode and the graphite offers better thermal conduction properties than lead. A constraint to lead-acid battery chemistry is the bounded reactive interface area between the active chemistry and the electrodes. The replacement of both the negative and positive battery plate grids with stable carbon-graphite foam grid material attempts to redress this active area constraint.

Table 23.5: Comparison of valve-regulated lead acid battery types

VALVE REGULATED LEAD ACID BATTERIES. GAS RECOMBINATION	When a charge current flows through a fully charged conventional lead acid cell, electrolysis of water occurs producing hydrogen from the negative plate and oxygen from the positive plate This results in electrolyte water loss from the cell and regular topping-up is needed. However, the evolution of oxygen and hydrogen gases does not occur simultaneously, because the efficiency of recharge of the positive plate is not as good as the negative plate. This means that during charging, oxygen is generated at the positive plate before hydrogen is evolved from the negative plate. At the same time that oxygen is generated from the positive plate, a substantial amount of highly active spongy lead exists on the negative plate before it commences hydrogen evolution. Therefore, provided oxygen can be transported to the negative plate, conditions are ideal for a rapid reaction between lead and oxygen: that is, the oxygen is electro-chemically reduced on the negative plate according to $2e + 2H+ + ½O_2 \rightarrow H_20$, where the final product is water. The external current flowing through the negative plate drives this reaction instead of hydrogen generation that would occur in a flooded cell. This process is called gas recombination. If the process were 100% efficient, no water would be lost from the cell. By careful design of the cell constituents, gas recombination of more than 95% is achieved. An efficient gas recombination cell can be made using either Absorptive Glass Mat (AGM) separators or gelled (gel) electrolyte.	 Conventional Cell
ABSORPTIVE GLASS MAT (AGM)	To achieve a high gas recombination efficiency it is necessary to create gas diffusion paths between the positive and negative plates. Then oxygen can migrate from the positive to the negative plate where it reacts with the spongy lead negative active material. In the AGM cell, a special, highly porous micro-fibre glass separator is used. By controlling the saturation level and the exact balance between electrolyte quantity and porosity, a continuous path for oxygen transport is formed. Optimum conditions for gas recombination achieve an efficiency of better than 99%. The special separator used in this battery produces a low resistance, making it good at delivering high currents. The VRLA AGM battery can be used for all discharge rates from switch closing/ tripping and engine starting which requires short duration high currents, to telecom and navigational systems requiring long, stable power occasionally for many days.	 AGM Cell
GELLED ELECTROLYTE (GEL)	For gelled electrolyte cells, a mixture of sulphuric acid with finely dispersed silica is use to produce a gel. By vigorously stirring, the thixotropic mixture stays fluid so that it can be poured into the cells. The mixture then stiffens and forms a firm gel. As the gel stiffens, it shrinks and this leads to the formation of numerous micro-fine cracks through which the oxygen generated at the positive plate can diffuse to the negative plate. Because a conventional microporous separator is used in the gelled electrolyte cell, the internal resistance of such cells is slightly higher than AGM equivalents. Therefore Gel cells are better suited to medium and long rate discharges. Tubular plate cells with gelled electrolyte are suitable for cyclic applications or where there is a need to supply power to equipment for several hours. The rod plate offers a compromise between tubular plate and pasted plate types by giving a higher power to weight and volume ratio than the tubular type but being more robust than the pasted type because the design is more corrosion resistant.	 GEL Cell

The ***bipolar lead-acid battery***
Virtually all lead-acid batteries are monopolar where a large number of plates are stacked in each cell, increasing the capacity of the battery, and cells are serially coupled to increase voltage. Since the current in a monopolar battery flows from one end of the battery to other, high currents will be unevenly distributed over the electrodes with maximum current flowing close to the posts, as shown in figure 23.13a.

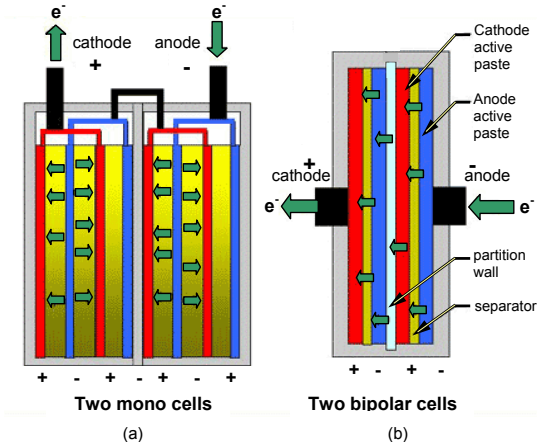


Figure 23.13. *Construction of lead-acid batteries: (a) conventional cells and (b) bipolar cells.*

A bipolar lead-acid battery is made up of a stack of series coupled bipolar electrodes. Each bipolar electrode, except the end ones, has one side of a electrical conducting partitioning wall covered with porous lead, which is the negative side of the bipolar electrode, and the other side - the positive side, covered with porous lead dioxide. Since current must pass through the end electrodes, it will flow perpendicular to all electrode surfaces, also at high currents, and the active materials are efficiently utilized with minimal internal resistance.
Lead-infused ceramic plates are coated with a positive paste on one side and a negative paste on the other. The plates are then stacked on top of each other with a separator between them, thus forming a battery in a bipolar design, as shown in figures 23.13b and 23.14. The partitioning walls are porous, lead-infiltrated ceramic plates. The bipolar plates have high corrosion resistance, and the lead surface on the bipolar plate enables good contact to the active material in the same way as standard lead acid technology. The battery contains only half the amount of lead per unit of output and offers 800W/kg discharge and 400W/kg recharge.

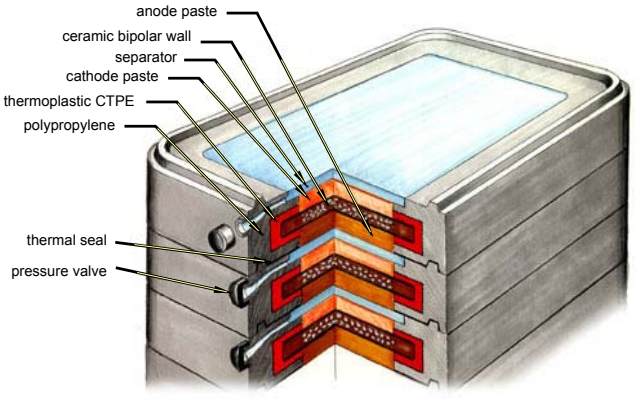


Figure 23.14. *Construction of the bipolar lead-acid battery.*

23.4.5 Lead acid battery charging and storage regimes

The charge algorithm for lead acid batteries differs from nickel-based chemistry in that voltage limiting rather than current limiting is used and VRLAs cannot be fully charged as quickly as nickel-based cells. Charge time of a valve-regulated lead acid (VRLA) cell is 12 to 16 hours, as shown in figure 23.15. With higher charge currents and multi-stage charge methods, the charge time can be reduced to less than 10 hours. A multi-stage charger applies a constant-current charge, then a topping charge, and finally a float charge. During the constant current charge period, the battery charges to 70 percent in about five hours; the remaining 30 percent is completed by a slow topping charge for a further five hours and is essential for battery long lifetime. If the battery is not completely saturated, the VRLA cell will eventually lose its ability to accept a full charge and battery performance deteriorates. The third stage is the long-term float charge, which compensates for the self-discharge after the battery has been fully charged.

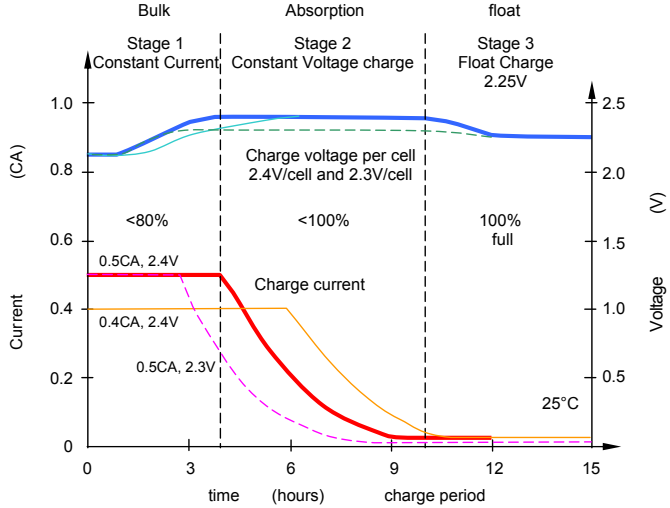


Figure 23.15. *Charge stages of a lead acid battery; constant-current charge, topping charge, and float charge.*

The cell-voltage limit is critical and typical voltage limits are from 2.30V to 2.45V. If a slow charge is acceptable, or the room temperature may exceed 30°C, the voltage limit is 2.35V/cell. If a faster charge is required, and the room temperature will remain below 30°C, 2.40V to 2.45V/cell is used. The charge voltage limit is a momentary voltage peak and the battery cannot dwell at that level. This voltage crest is only used when applying a full charge cycle to a battery that has been discharged. Once fully charged, a float charge is applied, which is a constant, lower voltage level. The float charge voltage of low-pressure lead acid cells is between 2.25 to 2.30V/cell and the optimal float charge voltage is temperature dependant. A higher temperature demands slightly lower voltages and a lower temperature demands higher voltages. Chargers with temperature sensors can optimize the float voltage.
Float charge techniques attempt to fully charge the cell to avoid sulphation on the negative plate, but cannot over-saturated the cell, which causes grid corrosion on the positive plate. In addition to grid corrosion, a high float charge contributes to loss of electrolyte. Cell aging variations result in differing optimum float charge voltages. The development of gas pockets within some cells over time causes hydrogen evolution from overcharging. Other cells undergo oxygen recombination in an almost starved state. Since battery cells are connected in series, controlling the individual cell voltages during charge is not possible. If the applied cell voltage is too high or too low, a weaker cell deteriorates further, becoming more pronounced with time. Individual cell-balancing devices can correct some of these problems if access to each cell is possible.
A ripple voltage imposed on the charge voltage causes problems for lead acid batteries, especially larger VRLA batteries. The peak of the ripple voltage constitutes an overcharge, causing hydrogen evolution; the troughs induce a brief discharge causing a starved state. Electrolyte depletion may result. Pulse-charging lead acid batteries reduces cell corrosion, but is a debated non-conclusive subject.

Non-optimal battery voltage thresholds cause a gradual decrease in capacity due to sulphation. The pressure relief valve allows some recombination of the gases during charge. A VRLA must be stored in a charged state. A topping charge should be applied every six months to avoid the voltage from dropping below 2.10V/cell.
By measuring the open cell voltage while in storage, an approximate charge-level indication can be obtained, as indicated in figure 23.21. A voltage of 2.11V, if measured at 25°C, indicates that the cell has a charge of 60 percent and higher. If the voltage is at or above this threshold, the battery is in good condition and only needs a full charge cycle prior to use. If the voltage drops below 2.10V, several discharge/charge cycles may be required to bring the battery to full performance. When measuring the terminal voltage of any cell, the storage temperature should be observed. A cool battery raises the voltage slightly and a warm one lowers it.
When charging a new lead acid battery with over-voltage, current limiting must be applied once the battery starts to draw full current. If a battery does not accept a normal charge after 24 hours under elevated voltage, a return to a normal condition is unlikely.
During prolonged float charge, a periodic topping charge, also known as an 'equalizing charge', is recommended to fully charge the plates and prevent sulphation. An equalizing charge raises the battery voltage for several hours to a voltage level above that specified. Loss of electrolyte through elevated temperature may occur if the equalizing charge is not controlled correctly. Because no liquid can be added to the VRLA cells, a reduction of the electrolyte causes irreversible damage.
Some exercise, or brief periodic discharge, is believed to prolong battery life of lead acid cells. If applied once a month as part of an exercising program, the depth of discharge should only be about 10 percent of the total capacity. A full deep discharge regular maintenance cycle decreases battery service life. Disconnecting the float charge while the VRLA is on standby is another method of prolonging battery life. Occasionally a topping charge is applied to replenish the energy lost through self-discharge. This lowers cell corrosion and prolongs battery life. Essentially the battery is maintained as if it was in storage. This is only applicable to applications that do not draw a load current during standby. In many applications, the battery acts as an energy buffer, thus needs to be under continuous charge.

Thermal runaway phenomenon
Thermal runaway is an abnormal phenomenon occurring during charging, which manifests as a distended battery. A warmer battery requires a reduced charging voltage. Thermal runaway means a state of operation where heat generation increases faster than heat dissipation, which results during severe overcharging or electrolyte dry-out. The result is an increase in battery temperature. At elevated temperature, the internal oxygen cycle is accelerated, and the developed heat causes a further increase in battery temperature. With this self-sustaining cycle, thermal runaway results, and the battery becomes physically deformed and bloated. Several precautions can avoid thermal runaway:

- Avoid battery dry-out: Do not charge at voltages higher than the gassing voltage (2.4V/cell) for too long a duration, for example, >12 hours.
- Any defective battery, for example, the short-circuited or aged battery, in a string of batteries should be removed to prevent overcharging of other batteries.
- The internal oxygen cycle reaction usually occurs in the overcharging stage, where the originally decreasing current density in the constant-voltage-charging mode may increase. If the charger cut-off condition is too low when the battery is aged, the charger overcharges the battery until thermal runaway results.

Gel batteries are much less susceptible to thermal runaway than AGM batteries. Batteries may become more susceptible with increasing age. Without a recombination reaction, flooded batteries convert most excess charging energy to gas, not heat, making them virtually immune to thermal runaway.

23.4.6 Valve-regulated battery discharge characteristics

i. Battery discharge characteristics
Through a series of discharge tests as in figure 2.16, curves can describe the expected cell capacity as a function of discharge rate and minimum allowed cell voltage. Figure 23.17 shows a typical set of battery discharge characteristic curves illustrating the effect of discharge rate on battery capacity. The minimum voltage, the discharge time, and the discharge rate are interrelated.
The discharge curves in figure 23.17 have several straight time lines (8hr, 5hr, and so on down to 1min) coloured green, radiating from the origin, with a series of voltage lines (1.75V, 1.80V, and so on up to 1.9V) shown coloured blue, intersecting the green straight time lines. The x-axis and y-axis have coordinates amperes per positive plate and ampere-hours per positive plate, respectively. A separate line, labelled the initial volts line, is shown in red at the top, with its reference y-axis, in volts, on the right. The inter-relation of these lines defines the cell's expected capability as a function of discharge rate and duration.

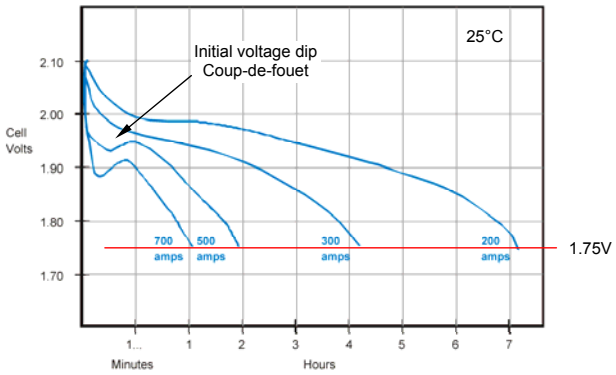


Figure 23.16. *Cell voltage during discharge with a 1.75V/cell end voltage.*

These curves are based on a specific cell size tested at a constant current discharge with a series of different discharge rates. Each tested cell initially has 100% capacity and characterisation is at the 25°C reference temperature. The total energy removed from the cell can be calculated from the constant current discharge rate and the discharge time, $E = I_{dis} \times t$.

Data is presented based either on the cell size or on the number of positive plates. Information on a per plate basis provides a common reference for a range of cell sizes in a family. Stationary cells typically have *n* positive plates and *n+1* negative plates. If data is presented in terms of the number of positive plates, then the rating of a particular cell size is the rating of a single positive plate times the number of positive plates, *n*.
There are 10 positive plates in the cell characterised in figure 23.16. The test discharge currents are divided by 10 to obtain the current per positive plate. The ampere-hours are calculated based on the discharge rate and the time to reach the specified end voltage limit, 1.75V in figure 23.16.
For any discharge rate, the cell voltage immediately falls to an initial cell voltage level, which decreases as the discharge rate increases. If the discharge rate is too high, the cell voltage will immediately fall to the defined end-point voltage, which sets the discharge limit for the defined end-point voltage.
The data is transposed onto a graph with x-coordinates of amperes per positive plate and y-coordinates of ampere-hours per positive plate. This shows the discharge capability of a single positive plate cell to a particular final voltage. The upper blue line in figure 23.17 shows the data graphed for a 1.75V final voltage. This line shows the relationship between amperes and ampere-hours for this plate. The initial volts line is drawn based on the observed initial voltage upon application of each discharge rate.
The 1.75V line shows the rate (amperes) and how much energy (ampere-hours) can be removed before the cell voltage falls to 1.75V. The process is repeated for each voltage level to create the family of capability curves shown in blue in figure 23.17.

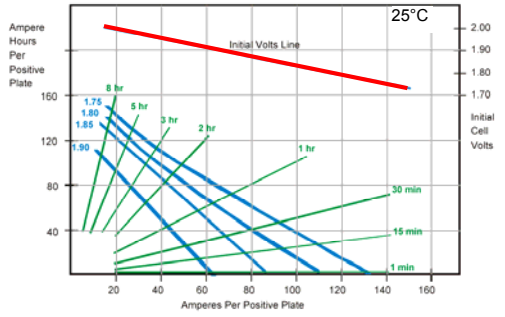


Figure 23.17. *Typical battery discharge characteristics.*

Next, the green radiating time lines are added to the plot. These time lines show how long the cell plate can provide a certain discharge rate before the voltage falls to the specified level. These green straight lines radiate radially from the origin and are based on how long it takes at a given discharge rate to remove a certain amount of energy (ampere-hours) from the positive plate.
For example, it takes 8 hours to remove 160Ahr at a discharge rate of 20A (8 hours times 20A = 160 Ahr). The 8-hour time line starts at the origin and goes through the intersection of 160Ahr and 20A.
Similarly, it takes 3 hours to remove 120Ahr at a discharge rate of 40A, or 2 hours to remove 120Ahr at a discharge rate of 60A. In each case, a time straight line can be drawn based on this relationship as shown by the green lines in figure 23.17.

The time lines overlaid onto the voltage capability curves give the characteristic discharge curves in figure 23.17.

Example 23.1: ***Lead-acid battery discharge characteristics***

Cell discharge tests produced the characteristic curves in figure 23.17, which describe the expected cell capacity as a function of discharge rate and minimum allowed cell voltage.

i. A battery must carry a load of 400A for one hour without the battery voltage falling below 1.75V per cell. How many positive plates must the battery have to fulfil the load requirement?
ii. What is the expected initial voltage for discharge rates of 70A and 110A, respectively? If a minimum cell voltage of 1.75V is allowed, what is the expected discharge rate per positive plate?
iii. Suppose 80Ahr per positive plate of energy has been removed from the cell. What is the expected voltage if the battery is continued to be discharged at a rate of 40A per positive plate?
iv. Estimate the cell internal resistance, thence the expected short circuit current, assuming a typical cell voltage of 2V. What is the internally generated power and the maximum power transfer?
v Estimate the expected short circuit current if 60 such cells are series connected to form a battery. What is the total internally generated power?

Solution

i. On figure 23.18a, follow the green one-hour time line radially out to the intersection of the blue 1.75V curve. The capability of this positive plate is 70A, or, this plate can provide 70A of current for one hour before its voltage falls to 1.75V. A total of 400A is needed for the load. By dividing the 400A load by 70A per positive plate, the required number of positive plates is 5.7.
Therefore, 6 positive plates fulfil the 400A load requirement.

ii. The red initial volts line in figure 23.18a shows the expected instantaneous cell initial voltage for a given discharge rate. This is the voltage produced immediately upon application of a load. As shown, the initial voltage is 1.89V for a discharge rate of 70A per positive plate. The voltage falls to about 1.82V for a discharge of 110A per positive plate.
There is a discharge rate that will cause the cell voltage to immediately fall to 1.75V. From figure 23.18a, this discharge rate is 140A per positive plate.

iii. The discharge curves can be used to predict the cell voltage at various times during a discharge. Figure 23.18b shows that the expected voltage can be interpolated from the voltage curves at the intersection of 80Ahr per positive plate and 40A per positive plate.
As shown, the expected voltage is about 1.86V.

iv. The initial volts line can also be used to estimate cell internal resistance. Each point on this line describes an expected instantaneous voltage for the associated discharge rate. Any two points on the initial volts line can be used to obtain a difference in voltage for a difference in discharge rate (current). Ohm's law can be applied to calculate the internal resistance.
In figure 23.18b, the initial voltage is about 1.94V for a discharge rate of 40A per positive plate and 1.83V for a discharge rate of 100A. The slope of the line is the effective internal resistance:

$$R = \frac{\Delta V}{\Delta I} = \frac{1.94\text{V} - 1.83\text{V}}{100\text{A-}40\text{A}} = 1.833\text{m}\Omega/\text{plate}$$

This is the internal resistance of this particular cell with a single positive plate (and two negative plates). As additional positive (and negative) plates are added to the cell to increase its capacity, the positive plates can be modelled as parallel resistances. If the cell has 10 positive plates (and eleven negative plates), there are 10 parallel resistances of 1.833mΩ each. The cell equivalent internal resistance is given by:

$$\text{cell resistance} = \frac{1.833\text{m}\Omega}{10} = 0.1833\text{m}\Omega$$

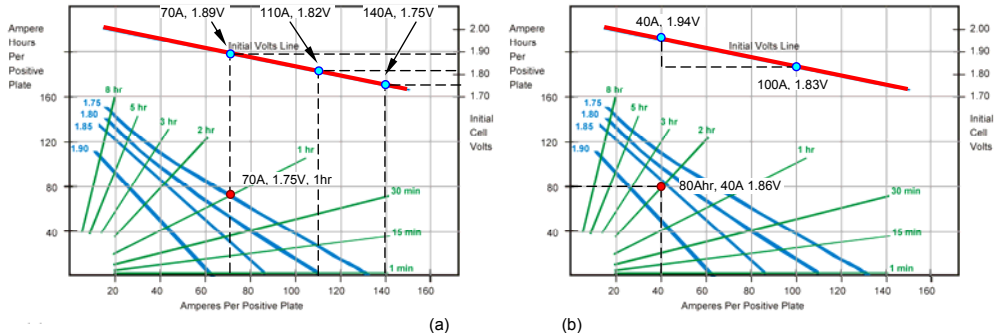


Figure 23.18. *Typical battery discharge characteristics, for Example 23.1.*

Given that the cell voltage is about 2V, the expected short circuit current from the cell is:

$$I = \frac{V}{R} = \frac{2\text{V}}{0.1833\text{m}\Omega} = 10,909\text{A}$$

The internally generated heating power is

$$P = I^2R$$
$$= 10,909\text{A}^2 \times 0.1833\text{m}\Omega = 21.81\text{kW}$$

The maximum power transfer, when the load resistance equals the internal resistance, is ¼ of 21.81kW, namely 5.45kW

v. Just as the cell voltages add when series connected, so too do the internal resistances. The short circuit current is

$$I = \frac{V}{R} = \frac{\text{number of cells} \times 2\text{V per cell}}{\text{number of cells} \times \text{internal resistance of one cell}}$$
$$= \frac{60 \times 2\text{V}}{60 \times 0.1833\text{m}\Omega} = \frac{2\text{V}}{0.1833\text{m}\Omega}$$
$$= 10,909\text{A}$$

The fault current for 60 series connected cells is the same as for each individual cell.
The total internally generated heating power P_T is 60 times that in each cell, that is

$$P_T = 60 \times I^2R$$
$$= 60 \times 21.8\text{kW} = 1.31\text{MW}$$

♣

ii. Self-discharge during storage – shelf life

LA batteries lose 2% to 3% capacity per month at 25°C when not in use, a phenomena termed self-discharge. Most batteries lose their stored energy when allowed to stand on open circuit, since the active materials are in a thermochemical unstable state. The rate of self-discharge is dependent on the cell chemistry, the quality of the active materials, as well as the temperature at which the battery is stored, as shown in figure 23.19. The use of pure raw materials decreases the rate of self-discharge and enhances storage life.
Loss of capacity during long storage or storage at high temperatures is compensated by periodically charging the battery, otherwise irreversible sulphation leads to permanent loss of capacity. If the open circuit voltage (OCV) falls below 2.1V at 20°C (approximately 60% residual capacity at a given temperature), at least every 6 months slow recharging to 2.4V per cell is necessary (for 96 hours or until the charging current remains constant for a three-hour period or falls below ½ percent of the battery's 20 hour rated capacity). If a battery is stored at temperatures above 25°C, then the boost charge interval should be more frequent. If the terminal voltage of any cell falls below 2.1V, there is a risk of open circuit corrosion or irreversible sulphation.

The impact of temperature on a battery and its rated capacity and life is based on an environment temperature of 25°C. When the environment temperature is below 25°C, the battery capacity decreases and the life extends; when the environment temperature is over 25°C, the battery capacity increases and the life shortens.

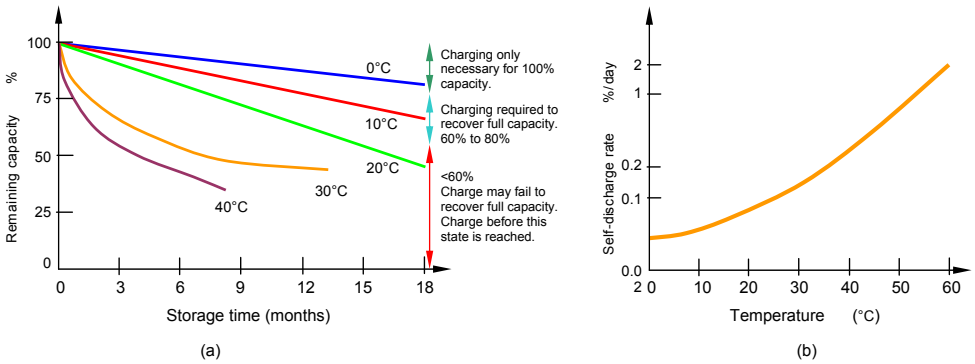


Figure 23.19. *State of charge dependency: (a) self-discharge with storage time characteristics and (b) self-discharge dependence on temperature.*

Typical self-discharge rates of lead acid batteries as a function of temperature are, 2% per month at 20°C, 4% per month at 30°C, 8% per month at 40°C, etc. That is, the self-discharge almost doubles by each 10°C rise in storage temperature.

iii. Cycle life

Primary reasons for short cycle life are grid corrosion of the positive plate, depletion of active material, and expansion of the positive plates, which are exasperated by higher temperatures; all are irreversible by charging/discharging cycling. Battery life is specified for 25°C and under float charge, halves for every 9.4°C above this temperature, as shown in figure 23.20a.

The temperature corrected number of expected years of battery life is given by

$$L_{tc} = \frac{\text{rated service life at } 25^{\circ}\text{C}}{\sum_i \frac{1}{L_i} \times months\,@\,T_i} \qquad (23.2)$$

where L_i is the % life at temperature T_i.

The expected number of cycles is dependant on depth of discharge, as shown in figure 23.20b.

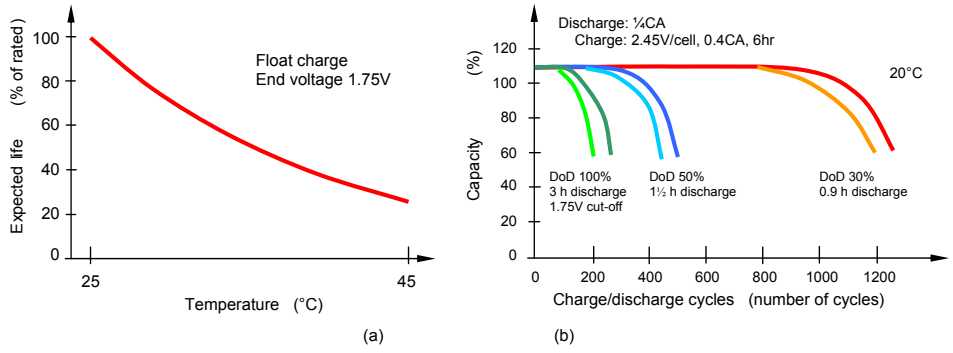


Figure 23.20. *Temperature and depth of discharge (DoD) effects on lead-acid cell life.*

Example 23.2: ***Lead acid battery life***

A battery rated for 20 years operates at 25°C for all but 3 months of the year, when it operates at 40°C. What is the expected service life of the battery?

Solution

From equations (23.2) and figure 23.20a:

$$L_{tc} = \frac{\text{rated service life at } 25^{\circ}\text{C}}{\frac{1}{L_1} \times months\,@\,T_1 + \frac{1}{L_2} \times months\,@\,T_2}$$

$$= \frac{240 \text{ months}}{\frac{1}{1} \times 9 \text{ months} + \frac{1}{0.32} \times 3 \text{ months}} = \frac{240}{18.375} = 13.06 \text{ years}$$

♣

iv. Relation between Open-circuit Voltage and State of Charge

The relation between open-circuit cell voltage and capacity retention is shown in Figure 23.21.
Open circuit voltage varies according to ambient temperature and the remaining capacity of the cell. Generally, open circuit voltage is determined by the specific gravity of the electrolyte, where discharging a battery lowers the specific gravity, as shown in figure 23.12. Consequently, it is possible to determine the approximate remaining capacity of a battery from the terminal voltage. In flooded lead-acid batteries with filling caps, with acid access, it is possible to estimate the residual capacity of the battery by measuring the density of the acid.

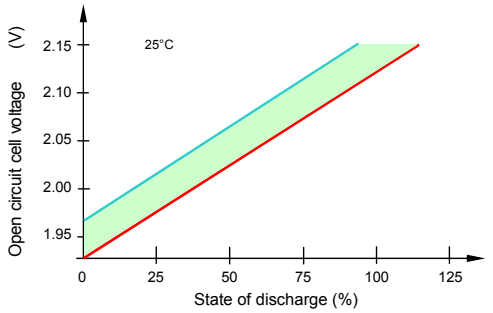


Figure 23.21. *Relationship between open-circuit voltage and state of charge (residual capacity), 24 hours after charge, or 10 minutes after discharge, at 25°C.*

However, this is not possible with valve-regulated batteries, thus leaving the value of the open circuit voltage as a method to approximate the residual capacity. The result of an open circuit voltage measurement, taken either 24 hours after a full charge or at least 10 minutes after discharge, in conjunction with figure 23.21, produces an approximation of the residual capacity.
The graph shows that a healthy, fully charged cell has an OCV of 2.14V or higher at 25ºC. Open circuit voltage varies according to ambient temperature and the remaining capacity of the battery.

v. Effects of temperature on capacity

Cell capacity is a function of ambient temperature and rate of discharge. At 20°C, rated capacity is 100%. Battery discharge is an electrochemical reaction between the electrodes (the plates) and the diluted sulphuric acid. When the discharge current is high, or the temperature is low, thereby causing a greater viscosity of the acid, the diffusion rate of the acid through the plates can no longer keep up with the discharge, reducing the capacity, as shown in figure 23.22. The capacity increases slowly above this temperature. Even at -40°C, however, a prime quality battery will still function at better than 40% of its rated capacity when discharged at the 20-hour rate, 0.05C. At any ambient temperature, the higher the rate of discharge, the lower the available capacity. This relationship is shown in Figure 23.22a. Batteries may be discharged at temperatures ranging from -40°C to 60°C, possibly 80°C when metal case cooled, and charged at temperatures from -20°C to 50°C. Whilst raising the ambient temperature increases capacity, it also decreases useful service life. Battery life is halved for each 10°C rise above normal room temperature. As temperature increases, internal resistance decreases as shown in figure 23.22b.

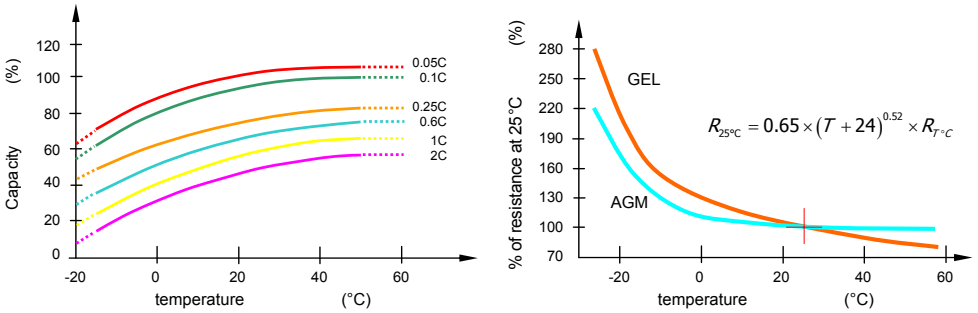


Figure 23.22. *Effect of temperature on capacity and internal resistance.*

23.4.7 Gassing and internal recombination

i. Gassing: Traditionally there are problems with the basic flooded-cell lead-acid battery design. If cell voltages exceed 2.39V, the water breaks down to hydrogen (at the anode) and oxygen (at the cathode). The chemical reactions are shown in Table 23.6. This 2.39V voltage is called the gassing voltage and is temperature and pressure dependent. Gassing requires replacement of the cell's water. Also, as the hydrogen and oxygen vent from the cell, too high a mixture concentration could cause an explosion. Another problem arising from an open system is that fumes from the acid solution can have a corrosive effect on the surrounding area.

Table 23.6: Lead-acid cell gassing and corrosion equations

undesirable plate chemical reactions		reaction
positive electrode cathode	oxygen evolution	$2H_2O \rightarrow O_2 + 4H^+ + 4e^-$
	grid corrosion	$Pb + 2H_2O \rightarrow PbO_2 + 4H^+ + 4e^-$
negative electrode anode	oxygen reduction	$O_2 + 4H^+ + 4e^- \rightarrow 2H_2O$
	hydrogen evolution	$2H^+ + 2e^- \rightarrow H_2$

These problems are basically solved with sealed cells. In the case of lead-acid cells, the term 'valve-regulated cells' is more accurate, because no rechargeable cell can be completely sealed. If sealed, the hydrogen gas pressure would build-up. Catalytic gas recombiners alleviate this problem by converting the hydrogen and oxygen back into water, with better than 90% efficiency. Although this does not entirely eliminate the hydrogen and oxygen gas, the water lost becomes so insignificant that no refill is needed for the life of the battery. For this reason, these cells are often referred to as maintenance-free batteries. Also, this cell design prevents corrosive electrolyte fumes from escaping.

ii. Theory of Internal Recombination
When an open flooded lead-acid cell is charged, a release of gas occurs. The charge current electrolyzes the water, decomposing it into its forming elements, initially oxygen from the positive electrode and subsequently hydrogen from the negative electrode. Thus water is lost from the cell, which must be replenished by means of frequent topping up with water. The nitrogen and carbon diode present, from air, are inactive.
The evolution of the two gases does not occur at the same time since the recharge efficiency of the positive electrode is less than that of the negative electrode. This means that oxygen is evolved from the positive plate before the negative plate can generate hydrogen. As oxygen is evolved from the positive plate, a significant quantity of highly active spongy lead exists on the negative electrode before the negative plate can generate hydrogen.
Should oxygen and hydrogen escape, a gradual drying out would occur, eventually affecting capacity and battery life. To maintain the chemical balance in the cell, the lost water must therefore be replaced periodically, involving time consuming verification and refilling of the electrolyte.

In valve-regulated batteries, however, the elements in the gases created are combined during the charge phase, with oxygen transported from the positive plate to the negative plate, through the so-called 'cycle of oxygen recombination', thereby producing water as described in the following cycle.

The internal design of the cell allows the oxygen generated at the positive plates to diffuse towards the negative plates, where hydrogen will be produced.

- The oxygen reacts chemically with the spongy lead of the active material to form lead oxide
- The sulphuric acid within the electrolyte then reacts with this lead oxide, forming lead sulphate and water
- The lead sulphate formed is transformed electrochemically into lead and sulphuric acid
- As long as the battery remains fully charged, this equilibrium is maintained.

The following internal recombination stages are summarised in figure 23.23.

1. At the end of charge or if overcharged, oxygen gas is generated and released at the positive plate due to water electrolysis:

$$2H_2O \rightarrow O_2 + 4H^+ + 4e^- \qquad (23.3)$$

and the oxygen is diffused through the microporous separators to the surface of the negative plates. In VRLA batteries, a densely porous medium is offered to the oxygen to facilitate its movement, namely, in AGM-batteries: pores in the glass mat; and in Gel batteries: cracks in the gel.

2. The process is reversed at the negative plate, with the recombination net effect of is a slight generation of heat, which is minimised by using high purity materials.

On the negative plates, the oxygen that has defused through the electrolyte from the positive plate chemically reacts, specifically absorption, and combines with a part of the spongy lead contained in these plates producing lead oxide:

$$2Pb + O_2 \rightarrow 2PbO \qquad (23.4)$$

3. The lead oxide combines with the sulphuric acid in the electrolyte, forming lead sulphate and water:

$$PbO + H_2SO_4 \rightarrow PbSO_4 + H_2O \qquad (23.5)$$

Part of the spongy lead is thus chemically discharged to a lead sulphate state and the water consumed at the positive plate is regenerated. Water is therefore regenerated on the positive plates, while lead sulphate is formed from the partially discharged negative plates.

4. The spongy lead which was chemically discharged at the negative plate is recharged chemically. The charge process recharges the partially discharged negative plates, thereby closing the cycle.

$$PbSO_4 + 2H^+ + 2e^- \rightarrow Pb + H_2SO_4 \qquad (23.6)$$

The recombination cycle, as described, is therefore theoretically complete as shown when equations (23.4) to (23.6) are added (see figure 23.23).

$$2H_2 + O_2 \rightarrow 2H_2O \qquad (23.7)$$

The constituent parts of water and sulphuric acid in the electrolyte, as well as the amount of negative plate lead, reappear at the end of the process in their original state, without having modified the charge conditions of the plates. The water content of the electrolyte thus remains unchanged unless the charging rate is too high. Recombination yields tend to be slightly less than complete, giving approximately 98% efficiency.
During overcharge or abnormal charge at less than the specific temperature, the amount of oxygen gas generated by reaction equation (23.3) cannot be fully-absorbed by the reaction on the negative plate, equation (23.7). The internal pressure increases, activating the safety valve, releasing the excessive gas including hydrogen generated (along with oxygen) at the negative plate during excessive overcharge.

$$\text{negative plate} \quad 2H^+ + 2e^- \rightarrow H_2 \qquad (23.8)$$

H_2 generation increases with increased float voltage, temperature, and altitude (decreased pressure), with lead-antimony generating far more gas than the equivalent calcium-lead grid cell, at the end of life.
When the safety valve operates, electrolyte is consumed and cell performance deteriorates.
In the flooded cell, oxygen and hydrogen escape to the environment, as shown in figure 23.24a. In the VRLA case, oxygen evolved from the positive plate transfers through the electrolyte to the negative plate and recombines with the generated hydrogen to form water, as shown in figure 23.24b.

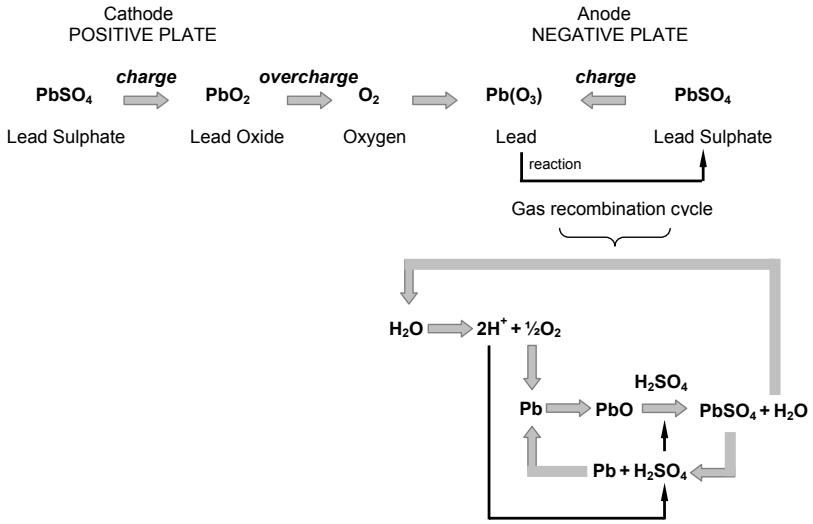


Figure 23.23. *Gas recombination reaction cycle at negative lead plate.*

iii. Necessary conditions for recombination

To facilitate the diffusion of oxygen, uniform and porous separators are used. To avoid saturating the available porosity of the separators, the quantity of electrolyte is carefully measured, ensuring that the electrolyte is completely contained inside the plates and the separators, leaving no free electrolyte inside the battery container. To prevent contact of the lead of the negative plates and the oxygen contained in the surrounding atmosphere, and the consequential chemical oxidation, the electrical elements are held in fully closed containers. At the same time, it is also necessary to allow the venting of any over-pressurization of gases, which may be generated within the container during anomalous and/or overly harsh charging conditions. The one-way valve allows excess gases to be vented when required, but does not permit outside air to enter. The presence of a one-way valve therefore gives rise to the 'valve-regulated' terminology, rather than the more commonly used, but inaccurate, 'sealed' terminology.

Since it is more volume efficient to fill a container with a liquid than a semi-solid, **AGM** batteries require less space between battery plates. The closer plate spacing gives the **AGM** battery a lower internal resistance, making it more charge efficient, with better power performance on discharge, especially at low temperatures.

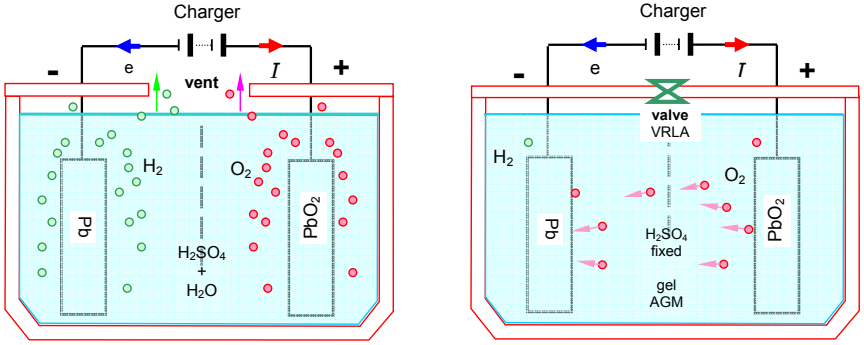


Figure 23.24. *Gassing and recombination in vented lead-acid and VRLA batteries.*

23.4.8 User properties and cell type comparisons

i. Lead-acid battery user properties

General lead-acid battery properties are summarised in Table 23.2 and in the points that follow.
Used mainly for engine batteries, the lead-acid cells represent over half of all battery sales. Some advantages are low cost, modest life cycle, and the ability to withstand electrical mistreatment. They also perform adequately in high and low temperatures and in high-drain applications.
Lead-acid cells have a low cycle life if deeply cycled and low energy densities, normally between 30 and 40Whr/kg. However, with a nominal cell voltage of 2V and power densities of up to 600W/kg, the lead-acid cell is adequate for automotive batteries. Lead acid chemistry predicts a battery theoretically capable of delivering approximately 170Whr/kg, but even the most efficient lead-acid battery produce energy densities, on average, of no more than 30-50Whr/kg.

Advantages

- Inexpensive and simple to manufacture.
- Mature, reliable and well-understood technology that is durable and dependable.
- Self-discharge is among the lowest of rechargeable battery systems. Low maintenance.
- Capable of high discharge rates, due to low internal impedance.
- No memory.

Limitations

- Low energy density limits use to stationary and wheeled applications.
- VRLA cells can never be charged to their full potential – if excessive gassing is to be avoided.
- Cannot be stored in a discharged condition - the cell voltage should never fall below 2.10V.
- A limited number of full discharge cycles, but is suited for standby applications.
- Lead content and acid electrolyte make the battery environmentally unfriendly, but recyclable.
- Transportation restrictions on flooded lead-acid cell due to environmental spillage concerns.
- Thermal runaway can occur if improperly charged.
- Heavy and bulky.

ii. Differences between VRLA batteries and traditional wet flooded batteries

Wet flooded batteries do not have pressurized sealing vents, as they do not operate on the recombination principle. They contain liquid electrolyte that can spill and cause corrosion if tipped or punctured. Therefore they are not air transportable without special containers and can only be installed 'upright'.
Wet batteries lose capacity and become permanently damaged if:

- left in a discharged condition for any length of time (due to sulphation). This is especially true of antimony and hybrid types.
- continually over-discharged, due to active material shedding, particularly automotive starting types.

The shelf life of a valve-regulated lead-acid ***VRLA battery*** is seven times longer than the shelf life of a deep cycle flooded antimony battery.
VRLA technology encompasses both gelled electrolyte and absorbed glass mat (AGM) batteries. Both have significant advantages over flooded lead acid products. A VRLA battery is a 'recombinant' battery. This means that the oxygen normally produced on the positive plates of all lead-acid batteries is absorbed by the negative plate. This suppresses the production of hydrogen at the negative plate. Water, H_2O is produced instead, retaining the electrolyte within the battery. It never needs watering, and should never be opened as this would 'poison' the battery with additional oxygen from the air.
VRLA AGM and to a lesser extent Gel technology, can be used in any orientation, but charging should be avoided in the inverted position, due to the vent orientation.

iii. Differences between gel batteries and absorbed glass mat (AGM) batteries

Both are recombinant batteries being valve-regulated lead-acid (VRLA). Both AGM and Gel batteries utilize oxygen recombination and pressure relief valves to minimize water loss and allow maintenance-free operation. AGM batteries and gel batteries are both considered 'acid-starved', between 90 and 95 percent - but not fully, saturated. The 'acid-starved' condition of gel and AGM batteries protects the plates during heavy deep-discharges. That is where the similarities end.
The **gel** battery is more starved, giving more protection to the plates; therefore, it is better suited for deep discharge applications. The electrolyte does not flow like a normal liquid; rather it has the consistency and appearance of petroleum jelly. It is a highly viscous, semisolid mixture of silica gel with dilute sulphuric acid in a colloidal suspension. The electrolyte is difficult to keep homogeneous and the solid silica can separate from the acid, creating a 'flooded' battery. Vibration is an operational factor that can cause the silica and acid mixture to separate, as there is no chemical bond. In high temperature environments, the semisolid electrolyte develops cracks and voids that reduce contact between the

plates and causes the battery to lose capacity. This same effect gradually occurs even at room temperatures.
Like gelled electrolyte batteries, absorbed electrolyte batteries **AGM** are also considered non-spillable, since the liquid electrolyte is trapped in the sponge-like matted glass micro-fibre mat separator material.

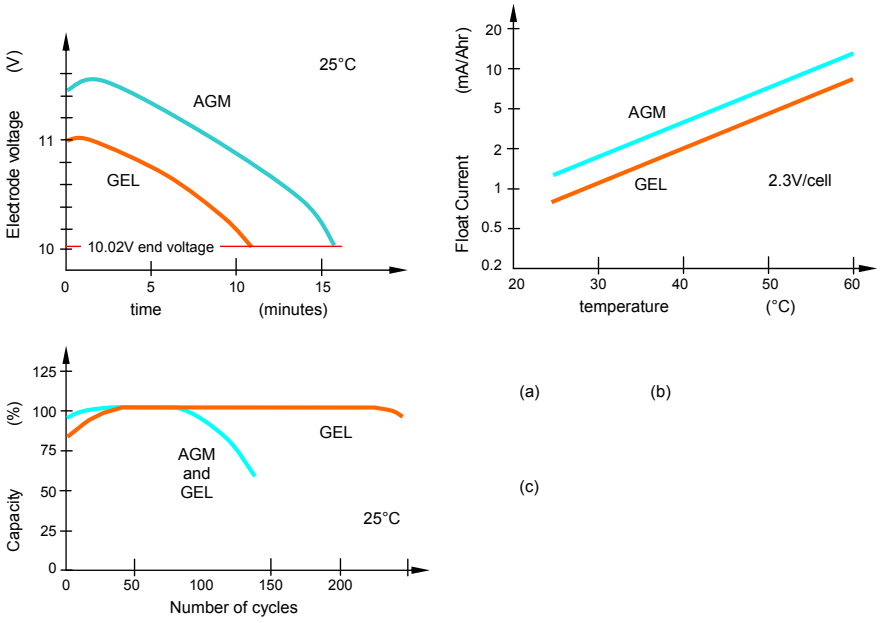


Figure 23.25. *Comparison between Gel and AGM batteries: (a) high discharge rate performance, (b) float current versus temperature; and (c) VRLA battery cycle life.*

Shrinkage of the separator does not occur as the battery ages and the electrolyte remains in direct contact with the plates. The electrolyte remains immobilized even when the battery is exposed to severe vibration, so electrolyte spillage or leakage is prevented.
Due to the physical properties of the gelled electrolyte, **gel** battery power declines faster than an **AGM** battery's as the temperature drops below 0ºC. **AGM** batteries excel for high current, high power applications, and in sub-zero environments. **AGM** batteries have the advantage of being mountable in any orientation without capacity loss, while **gel** batteries preferably should be mounted upright to prevent air pockets from forming that will burn out the plates. They have inferior performance at high discharge rates and low temperatures.
Gel batteries are more sensitive to charging voltage. If the charging voltage is not controlled within a tight range relative to the battery's temperature, the life of the battery will be adversely affected. For example, if the charging voltage is 60mV/cell higher than the recommended level, the cycle life is reduced by 60 percent. The reason for this effect is the limited oxygen recombination capability of gelled batteries. **AGM** batteries are more forgiving in overcharge conditions and their ability to recombine the hydrogen and oxygen gases back into water is more efficient. With **AGM** batteries, increasing the charging voltage 85mV/cell above the recommended charging voltage results in only a 23% reduction in the cycle life.
The charge acceptance of **gel** batteries is also less than that of **AGM** batteries. This means it takes longer to recharge **gel** batteries. As an example, when discharged to 50% of rated capacity (common in a deep cycle applications), gel batteries take twice as long to reach full charge as compared to AGM batteries.
The **AGM** VRLA battery has a slightly more efficient oxygen recombination cycle and a lower resistance than the gelled electrolyte VRLA battery. As a result, it draws slightly more float charge current, as shown in figure 23.25b, resulting in greater internal heating. The **gel**led electrolyte is in complete contact with the plates and the container walls where the heat is dissipated. The electrolyte, in the AGM case is

not in complete contact with the heat-dissipating casing, resulting in 15% less heat conduction from the plates.
In the case of the **gel**led electrolyte, the addition of phosphoric acid to the electrolyte, with special separators, minimise paste shedding, making the cell more capable of enduring the stresses of deep cycling, as shown in figure 23.25c. Adversely, the phosphoric acid reduces the initial capacity by about 10% of AGM and standard gelled cells.

The following table, 23.7 provides a one-to-one comparison between AGM and GEL batteries.

Table 23.8: Advantages and disadvantages of the different types of lead acid battery designs

	The different types of lead acid battery designs		
	Gelled Electrolyte	***Absorbed Electrolyte***	***Flooded Electrolyte***
Advantages	• Totally maintenance-free • Air transportable • Spill-proof/leak-proof • No corrosion • Installs upright or on side (side installation may lose about 10% capacity) • Compatible with sensitive electronic equipment • Very low to no gassing (unless overcharged) • Superior shelf life • Very safe at sea with no chlorine gas in bilge (due to sulphuric acid and salt water mixing) • Operates in wet environments - even under 10m of water • Rugged and vibration-resistant • Versatile: Starting, Deep Cycle, Stationary • Superior deep cycle life • Will not freeze to -30°C (if fully charged) • Superior rechargeability (from 0% to 90% in 3½ hours) • No recharge current limitation @ 2.3V/cell • Lowest cost-per-month (cost/months of life) • Lowest cost-per-cycle (cost/life cycles)	• Totally maintenance-free • Air transportable • Spill-proof/leak-proof • No corrosion • Installs upright or on side • Compatible with sensitive electronic equipment • Very low to no gassing (unless overcharged) • Superior shelf life • Very safe at sea with no chlorine gas in bilge (due to sulphuric acid and salt water mixing) • Operates in wet environments - even under 10m of water • Rugged and vibration-resistant • Excellent for starting and stationary applications • Superior for shorter duration/higher rate discharges • High charge acceptance rate • Superior under extreme cold conditions when fully charged • Superior rechargeability (from 0% to 90% in 3½ hours) • Low self-discharging rate • Lower cost than gel cell batteries	• Higher cranking amps • Certain designs are good for deep cycle applications • Excellent for starting applications • Good under extreme cold conditions when fully charged • More resilient and tolerant of improper recharge voltage • Water can be added (if accessible) • Lowest initial cost • Replacements readily available • lighter in weight
Disadvantages	• Higher initial cost • Heavier weight • Water cannot be replaced if continually overcharged • Automatic temperature-sensing, voltage-regulated chargers must be used • Charge voltage must be limited to extend life (2.3 to 2.35V/cell maximum at 25°C)	• Expensive and heavy • Shorter cycle life than gel when very deep cycled • Water cannot be replaced if continually overcharged • Automatic temperature-sensing, voltage-regulated chargers must be used • Charge voltage must be limited (2.4 to 2.43V/cell maximum at 25°C)	• Require maintenance • Spillable • Operates upright only • Shorter shelf life • Fewer shipping options • Cannot be installed near sensitive electronic equipment • Watering may be required (if accessible) • gases are generated

Table 23.7: Comparison between AGM and GEL VR LA batteries

Battery Characteristic	VRLA AGM	VRLA Gel
Electrolyte Stability	Has all of its electrolyte absorbed in separators consisting of a sponge-like mass of matted glass fibres.	Uses thixotropic gelled electrolyte.
	Excellent AGM acts like a flexible sponge. Lower acids density gives longer service life. Because acid is not immobilised, stratification can occur.	Prone to solid / liquid separation leading to spillage / spewage of acid and premature failure. Electrolyte looses contact with plates due to cracks and voids as the battery ages, especially at higher ambient temperatures.
High Rate Performance	Excellent due to low internal impedance due to close proximity plates.	Inferior Plate spacing must be greater to allow for gel passage during filling. Gel adds to impedance (presence of SiO_2), especially at low temperatures.
Sensitivity to Charging Voltage Levels	Moderately sensitive Life is reduced if charged outside of recommended charge voltage levels.	Very Sensitive Life is greatly reduced if charged outside of recommended charge voltage levels.
Charge Acceptance Rate	Excellent Battery can be fully charged (due to high acid density) in two hours if high inrush current is available.	Inferior Must limit in rush current and charge time is at least twice as long to reach full charge.
Similarities	'Sealed' using pressure valves and should never be opened. Maintenance-free. Non-spillable, and therefore can be operated in virtually any position. However, upside-down installation is not recommended, during charging. Uses a recombination reaction to avoid the escape of hydrogen and oxygen gases normally lost in a flooded lead-acid battery (particularly in deep cycle applications).	

23.5 The nickel-cadmium battery

Anode: Cadmium hydroxide (plus iron oxide for capacity stabilisation and to enhance cycle life)
Cathode: Nickel oxyhydroxide NiO-OH (plus graphite for improved conductivity) - positive plate
Electrolyte: Aqueous potassium hydroxide KOH 7M (a base solution), plus lithium hydroxide
Applications: Calculators, digital cameras, pagers, laptops, tape recorders, flashlights, medical devices (for example, defibrillators), electric vehicles, space applications
Typical ratings: Specific energy density: 45 to 80 Wh/kg
Volumetric energy density: 50 to 150 Wh/l

The cathode is nickel-plated steel, woven mesh, and the anode is a cadmium-plated steel net, termed a *pocket plate*. The active materials are retained in the pockets formed by the mesh, as shown in figure 23.26. Since the cadmium is just a coating, this cell's negative environmental impact is often overstated. Separation between plates is provided by injection moulded micro-porous polymer separator grids, integrating both plate functions of edge insulation and plate separation. Retainers are not necessary since shedding and loss of active material do not occur.

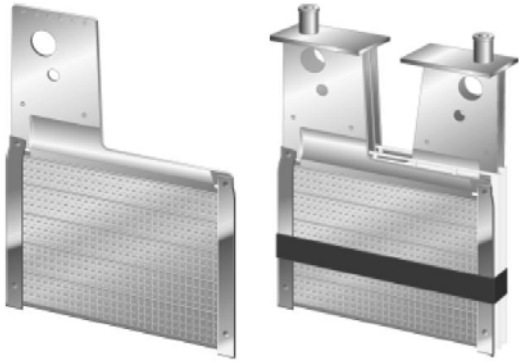


Figure 23.26. *NiCd plates.*

The electrolyte is an aqueous solution of potassium hydroxide containing small quantities of lithium hydroxide to improve cycle life and to facilitate higher temperature operation. Nickel-cadmium electrode material is less reactive with the alkaline electrolyte than lead is with acid electrolytes. During charging and discharging the electrolyte in alkaline batteries functions mainly as a carrier of oxygen or hydroxyl ions from one electrode to the other; hence the composition or the concentration of the electrolyte does not change noticeably. A cell level comparison between NiCd and lead-acid cell design characteristics is presented in Table 23.9.
Rather than venting, units are valve-sealed and the internal gases generated during charge, recombine. The casing is usually of durable polypropylene or flame retardant polyamid, nylon.

Table 23.9: Comparison of Nickel-Cadmium (NiCd) and Lead-Acid cell constructions

Parameter/Cell Dimension		Sealed Ni-Cd	Sealed Lead-Acid
Separator Thickness	mm	<1 mm	1-2 mm
Separator Material		Nylon or polypropylene	Glass microfiber
Separator Porosity	%	85-95	85-95
Electrolyte Volume	cm^3/Ah	~4	~8-10
Electrolyte in Separator	%	~10	~75
Electrolyte in Plates	%	~90	~25
Saturation Level of Separator	%	20-30	80-90
Saturation Level of Negative Plate	%	70-80	50-60
Total Cell Pore Filling	%	50-60	70-90
Negative Plate Film Thickness	mm	~0.003	~0.1
Positive Plate Film Thickness	mm	~0.01	~0.01
Electrolyte Composition	Mol	~7 M KOH	~5 M H_2SO_4
O_2 Diffusion Coefficient in Electrolyte	cm^2/s	6×10^{-6}	9×10^{-6}

The NiCd half-cell chemistry reactions are shown in Table 23.10.

Table 23.10: Half-cell electro-chemical equations for the nickel-cadmium cell

Location		Half-Reaction (discharge) 165A-hr/kg	Potential
Anode	Negative terminal	$Cd_{(s)} + 2OH^-_{(aq)} \rightarrow Cd(OH)_{2(s)} + 2e^-$	$E^-_{½cell} = -0.81V$
Cathode	Positive terminal	$2NiOOH_{(s)} + 2H_2O + 2e^- \rightarrow 2Ni(OH)_{2(s)} + 2OH^-$	$E^+_{½cell} = 0.49V$
Net REDOX reaction		$Cd_{(s)} + 2NiOOH_{(s)} + 2H_2O \rightarrow Cd(OH)_{2(s)} + 2Ni(OH)_{2(s)}$	$E^o_{cell} = 1.30V$

i. NiCd battery charging
New, unused NiCd batteries should be slow-charged for 24 hours before use. The slow charge brings each cell within a battery to an equal charge level since each cell self-discharges to a different capacity level. During long storage, the electrolyte tends to gravitate to the bottom of the cell. The initial trickle charge aids electrolyte redistribution to remedy any dry spots on the separator. Batteries are not fully form when shipped and only reach their full potential after priming through several charge/discharge cycles. 50 to 100 discharge/charges may be needed to fully form a nickel-based battery, while better quality cells perform to rated specification after as few as 5 to 7 discharge/charge cycles. Early charger readings may be inconsistent, but the capacity levels become stable once fully primed.
Nickel-cadmium prefers fast charge (10 minutes to 2 hours) to slow charge and pulse charge to dc charge. Fast charging is possible because the recharging chemical reaction is endothermic which counters the recharging I^2R losses, giving a lower cell temperature during charging. A slight capacity peak is observed between 100 and 300 cycles.

Most rechargeable cells are equipped with a safety vent to release excess pressure if incorrectly charged. The safety vent on a NiCd cell opens at 1000 to 1400 kPa (150 to 200 psi). With a resealable vent, no damage occurs on venting but some electrolyte is lost and the seal may leak afterwards, with a white salt powder accumulating over time at the vent opening.
Simple inexpensive charge termination by temperature sensing alone is not accurate. The thermistors used commonly exhibit broad tolerances; their positioning with respect to the cells is not consistent. Ambient temperatures while charging also affect the accuracy of full-charge detection. To prevent the risk of premature cut-off and assure full charge under most conditions, 50°C is used as a cut-off temperature.
More advanced NiCd charging methods sense the rate of temperature increase, defined as ***dT/dt***, or the change in temperature over charge time, rather than responding to an absolute temperature (*dT/dt* is defined as delta temperature / delta time). Although better than fixed temperature cut-off, the cells still need to generate heat to trigger detection. A temperature increase of 1°C per minute with an absolute temperature cut-off, TCO, of 60°C is used to terminate the charge. Because of the relatively large mass of a cell and the sluggish propagation of heat, the delta temperature enters a brief overcharge condition before the full-charge is detected. The *dT/dt* method is only applicable to fast charging.

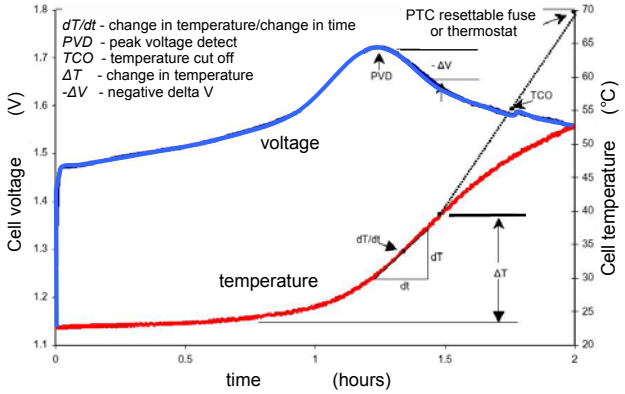


Figure 23.27. *NiCd charge termination terminology.*

Harmful overcharge occurs if a fully charged battery is repeatedly inserted for topping charge. Repetitive connection to power affects mostly 'dumb' nickel-based batteries. A 'dumb' battery contains no electronic circuitry to communicate with the charger. Li-ion chargers detect the SoC by voltage only and multiple reconnections will not confuse the charging regime. More precise full charge detection of nickel-based batteries can be achieved with electronic monitoring of the battery voltage and terminates the charge when a certain voltage signature occurs. A drop in voltage signifies that the battery has reached full charge, as shown in figure 23.27. This is known as Negative Delta V, -ΔV, NDV.
NDV, is applicable for full-charge detection for 'open-lead' NiCd chargers because of its fast response time. NDV charge detection is also applicable to a partially or fully charged battery. If a fully charged battery is inserted, the terminal voltage rises quickly, then drops sharply, triggering the ready state. Such a charge lasts only a few minutes and the cells remain cool. NiCd chargers based on NDV full charge detection typically respond to a voltage drop of 10 to 30mV per cell.
To obtain sufficient voltage drop, the charge rate must be ½C and higher. Lower charge rates produce a shallow voltage decrease that is difficult to detect, especially if the cells are slightly mismatched. In a battery that has mismatched cells, each cell reaches the full charge at a different time and the curve becomes distorted. Failing to achieve a detectable negative slope allows the fast-charge to continue, causing excessive heat build-up due to overcharge. Chargers using NDV must include other charge-termination methods, mostly battery temperature, to provide safe charging under all conditions.

The charge efficiency factor of a standard NiCd cell is better on fast charge than slow charge. At a 1C charge rate, the typical charge efficiency is 1.1 or 91%. On a slow charge, $\frac{1}{10}$C, the efficiency drops to 1.4 or 71%. At a rate of 1C, the charge time is 66 minutes at an assumed charge efficiency of 1.1. The charge time of a NiCd battery that is partially discharged or cannot hold full capacity due to memory or other degradation, is shorter. At a $\frac{1}{10}$C charge rate, the charge time of a discharged NiCd cell is about 14 hours, which relates to a charge efficiency of 1.4.

During the first 70% of the charge cycle, the charge efficiency is almost to 100%. Virtually all the energy is absorbed and the NiCd battery remains cool. Currents of several times the C-rating can be injected into a NiCd battery designed for fast charging without causing heat build-up. Ultra-fast chargers use this unique phenomenon and charge a battery to the 70% charge level within a few minutes. The charge then continues at a lower rate until the battery is fully charged.
Once the 70% charge threshold is reached, the battery gradually loses ability to accept charge, as illustrated in figure 23.28a. The cells start to generate gases, the pressure rises and the temperature increases. The charge acceptance drops further as the battery reaches 80% to 90% SoC. Once full charge is reached, the battery goes into overcharge. In an attempt to gain extra capacity, a measure of overcharge is acceptable. Figure 23.28b illustrates the relationship between cell voltage, pressure and temperature while a NiCd cell is being charged. Ultra-high capacity NiCd batteries tend to heat up more than the standard NiCd cells if charged at 1C and higher. This is partly due to the higher internal resistance of the ultra-high capacity battery. Optimum charge performance can be achieved by applying higher current at the initial charge stage, then tapering it to a lower rate as the charge acceptance decreases. This avoids excess temperature rise and yet assures a fully charged battery.

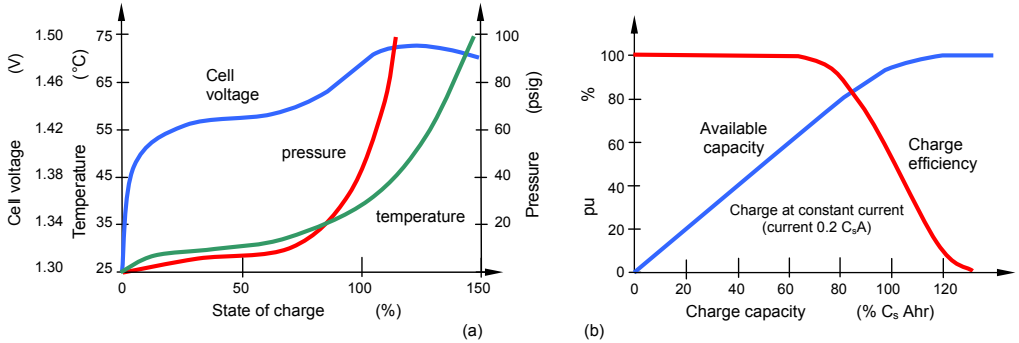


Figure 23.28. *Charge characteristics of a NiCd cell:*
(a) the cell voltage, pressure and temperature characteristics are similar to a NiMH cell and
(b) charge efficiency as a function of state of charge.

Interspersing discharging and charging pulses improves the charge acceptance of nickel-based batteries. Commonly referred to as 'burp' or 'reverse load' charge, this charge method promotes high surface area on the electrodes, resulting in enhanced performance and increased service life. Interspersed discharging also improves fast charging because it helps to recombine the gases generated during charge. Charging with the reverse load method minimizes crystalline formation. The result is a cooler and more effective charge than with dc charging with an added 15% to the life of the NiCd battery.
After full charge, the NiCd battery is maintained with a trickle charge to compensate for the self-discharge. The trickle charge for a NiCd battery ranges between 0.05C and $\frac{1}{10}$C, with lower trickle charge currents reducing the memory phenomenon. This 'floating' voltage should be in the range 1.42V to 1.45V to ensure that the battery continues to accept a small level of charge. Temperature compensation reduces battery water consumption.

ii. Balancing: Cell-to-cell balance in batteries is a major concern. Imbalances can drive one or more cells in a battery into reversal, thus causing damage and possibly resulting in initially oxygen generation at the negative electrode and then hydrogen at the positive nickel electrode. The oxygen will eventually recombine but the hydrogen will lead to pressure build-up. Although nickel-cadmium cells do generate hydrogen on normal over-charge and do gas, these occurrences are minor compared to VRLA systems.

iii. Gassing: Sealed nickel-cadmium cell technology has been developed to optimize the efficiency of the oxygen-recombination process. The chemistry is such that the cells can be operated in a starved condition (relative to valve-regulated lead-acid, VRLA, systems) and under normal operating conditions, there is no venting of gases because the cells have a thin, oxygen-permeable separator with a high void volume and an overbuilt active spongy cadmium-negative electrode with a thin electrolyte film. Unlike the lead-acid system, the primary function of the electrolyte is to provide good conductivity within the cell and only water is involved in the overall cell reaction, leaving the KOH electrolyte relatively unchanged during charge/discharge cycling. Table 23.11 compares lead acid and NiCd chemistries. Sealed nickel-cadmium cells have self-resealing safety vents that release gas due to any pressure build-up, but they are normally intended to operate at high internal pressures with minimal gassing. The positive plate is

designed to enter overcharge first, thus generating oxygen, and transport to, and recombination at the negative plate, is promoted. Because it is overbuilt, relative to the positive electrode and constantly being oxidized by oxygen, the cadmium anode (negative) electrode does not normally reach a potential where hydrogen is generated. This is also facilitated by a carefully controlled, narrow fill-weight range that is enough to provide good conductivity and small enough so the separator and plate pores are not flooded, which would lead to a pressure build-up. As summarised in Table 23.11, the reactions become:

Positive plate

$$4OH^- \rightarrow O_2 + 2H_2O + 4e^-$$

Negative plate

$$2Cd + O_2 + 2H_2O \rightarrow 2Cd(OH)_2$$

$$Cd(OH)_{2(s)} + 2e^- \rightarrow Cd_{(s)} + 2OH^-$$

Because no gases are usually given off, all of the overcharge current goes into heat generation. Therefore, charging and thermal management are critical NiCd issues; only constant-current charging is recommended for nickel-cadmium cells and only at moderate and low continuous levels, less than ⅓C (a current of one third the Ah rating for three hours).

Table 23.11: Comparison of Nickel-Cadmium and Lead-Acid Chemistries

Chemistry	Nickel-Cadmium	Lead-Acid
Negative	$Cd(OH)_{2(s)} + 2e^- \rightleftarrows Cd_{(s)} + 2OH^-$	$PbSO_{4(s)} + 2e^- + H^+ \rightleftarrows Pb_{(s)} + HSO_4^-$
Overcharge	$2H_2O + 2e^- \rightleftarrows H_2 + 2OH^-$	$2H^+ + 2e^- \rightleftarrows H_2$
Positive	$Ni(OH)_{2(s)} + OH^- \rightleftarrows NiOOH_{(s)} + e^-$	$PbSO_{4(s)} + 2H_2O \rightleftarrows PbO_{2(s)} + 3H^+ + HSO_4^- + 2e^-$
Overcharge	$4OH^- \rightleftarrows 2H_2O + O_2 + 4e^-$	$2H_2O \rightleftarrows O_2 + 4H^+ + 4e^-$
Overall Cell Process	$Cd_{(s)} + 2NiOOH_{(s)} + 2H_2O \rightleftarrows Cd(OH)_{2(s)} + 2Ni(OH)_{2(s)}$	$Pb_{(s)} + PbO_{2(s)} + 2H_2SO_4 \rightleftarrows 2PbSO_{4(s)} + 2H_2O$
Recombination Reaction	$2Cd_{(s)} + O_2 + 2H_2O \rightleftarrows 2Cd(OH)_{2(s)}$	$2Pb_{(s)} + O_2 + 2H_2SO_4 \rightleftarrows 2PbSO_{4(s)} + 2H_2O$

iv. Water consumption and gas evolution:- During charging, more ampere-hours are supplied to the battery than the discharge capacity. These additional ampere-hours must be provided to return the battery to the fully charged state and, since they do not all contribute directly to the chemical changes in the plate active materials, they are dissipated. The current associated with this surplus charge, or overcharge, breaks down the water content of the electrolyte into oxygen and hydrogen, with pure distilled water having to be added to replace this loss.
A battery that is constantly cycled, that is, charged and discharged on a regular basis, will consume more water than a battery on standby operation.

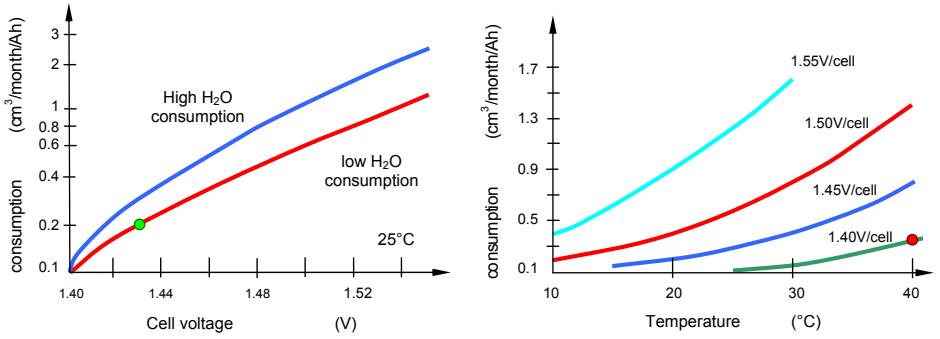


Figure 23.29. *NiCd discharge water consumption for different voltages and different (a) plate types and (b) temperatures.*

In theory, the quantity of water used can be found by the Faradic equation that each ampere-hour of overcharge breaks down 0.366cc of water. However, in practice, the water usage is less, as the overcharge current is also needed to support self-discharge of the electrodes. The overcharge current is a function of both voltage and temperature and so both influence the water consummation.
Figure 23.29 gives typical water consumption values over a range of voltages and temperatures. Gas evolution is a function of the amount of water electrolysed into hydrogen and oxygen that are predominately given off at the end of the charging period. A battery gives off no gas during discharge.
Electrolysis of 1cc of water produces about 1865cm^3 of gas mixture in the proportion ⅔ hydrogen to ⅓ oxygen. That is, the electrolysis of 1cc of water produces about 1243cm^3 of hydrogen.

Example 23.3: ***NiCd battery electrolyte life***

A NiCd, 161Ahr, low water consumption battery floats at 1.43V per cell. How long before electrolyte reserve depletion, at 25°C, if the electrolyte reserve for each cell is 500cm^3. How long before depletion if the float voltage is reduced to 1.40V/cell in an attempt to compensate for 40°C ambient conditions?

Solution

From figure 23.29a (the lower curve represents the cell with lowest water consumption), a NiCd cell at 1.43V per cell will use 0.20cm^3/month for 1Ah of capacity. Thus each cell will use 0.20cm^3/month/Ah x 161Ahr = 32.2 cm^3 per month and the electrolyte reserve will be used in

$$\frac{500\text{cm}^3}{32.2\text{cm}^3/\text{month}} = 15.5 \text{ months}$$

Gas evolution is a function of the amount of water electrolyzed into hydrogen and oxygen, which are predominantly given off at the end of the charging period. The battery gives off no gas during a normal discharge.
From figure 23.29b, at 1.4V/cell and 40°C, each cell will use 161Ahr x 0.35cm^3/month/Ah = 56.4 cm^3 / month, and the reserve will be depleted in 500cm^3 / 56.4 = 8.9 months.

♣

v. Discharge: Nickel-cadmium performs well in high-discharge and low-temperature applications, as indicated in the parts of figure 23.30. It also has a long shelf and use life but costs more than the lead-acid battery and has a lower power density but higher energy density. The nickel-cadmium is the only battery type that performs well under rigorous working conditions. All other chemistries prefer a shallow discharge and moderate load currents. Discharges of 10C can be tolerated for short durations. Although there is shift towards batteries with higher energy densities and less toxic metals, alternative chemistries cannot always match the superior durability and low cost of nickel-cadmium.
The typical short circuit current is approximately 15 to 30 times the ampere-hour capacity. The nickel-cadmium battery is tolerant to high ripple current and the only effect is increased water usage.

vi. NiCd High Current Discharge: High rate nickel-cadmium cells will deliver high currents. If the cells are discharged continuously under short circuit conditions, self-heating may incur irreparable damage. The heat problems involve the internal metal strip tab connectors overheating or the electrolyte boiling, or both. Overheating can be prevented by using the battery surface temperature to determine when to reduce the loading to allow cooling. A cut-off temperature of 60°C during discharge is acceptable. Overheating of the internal connectors ise difficult to detect, as it takes place in a few seconds or less, producing minimal overall cell temperature increase. Output capacity that is composed of pulses make it difficult to predict temperature effects accurately because there are infinite combinations of current, 'on' time, rest time, and endpoint voltage. Testing on a specific cycle is the simplest way to benchmark temperature issues.

vii. NiCd Over Discharge: When cells are connected in series and discharged completely, small cell capacity differences cause one cell to reach complete discharge first. This cell may be driven into reverse by the others. When this happens in an ordinary nickel-cadmium sealed cell, oxygen is evolved at the cadmium electrode and hydrogen at the nickel electrode. The gas pressure increases if current discharge continues, eventually the cell vents. This condition is minimized by using a reducible material in the positive plates in addition to the nickel hydroxide, to suppress hydrogen evolution when the positive plate expires. Discharging to the point of reversal should be avoided.

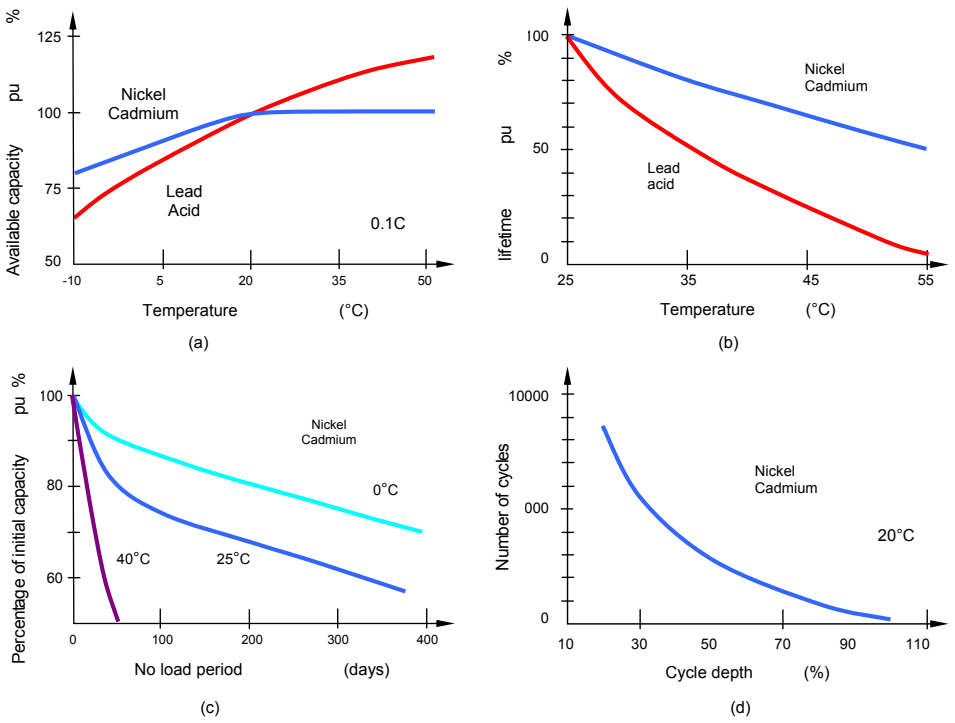


Figure 23.30. *NiCd discharge characteristics: (a) effect of operating temperature on cell capacity at 0.1C cell discharge rate; (b) effect of temperature on lifetime at 25°C; (c) no-load capacity loss; and (d) cycle life dependence on depth of discharge as a percentage of rated capacity.*

viii. Polarity reversal during over-discharge

Most applications employ multi-cell, series connected batteries. When discharging, the lowest capacity cell will be the first to experience a voltage drop. If the battery discharge continues, this unit cell will be driven into an over-discharged condition. When the cell voltage drops below 0V, its polarity is effectively reversed. The cell reactions, at different stages, as shown in figure 23.31, are as follows:

Stage 1: Initially, the positive and negative electrodes, as well as the discharge voltage, are normal.

Stage 2: The active material on the positive electrode becomes completely discharged and evolution of hydrogen occurs. Cell pressure builds up. Since the battery is designed with excess negative capacity (discharge reserve), the discharge continues; discharge voltage is around -0.2V to -0.4V.

Stage 3: The active material on both electrodes is depleted and oxygen generation starts at the negative electrode. Formation of gases at both electrodes leads to high internal cell pressure and opening of the safety vent, resulting in cell performance deterioration if this scenario occurs repeatedly.

viii. Memory: A known NiCd limitation is the *memory effect*, where the cell retains the characteristics of the previous cycle, gradually losing its useful capacity when subjected to repeated shallow cycling without being fully discharged. The cell remembers the level of discharge and the voltage of the cell emulates that of a fully discharged cell. This causes the normally microscopic cadmium hydroxide crystals to grow large, passivating the anode electrode, or the battery to wear out. In the former case, a few cycles of discharging and charging the cell will correct the problem, but may shorten the lifetime. Deep discharge is not a discharge to zero volts, but to about 1V per cell.

Nickel-cadmium is best used for deep cycling applications, and should not be used in a standby mode since it does not like being float charged.

ix. Disposal: NiCd contains toxic metals. Furnace heat treatment recovers the cadmium and iron-nickel that can be used in steel production.

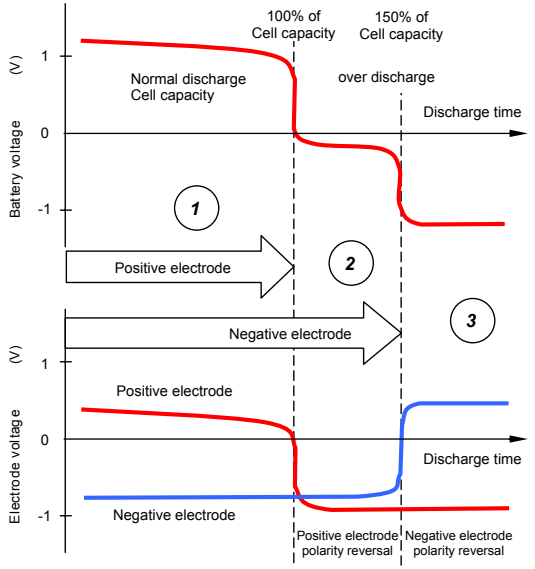


Figure 23.31. *NiCd discharge polarity reversal characteristics.*

Example 23.4: ***NiCd battery requirement***

An application requires autonomy of 3 days, but is usually only discharged to its design limit once every 2 weeks, has a normal ambient temperature of 30°C and a normal average daily load of 100W @ 48V. Determine NiCd battery requirements.

Solution

$$\text{Capacity} = A \times L \times k_t \times k_d \times k_a$$

where: A = required autonomy, hours
L = daily load, A
k_t = temperature compensation factor (see figure 23.30a)
k_d = compensation factor for maximum allowable depth of discharge (see figure 23.30d)
k_a = compensation factor for ageing (see figure 23.30b)

Autonomy period: 3 days x 24 hours = 72 hours
Daily load is: 100W/48V = 2.1A
Temperature compensation factor (figure 23.30a): 1/1.04 = 0.96
Maximum allowable depth of discharge is calculated from life requirement (20 years) and number of cycles (26 per year) = 520
Discharge depth allowable (Figure 23.30d) is 90% so the factor = 1/0.9 = 1.11
Compensation factor for aging at +30°C (figure 23.30b): 20/18 = 1.11

Battery capacity = 72hr x 2.1A x 0.96 x 1.11 x 1.11 = 178.8Ah

Number of series cells required: 48V/1.2V = 40

♣

23.5.1 Nickel-Cadmium battery properties

General nickel-cadmium battery properties are summarised in Table 23.3, illustrated in figure 23.32, and in the following points.

Advantages

- Fast and simple charge, even after prolonged storage.
- High number (over 1000) of charge/discharge cycles.
- Good load performance with recharging at low temperatures and up to 70°C.
- Not harmed by ripple current.
- Long five-year storage shelf life, in any state-of-charge, requiring priming prior to use.
- Simple storage and transportation.
- Forgiving if abused - nickel-cadmium is one of the most rugged rechargeable batteries.
- Economically priced - nickel-cadmium is lowest in terms of cost per cycle.
- Available in a wide range of sizes and performance options; most cells are cylindrical.

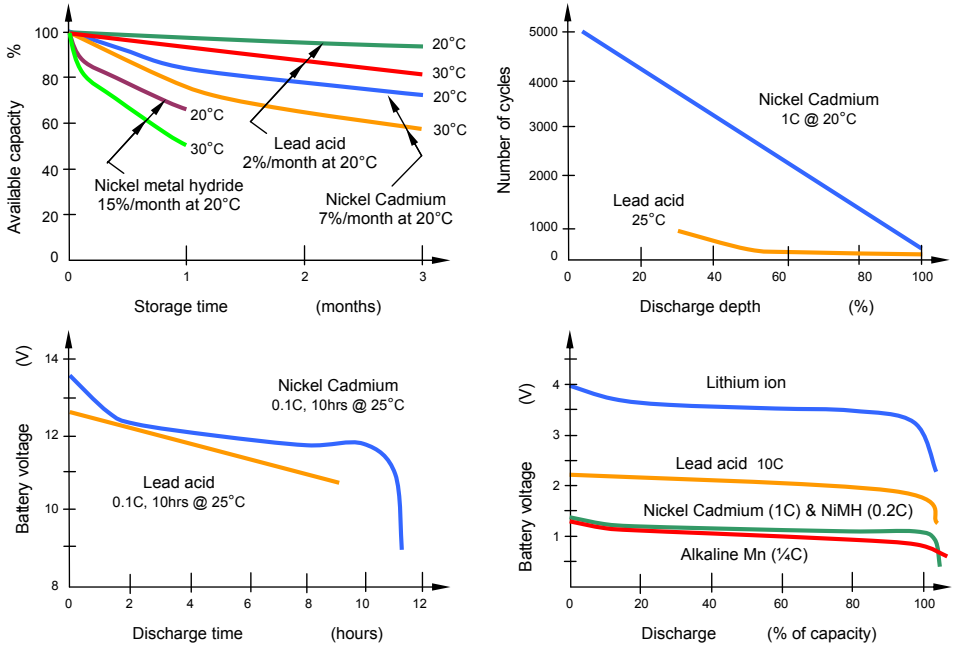


Figure 23.32. *NiCd discharge characteristics:*
(a) capacity versus storage time; (b) cycle life versus discharge depth; (c) voltage characteristics during discharge; and (d) discharge curves for five cell types.

Limitations

- Relatively low energy density, compared with new technologies.
- Memory effect but prevented by periodically discharge/charge or by the use of pocket plate technology.
- Environmentally unfriendly since it contains toxic metals.
- Relatively high self-discharge thus needs recharging after storage.
- Over-charging causes damage.

Table 23.12: Comparison between lead-acid and nickel-cadmium battery technologies

Parameter	Lead acid	Nickel cadmium
Life: float at 25°C, occasional discharge	Lifetime and performance are inter-related Thin plate: 8-10 years Thick plate (Plante) 20 years	Lifetime is independent of performance. 20 years under float conditions.
Cycle life	VRLA is not designed for cycling. Tubular plate: 700 Flat plate: 200	Pocket cell: > 1000cycles
Reliability	Cell open circuit, unpredictable	Cell short circuit, continued but decreased battery performance.
Performance	Current over 7 times Ah capacity	Current over 15 times Ah capacity
Physical size	VRLA compact, Flooded cells three times larger	NiCd larger than VRLA and smaller flood LA cells
Water	VRLA – no water addition	Occasional water addition, yearly
maintenance	VRLA cleaning and testing	Cleaning, inspection, and occasional water additions
High temperature operation > 20°C	50% lifetime reduction for every 10°C temperature increase	20% lifetime reduction for every 10°C temperature increase
low temperature operation < 0°C	Decreased performance and capacity. Possible plate damage.	Less performance fall off. No damage down to -30°C.
Electrical abuse	Not ripple tolerant. Over discharge can cause permanent damage.	Ripple tolerant Over charge and discharge possible
Mechanical abuse	Vibration cause shedding and capacity loss	High vibration resistance
Relative costs Initial and life cycle	Low initial cost but restricted lifetime thus reducing life cycle cost.	Higher initial cost but restricted lifetime but superior lifetime and characteristics giving lower life cycle cost.

23.6 The nickel-metal-hydride battery

Anode: Rare-earth or nickel alloys with various intermetals
Cathode: Nickel oxyhydroxide $Ni(OH)_2$
Electrolyte: Potassium hydroxide KOH 6M
Applications: Mainly satellite application, formerly cellular phones, camcorders, emergency backup lighting, power tools, laptops, portable, electric vehicles
Typical ratings: Specific energy density: 60 to 120 Wh/kg
Volumetric energy density: 150 to 300 Wh/l

The original battery chemical form was nickel-hydrogen. Although hydrogen has superior anodic qualities, exploitation was restricted since nickel-hydrogen requires cell pressurization. Nickel-hydrogen not only require high-pressure steel canisters batteries but are bulky, and cost thousands of dollars per cell. Nickel-hydrogen is used mainly for satellite applications.
Some metal alloys (hydrides such as $LaNi_5$ or $TiNi_2$) can store hydrogen atoms, which then can participate in reversible chemical reactions. In NiMH batteries, the anode consists of metal alloys, including AB_2 (***A***:-V, Ti; ***B***:- Zr, Ni, plus Mn, Cr, Co, and Fe giving $TiNi_2$) and rare earth/nickel alloys AB_5 (***A***:-La, cerium, neodymium, praseodymium; ***B***:- Ni, Mn, and Co giving $LaNi_5$).
Metal hydride cell chemistry depends on the ability of some metals to absorb large quantities of hydrogen. These metallic alloys, termed hydrides, can provide a storage sink for hydrogen that can reversibly react in battery cell chemistry. Such metals or alloys are used for the negative electrodes. The metal hydride electrode has a theoretical capacity approximately 40% higher than the cadmium electrode used in the NiCd cell.

The positive electrode is nickel hydroxide as in NiCd cells.

Except for the anode, the NiMH cell closely resembles the NiCd cell in construction. Even the voltage is virtually identical, at 1.2V, making the cells interchangeable in many applications. Thus the sealed cell construction is a hybrid of the NiCd and NiH_2 cell structures.

The cell chemistry reactions are shown in Table 23.13.

Table 23.13: Half-cell electro-chemical equations for the Nickel-metal-hydride cell

Location		Half-Reaction (discharge)	Potential
Anode	Negative terminal	$MH_{(s)} + OH^- \rightarrow M_{(s)} + H_2O + e^-$	$E^-_{½cell} = -0.83V$
Cathode	Positive terminal	$NiOOH + H_2O + e^- \rightarrow Ni(OH)_2 + OH^-$	$E^+_{½cell} = 0.52V$
Net		$NiOOH + MH_{(s)} \rightarrow Ni(OH)_2 + M_{(s)}$	$E^o_{cell} = 1.35V$

When a NiMH cell is charged, the positive electrode releases hydrogen into the electrolyte. The hydrogen in turn is absorbed and stored in the negative electrode. The reaction begins when the nickel hydroxide, $Ni(OH)_2$, in the positive electrode and hydroxide, OH^-, from the electrolyte combine. This produces nickel oxyhydroxide, NiOOH, within the positive electrode, water, H_20, in the electrolyte, and one free electron, e^-. At the negative electrode the metal alloy, M, in the negative electrode, water, H_20, from the electrolyte, and an electron, e^-, react to produce metal hydride, MH, in the negative electrode and hydroxide, OH^-, in the electrolyte. See Table 23.13 - chemical equations and figure 23.33 - transport diagram. Because heat is generated as a part of the overall chemical reaction during NiMH cell charging, the charging reaction is exothermic. As a cell is charged, the generation of heat does not accumulate if it is effectively dissipated. Extreme elevated temperatures may be experienced if a cell is excessively overcharged.
The electrolyte, which is a hydrogen absorbent aqueous solution such as potassium hydroxide, takes no part in the reaction but serves to transport the hydrogen between the electrodes.
Some of the alloy metals absorb heat when absorbing hydrogen, while others give off heat when absorbing hydrogen. Both are undesirable for a cell, since hydrogen transfer should occur without any energy transfer. The better alloys are combinations of exothermic and endothermic metals in order to achieve a zero thermal specification. The success of NiMH battery technology comes from the rare earth, hydrogen-absorbing alloys (commonly known as Misch metals) used in the negative electrode. These metal alloys contribute to the high energy density of the NiMH negative electrode thus increasing the volume available for the positive electrode.

Table 23.14: Hydrogen Storage Metals Comparison

Material	Density	H_2 Storage Capacity
	g/cc	g/cc
$LaNi_5$	8.3	0.11
FeTi	6.2	0.11
Mg_2Ni	4.1	0.15
Mg	1.74	0.13
MgNi Eutectic	2.54	0.16
liquid H_2	0.07	0.07

NiMH batteries are mostly of the rare earth-nickel type, of which $LaNi_5$ is representative, as shown in Table 23.14. These alloys can store six hydrogen atoms per unit structure cell giving $LaNi_5H_6$. The electrolyte of NiMH batteries is typically 6M KOH.
The NiMH cell cost more and has half the service life of the NiCd cell, but it also has 30% more capacity, the increased power density is theoretically 50% more, while the energy density is 40% higher. The memory effect, which was at one time thought to be absent from NiMH cells, is present if the cells are sequenced just right. The memory effect is avoided if the cell is fully discharge once every 30 cycles.

The success of nickel-metal-hydride has been driven by high energy density and the use of environmentally friendly metals.

23.6.1 Nickel-metal-hydride battery properties

i. Charging: Although a battery may be able to operate at cold temperatures, this does not mean that charging is possible under such conditions. The charge acceptance for most batteries at low temperatures is confined and must be brought up to temperatures above 0°C for charging. Nickel-cadmium can be recharged at below 0°C provided the charge rate is reduced to $\frac{1}{10}$C.

A charger designed for NiMH batteries can also charge NiCd cells, but not the other way around, since a charger only made for NiCd batteries could overcharge a NiMH battery.
Nickel-cadmium cells are endothermic on charge while nickel-and traditional nickel-cadmium cells are physically similar, there are significant differences in behaviour on charge between the two cell types that relate to the electrochemical fact that metal hydride cells are exothermic.

The effect of temperature on charging efficiency (the increase in cell capacity per unit of charge input) is an area of difference between nickel-metal hydride and nickel-cadmium cells. Specifically charge acceptance in the nickel-metal hydride cell (as shown in figure 23.34a) decreases monotonically with rising temperature beginning below 20°C and continuing through the upper limits of normal cell operation. This contrasts with the nickel-cadmium cell, which has a peak in charge acceptance at near room temperature. With either cell type, the drop in charge acceptance at higher temperatures remains a significant concern for cells in close proximity to heat sources or with limited cooling or ventilation.
Figure 23.34b indicates that the charge acceptance efficiency for the nickel-metal hydride cell is improved as the charging rate is increased.

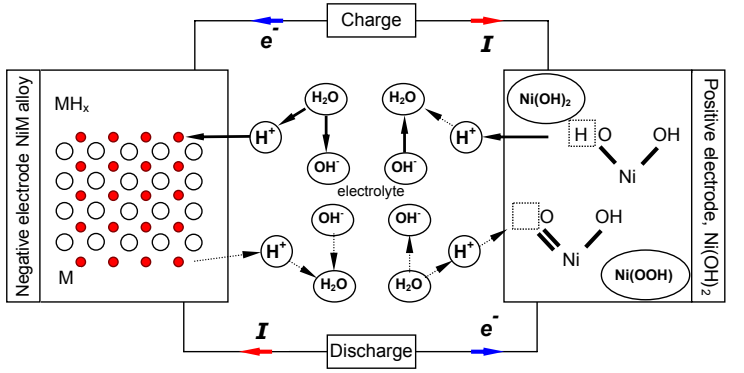


Figure 23.33. *NiMH cell transport diagram for NiMH charging and discharging.*

ii. Overcharge: Nickel Metal Hydride cells are designed with an oxygen-recombination mechanism that slows the build-up of pressure caused by overcharging. The overcharging of a cell occurs after the positive electrode

- no longer has any nickel hydroxide to react with the hydroxide from the electrolyte and
- begins to evolve oxygen.

The oxygen diffuses through the separator where the negative electrode recombines the oxygen with stored hydrogen to form excess water in the electrolyte. If this oxygen-recombination occurs at a slower rate than the rate at which oxygen is evolved from the positive electrode, the result is in a build-up of excess oxygen gas resulting in an increase in pressure inside the cell. To protect against the first stages of overcharge, NiMH cells are constructed with the negative electrode having a capacity (or active material) greater than the positive electrode. This slows the build-up of pressure by having more active material available in the negative electrode to effectively recombine the evolved oxygen.
Excessive overcharging of a NiMH cell can result in permanent loss in capacity and cycle life. If a cell is overcharged to the level at which the pressure begins to build up, elevated temperatures are experienced causing the separator to lose electrolyte. The loss of electrolyte within the separator (or 'separator dry out') inhibits proper transport of hydrogen to and from the electrodes. Furthermore, if a cell is severely overcharged and excessive amounts of oxygen gas are evolved, the pressure may be released through the safety vent in the positive terminal. This removes elements from within the cell that are in fact needed for proper operation. To protect against the damaging effects of overcharging, specific charge terminations are used.

The oxygen recombination at the negative electrode occurs simultaneously via two reactions:

$$4MH + O_2 \rightarrow 4M + 2H_2O$$
$$O_2 + 2H_2O + 4e^- \rightarrow 4OH^-$$

The first equation represents a direct combination of the O_2 gas with MH, which is present in significant amounts at the negative electrode of a fully charged cell. The second equation is a reverse of the electrolysis reaction that originally generated the O_2 at the positive electrode. The result of these two equations is that gaseous O_2 is reabsorbed by the negative electrode, thereby preventing unacceptably high internal pressure during the charging reactions.

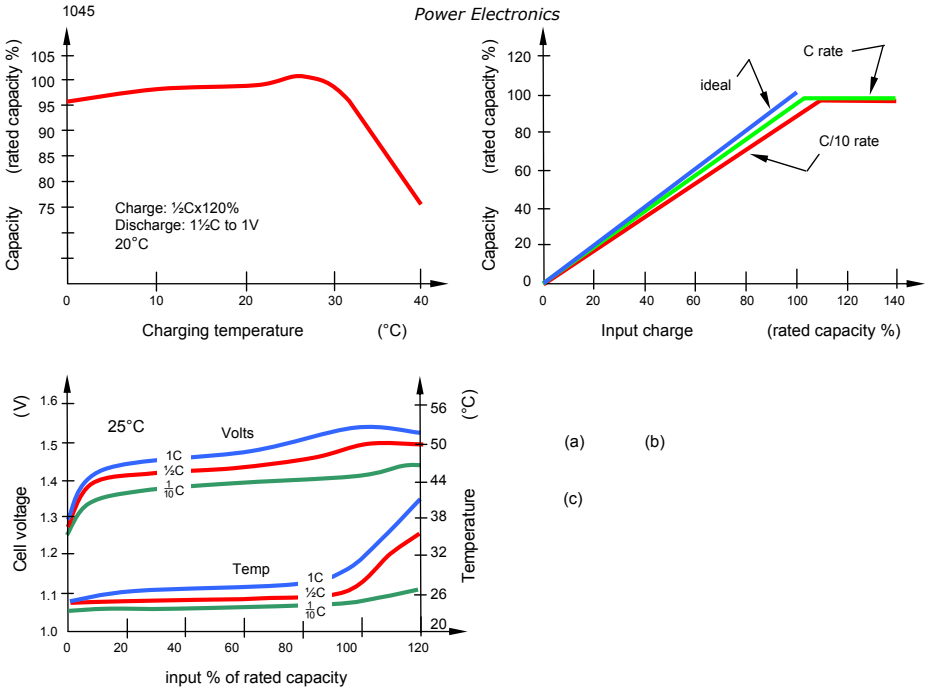


Figure 23.34. *Effect of charge temperature and charge rate on charge capacity.*

iii. Discharge: When a NiMH cell discharges, the chemical reactions are the reverse of what occurs when charging. Hydrogen stored in the metal alloy of the negative electrode is released into the electrolyte to form water. This water then releases a hydrogen ion that is absorbed into the positive electrode to form nickel hydroxide. See Table 23.13 - chemical equations and Figure 23.33 - transport diagram. For NiMH cells, the process of moving or transporting hydrogen from the negative electrode to the positive electrode absorbs heat and is therefore endothermic. Heat continues to be absorbed until the cell reaches a state of over discharge, where a secondary reaction occurs within the cell resulting in a temperature rise.

Nickel-metal-hydride is less durable than nickel-cadmium. Cycling under heavy load and operation and storage at high temperature reduces the service life and rapidly degrades performance. If charged and discharged at 45°C, the cycle life is half of that at room temperature. Nickel-cadmium is also affected by high temperature operation, but to a lesser extent.

At low temperatures, the performance of all battery chemistries drops drastically, as shown in figure 23.35. While -20°C is the threshold at which the nickel-metal-hydride, sealed lead-acid and lithium-ion battery cease to function, the nickel-cadmium can go down to -40°C. At that frigid temperature, the nickel-cadmium is limited to a discharge rate of $\frac{1}{5}$C (5 hour rate).

While NiCd and Li-ion are robust and show minimal deterioration when pulse discharged, NiMH exhibits a reduced cycle life when powering a pulsed load.

Nickel-metal-hydride cells suffer from high self-discharge, higher than that of nickel-cadmium cells.

iv. Over Discharge: There are two phases to the over discharging of a NiMH cell. The first phase involves the active material of the positive electrode becoming fully depleted and the generation of hydrogen gas begins. Since the negative electrode has more active material (metal hydride), it has the ability to absorb some of the hydrogen gas evolved by the positive electrode. Any hydrogen not absorbed by the negative electrode begins to build up in the cell generating pressure. The second phase begins when the entire negative electrode is fully depleted of active material. Once both electrodes are fully depleted, the negative electrode absorbs oxygen contributing to the loss of useable capacity. Extreme over discharge of a NiMH cell results in excessive gassing of the electrodes resulting in permanent damage in two forms. First, the negative electrode is reduced in storage capacity when oxygen permanently occupies each hydrogen storage site, and second, excess hydrogen is released through the safety vent reducing the amount of hydrogen inside the cell. To protect against the damaging effects of over discharging, proper end of discharge terminations must be used.

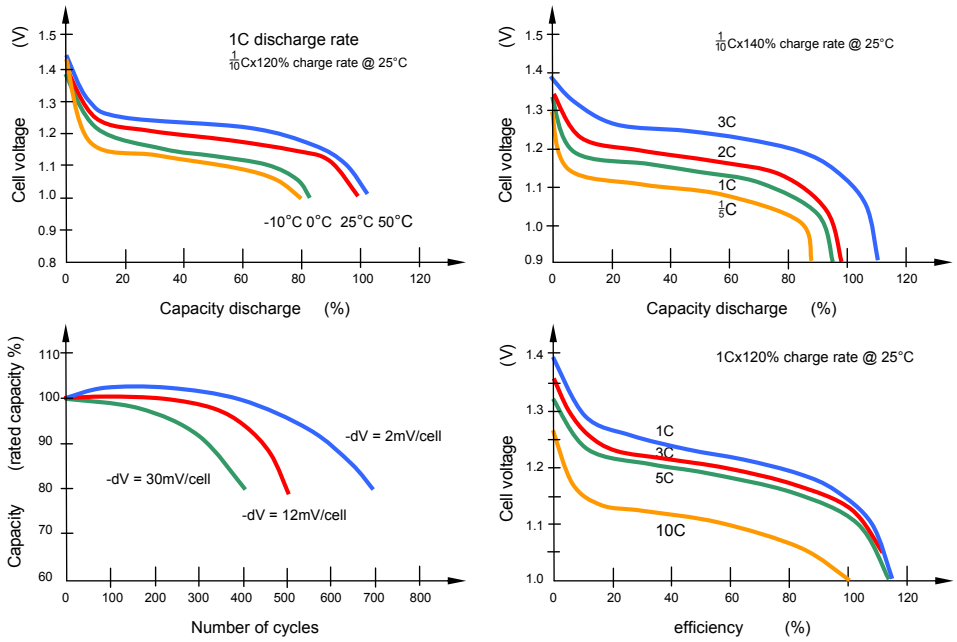


Figure 23.35. *Effect of charge temperature and discharge rate on discharge capacity.*

v. NiMH cell charging

Based on figure 33.36, NiMH battery charging is similar to that for the NiCd battery but involves slightly more complex control. A number general charging strategies have evolved.

The ***Two-Stage*** approach uses a timer to switch from the initial charge rate to the maintenance charge rate. Because there is no sensing of the cell's transition into overcharge, the charge rate must be kept low ($\frac{1}{10}$C) to minimize overcharge-related impact on cell performance and life. Charge durations are typically set at 16 to 24 hours to ensure full recharge in cases of complete discharge. Although economical, since this scheme makes no allowance for the degree of discharge or for environmental conditions, its use is rarely recommended for typical nickel-metal hydride cells.

A ***Three-Stage*** fast charge restores approximately 90 percent of the discharged capacity, an intermediate timed charge completes the charge and restores full capacity, then a maintenance charge provides a continuous trickle current to balance the cells and compensate for self-discharge. The fast charge (with currents in the 1C range) is typically switched to the intermediate charge using a temperature-sensing technique that triggers at the onset of overcharge. The intermediate charge normally consists of a $\frac{1}{10}$C charge for a timed duration selected based on battery configuration. This intermediate-charge replaces the need to fast-charge deeply into the overcharge regime to ensure that the cell has received a full charge. Three-step charging requires greater charger complexity (to incorporate a second switch point and third charge rate) but reduces cell exposure to life-reducing overcharge.

Negative Delta Voltage charging. The NiMH cell produces a small voltage drop at full charge. This Negative Delta Voltage, NDV or -ΔV, is almost non-existent at charge rates below ½C and elevated temperatures. Aging, increased cycle count, and cell mismatch masks the already minute voltage delta, 16mV or less, which makes the use of NDV difficult for charging control. Voltage fluctuations and noise induced by the battery and charger can corrupt the NDV detection circuit if set too precisely.

Most NiMH fast chargers use a combination of NDV, voltage plateau, rate-of-temperature-increase (*dT/dt*), temperature threshold, and timeout timers. The charger utilizes whatever comes first to terminate the fast-charge.

NiMH batteries that use the NDV method or thermal cut-off control tend to deliver higher capacities than those charged by less aggressive methods. The gain is approximately 6 percent on a good battery. This capacity increase is gained if the battery is exposed to a brief overcharge, with the adverse consequence of a shorter cycle life. Rather than expecting 350 to 400 service cycles, the battery may be

depleted within 300 cycles. Similar to NiCd charge methods, most NiMH fast-charging is based on the rate-of-temperature increase (dT/dt), with a 1°C per minute temperature raise used to terminate the charge. The absolute temperature cut-off, TCO, is 60°C. A topping charge of $\frac{1}{10}$C is added for about 30 minutes to maximize the charge. Subsequent continuous trickle charging maintains the battery in a full charge state.
An initial fast charge of 1C is possible. Cooling periods of a few minutes are interjected when certain voltage peaks are reached. The charge then continues at a lower current. When reaching the next charge threshold, the current reduces further. This process is repeated until the battery is fully charged.

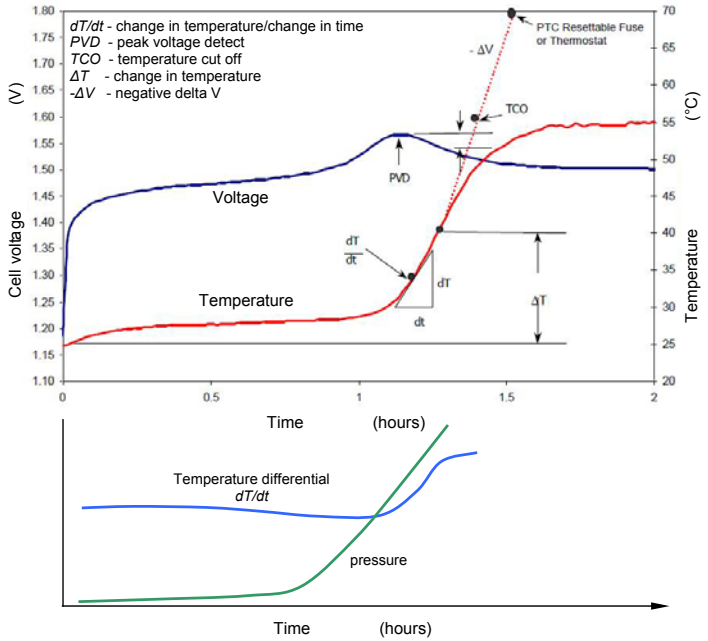


Figure 23.36. *NiMH cell charge characteristics and charge termination nomenclature.*

Step-differential charge is applicable to NiMH and NiCd batteries. The charge current adjusts to the SoC, allowing high current at the beginning and then moderate current towards the end of charge. This avoids excessive temperature increase towards the end of the charge cycle when the battery is less capable of accepting charge. NiMH batteries should be rapid charged rather than slow charged. The amount of trickle charge applied to maintain full charge is critical. Because NiMH does not absorb overcharge well, the trickle charge must be set lower than that for the NiCd cell. The trickle charge for the NiMH battery is a low 0.025C (C/40), hence a NiCd charger cannot be used to charge NiMH batteries, although the lower trickle charge rate is acceptable for NiCd cells.

It is virtually impossible to slow-charge a NiMH battery. At a C-rate of $\frac{1}{10}$C and 0.3C, the voltage and temperature profiles fail to exhibit defined characteristics to measure the full charge state accurately and charging depends on a timer. Harmful overcharge can occur if a partially or fully charged battery is charged on a charger with a fixed timer. The same occurs if the battery has lost charge acceptance due to age and can only hold 50% of charge. A fixed timer that delivers a 100% charge without regard to the battery condition ultimately applies too much charge. Overcharge can occur even though the NiMH battery does not heat up.

vi. Disposal: The individual materials are mechanically separated in a vacuum to prevent any hydrogen from escaping. The nickel component is processed for use in the manufacture of stainless steel.

vii. Uses: Nickel-metal hydride has been replacing nickel-cadmium in markets such as wireless communications and mobile computing.

23.6.2 Nickel-metal-hydride battery characteristics

Most nickel-metal-hydride shortcomings are intrinsic to nickel-based technology hence are shared with nickel-cadmium. General nickel-metal-hydride battery properties are summarised in Table 23.3 and the following comparison features are based on an equivalent NiCd cell.

Advantages

- 30 to 40%, and potentially higher energy capacity than standard nickel-cadmium.
- Less prone to memory than nickel-cadmium, requiring fewer exercise cycles.
- Simple storage and non-regulatory transportation control.
- Environmentally friendly since it contains only mild toxins; profitable for recycling.

Other general advantages are

- Can be deep cycled
- Low internal impedance
- Rapid one hour charging

Limitations

- Limited service life with performance deteriorating after 200 to 300 cycles.
- Relatively short storage of three years, increasing with cooler temperature and partial charging, 40% state of charge.
- Limited discharge current, although capable of delivering high currents, heavy load reduces battery cycle life. Optimum at $\frac{1}{5}$C to ½C (of rated capacity).
- Complex charging since nickel-metal-hydride generates more heat during charge and requires longer charge times than nickel-cadmium. Trickle charge settings are critical because the battery cannot absorb overcharge.
- High self-discharge, 10% in the first hour, then typically 50% higher than nickel-cadmium. Self-discharge is 50% higher than NiCd.
- High maintenance since it requires regular full discharge to prevent crystalline formation. Nickel-cadmium should be exercised monthly, nickel-metal-hydride quarterly.
- Performance degradation if not stored at cool temperatures in a 40% state-of-charge.
- Bulky, requires high-pressure steel canisters, and is expensive.

Table 23.15: Comparison between NiCd and NiMH Cells

Application Feature	Comparison between Nickel-Metal Hydride and Nickel-Cadmium Batteries
Nominal Voltage	Same (1.20V)
Discharge Capacity	Ni-MH up to 40% greater than NiCd
Discharge Profile	Equivalent
Discharge Cut-off Voltages	Equivalent
High Rate Discharge Capability	Effectively the same rates
High Temperature Discharge Capability	Ni-MH slightly better than standard NiCd cells
Charging Process	Generally similar; multiple-step constant current with overcharge control recommended for fast charging Ni-MH
Charge Termination Techniques	Generally similar but Ni-MH transitions are more subtle. Backup temperature termination recommended.
Operating Temperature Limits	Similar, but with Ni-MH, cold temperature charge limit is 15°C.
Self-Discharge Rate	Ni-MH slightly higher than NiCd
Cycle Life	Generally similar, but Ni-MH is more application dependent.
Mechanical Fit	Equivalent
Mechanical Properties	Equivalent
Selection of Sizes/Shapes/Capacities	Ni-MH product line more limited
Handling Issues	Similar
Environmental Issues	Reduced with Ni-MH because of elimination of cadmium toxicity concerns.

23.6.3 Comparison between NiCd and NiMH Cells

Nickel-metal hydride cells are essentially an extension of the proven sealed NiCd cell technology with the substitution of a hydrogen-absorbing negative electrode for the cadmium-based electrode. While this substitution increases the cell electrical capacity (measured in ampere-hours) for a given weight and volume and eliminates the cadmium which raises toxicity concerns, the remainder of the nickel-metal hydride cell is quite similar to the nickel-cadmium cell. Many application parameters vary little between the two cell types, and replacement of nickel-cadmium cells in a battery with nickel-metal hydride cells usually involves few significant design issues. Table 23.15 compares key design features between the two cell chemistries.

23.7 The lithium-ion battery

Anode: Carbon compound, graphite
Cathode: Cobalt-oxide/lithium-iron-phosphate/manganese-oxide
Electrolyte: $LiPF_6$, liquid lithium salts in an organic solvent (inflammable)
Applications: Laptops, cellular phones, electric vehicles
Typical ratings: Specific energy density: 150 to 200 Wh/kg (540 to 720 kJ/kg)
Volumetric energy density: 250 to 550 Wh/l (1000 to 2000 MJ/m³)
Specific power density: 300 to 1500 W/kg, @ 20 seconds and 300 Wh/l

Lithium is the lightest of all metals, has the greatest electrochemical potential, and provides the largest energy density for a given weight.
As with NiH, the original lithium battery used pure lithium metal, thus both chemistries suffered safety limitations when used as secondary (rechargeable) energy sources. For this reason, a series of cell chemistries have been developed using lithium compounds, which are slightly lower in energy density, instead of pure highly reactive lithium metal. The positive electrode, the cathode, is typically made of Lithium cobalt oxide, $LiCoO_2$. The negative electrode, the anode, is made of carbon. These are called generically lithium-ion batteries. The origin of the cell voltage is then the difference in free energy between Li+ ions intercalated in the crystal structures of the two electrode materials.

i. Cell Materials
The **anode** is a form of carbon mixture, graphite or, formerly, coke. All materials in a cell have a theoretical energy density. With lithium-ion, the anode has been optimized and little improvements can be gained in terms of design changes. The graphite is conductive, dilutes the lithium ion for safety, is reasonably cheap, and does not allow dendrites or other unwanted crystal structures to form.
Anodes consist of a layered crystal graphite into which the lithiated metal oxide such as $LiCoO_2$, $LiNi_{0.3}Co_{0.7}O_2$, $LiNiO_2$, LiV_2O_5, LiV_6O_{13}, $LiMn_4O_9$, $LiMn_2O_4$, $LiNiO_{0.2}CoO_2$ is intercalated.
Lithium titanate spinel oxide (in conjunction with manganese) nanomaterials ($Li_4Ti_5O_{12}$: 1.5V, 160mAh/g, 200Wh/kg) has been introduced in to the anode to eliminate lithium-ion/anode-graphite interaction, thereby providing high-power thermally stable cells with improved cycle life, increased power output, faster charge, wider temperature operating range, but at the expense of halved energy density. The higher potential offers inbuilt overcharge protection. The 'zero-strain' solid insertion material does not form a passivating interfacing layer with the solid electrolyte, resulting in a long cycle life.
During charge, the positive material is oxidized and the negative material is reduced, where lithium ions are de-intercalated from the positive material and intercalated into the negative material. The reverse process is present during a discharge cycle.

Location		**Half-Reaction** (charge)
Anode	Negative terminal	$C + xLi^+ + xe^- \rightarrow Li_xC$
Cathode	Positive terminal	$LiXXO_2 \rightarrow Li_{1-x}XXO_2 + Li^+ + xe^-$
Net		$C + LiXXO_2 \rightarrow Li_{1-x}XXO_2 + Li_xC$

Capacity calculation on a typical anode, using Faraday's constant of 96485 coulombs per mole of electrons:

$$8Li^+ + Co_3O_4 + 8e^- \underset{\text{charge}}{\overset{\text{discharge}}{\rightleftarrows}} 4Li_2O + 3Co^o$$

$$\frac{8e^- \times 96485\text{As}}{1\text{mole}} \times \frac{1\text{hour}}{3600\text{s}} \times \frac{1\text{mole}}{240.8\text{g}}$$

The anode theoretical gravimetric capacity is 0.890 Ah/g.

The **cathode** is a combination of lithium salts and other specific metals and shows promise for further enhancements. Battery research is therefore focussed on cathode variants, where Table 23.16 is a sample of the types of cathode materials used. However, the cathode is generally one of three lithium liberating compounds: a layered oxide, such as cobalt oxide, a polyanion, such as lithium iron phosphate, or a spinel, such as manganese oxide. Three typical electro-active oxide materials are

- Lithium Cobalt-oxide $LiCoO_2$,
- Lithium Manganese-oxide $LiMn_2O_4$, spinel, and
- Lithium Nickel-oxide $LiNiO_2$, all as shown in Table 23.16.

Table 23.16: Cathode chemistry active variants

Cathode Material	Average Voltage	Gravimetric Capacity - Energy Theoretical/practical		structure	Diffusion coefficient	conductivity	O_2 evolution at overcharge	Metal cost
	V	Ah/kg	Wh/kg		cm^2/s	S/cm		pu
$LiCoO_2$	3.7	274/140	520	hexagonal	10^{-7}~10^{-9}	10^{-2}	yes	25
$LiMn_2O_4$	4.0	148/120	580	cubic (spinel)	10^{-9}~10^{-11}	10^{-6}	no	1
$LiCo_{1/3}Ni_{1/3}Mn_{1/3}O_2$	3.6	-/160	580					
$Li(Li_aNi_xMn_yCo_z)O_2$	4.2	-/220	920					
$LiFePO_4$	3.6	-/170	570					
Li_2FePO_4F	3.6	-/115	420					
$LiNiO_2$	3.5	275/150	-	hexagonal	10^{-7}~10^{-9}	10^{-1}	yes	4
CuF_2	3.6	-/520	1875					
Li_2NiSiO_4	4.7	-/320	1550					

Doping with transition metals changes the nature of the active materials and enables the specific capacity to be regulated and internal impedance of the cell to be reduced. Cell voltages in the range 2.1V to 5V are possible.
With an iron phosphate cathode material, the batteries retain the features of conventional lithium-ion batteries, but eliminates the threat of thermal runaway.

Table 23.17: Comparison of cobalt and manganese as positive electrodes

	Cobalt $LiCoO_2$	**Manganese (Spinel)** $LiMn_2O_4$
Energy density	140 Wh/kg	120 Wh/kg
Safety	On overcharge, the cobalt electrode provides extra lithium, which can form into metallic lithium, causing a potential safety risk if not protected by a safety circuit.	On overcharge, the manganese electrode runs out of lithium causing the cell to warm. No safety circuits for smaller cell numbers.
Temperature	Wide temperature range. Best suited for operation at elevated temperature.	Capacity loss above +40°C. Not as durable at higher temperatures.
Aging	Short-term storage possible. Impedance increases with age. Newer versions offer longer storage.	Slightly less than cobalt. Impedance changes little over the life of the cell. Due to continuous improvements, storage time is difficult to predict.
Life Expectancy	300 cycles, 50% capacity at 500 cycles.	Shorter than cobalt.
Cost	Raw material relatively high; protection circuit adds to costs.	Raw material 30% lower than cobalt. Cost advantage due to simplified protection circuit.

Another cell part that has development potential is the **electrolyte**. The electrolyte serves as a reaction medium between the anode and the cathode. It acts as an ion conducting media, blocks electrons, and does not take part in the chemical reaction.

Table 23.18X. Lithium salts as electrolyte solutions

salt	Mole wt	$T_{decomposition}$	Aℓ corrosion	Conductivity σ 1M@25°C PC	Conductivity σ EC/DMC
		°C in solution		mS/cm	mS/cm
$LiBF_4$	93.9	>100	N	3.4	4.9
$LiBF_6$	151.9	>80	N	5.8	10.7
$LiAsF_6$	195.9	>100	N	5.7	11.1
$LiCℓO_4$	106.4	>100	N	5.6	8.4
Li Triphlate	155.9	>100	Y	1.7	
Li Imide	286.9	>100	Y	5.1	9.0

Liquid or gel electrolytes in Li-ion batteries consist of solid lithium-salt electrolytes, such as $LiPF_6$, (although this has a problem with aluminium corrosion), $LiBF_4$, or $LiCℓO_4$, and liquid organic solvents, such as ether. The exact composition of the non-aqueous solvents varies (see Table 23.18X), but propylene carbonate (PC) – dimethyl ether is used for primary cells and ethylene carbonate (EC) with linear organic carbonates such as dimethyl carbonate (DMC), diethyl carbonate, and ethylmethyl carbonate are used in secondary cells. A liquid electrolyte conducts Li ions, which act as a carrier between the cathode and the anode when a battery produces an electric current through an external circuit. However, solid electrolytes and organic solvents are easily decomposed on the highly reactive anodes during charging, thus preventing battery activation. Nevertheless, when appropriate organic solvents are used for electrolytes, they are decomposed and form a solid electrolyte anode passivation interface layer (solid electrolyte interface) at first charge that is electrically insulating and imposes resistance to Li-ion conduction. The interface prevents decomposition of the electrolyte after the second charge but the layer limits the discharge rates and renders the battery unchargeable at low cold temperatures. For example, microporous polyethylene membranes such as ethylene carbonate, are decomposed at a relatively high voltage, (in contrast to only 0.7V for lithium), and forms a dense and stable interface, which separate the electrons from the ions. The electrolyte interface breaks down at about 120°C, at which temperature the highly reactive anode reacts with the electrolyte, producing excessive heat in a thermally runaway process.
Since lithium reacts with water, the electrolyte is a non-aqueous organic lithium salt. Thus, unlike the lead-acid cell that uses water, no hydrogen or oxygen gases are produce by the Lithium-ion cell.
Depending on the choice of material for the anode, cathode, and electrolyte, the voltage, capacity, life, and safety of a lithium-ion battery can vary dramatically. Adding more nickel in lieu of cobalt increases the ampere/hours rating and lowers the manufacturing cost but makes the cell less stable.

ii. Construction
As with most batteries there is an outer case made of a non-reactive metal. The use of metal is particularly important because the battery is pressurized. This metal case has a pressure-sensitive vent hole. If the battery ever gets so hot that it risks exploding from over-pressure, this vent releases the excess pressure. The battery becomes ineffective, so venting is to be avoided. The vent is a safety measure, as is the resettable terminal Positive Temperature Coefficient (PTC) switch, which prevents the battery from overheating.
This cell chemistry and construction permits very thin separators between the electrodes that are made with high surface areas. This enables the cells to handle the high current rates necessary in high power applications.
The metal case holds a long spiral comprising three thin sheets pressed together, not unlike a capacitor:

- A positive electrode
- A negative electrode
- A separator

The sheets are immersed in an organic solvent like ether that acts as the electrolyte. The separator is a thin sheet of microperforated polyethylene plastic, which separates the positive and negative electrodes while allowing ions to pass, but blocks internal electron migration.

Spiral Wound Cylindrical Cell
In order to increase current carrying capacity, it is necessary to increase the active surface area of the electrodes, however the cell case size sets limits on the size of electrodes which can be accommodated. One way of increasing the electrode surface area is to make the electrodes and the separator from long strips of foil and roll them into a spiral cylindrical shape. This provides low internal resistance cells. But

since the electrodes take up more space within the can there is less room for the electrolyte and so the potential energy storage capacity of the cell is reduced. This construction is used extensively for secondary cells. Figure 23.37 shows a Lithium-ion cell but this technology is also used for NiCd, NiMH and even some Lead-acid secondary cells designed for high rate applications.
A spiral wound construction not limited to cylindrical shapes. The electrodes can be wound onto a flat mandrel to provide a flattened shape, which can fit inside a prismatic case. The cases may be made from aluminium or steel. This capacitor type of construction is ideally suited for production automation.

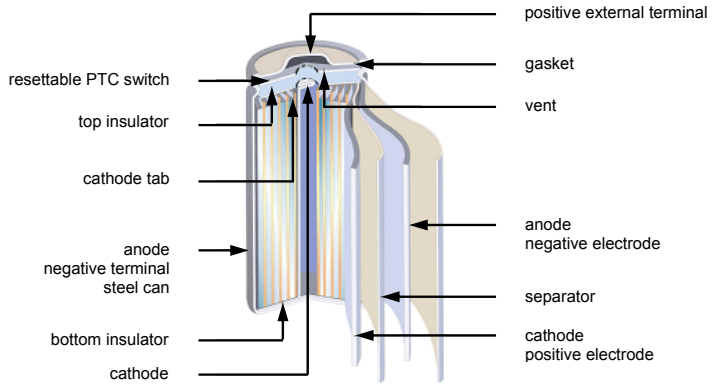


Figure 23.37. *Spiral Cylindrically Wound Electrodes.*

23.7.1 Cathode variants cells

i. The ***Lithium-Cobalt oxide*** Cell - $LiCoO_2$
As shown in Table 23.16, the lithium-cobalt oxide cell voltage is typically 3.7V to 3.9V, and the half-cell reactions are shown in the following table.

Location		**Half-Reaction** (charge)
Anode	Negative terminal	$C_6Li + xLi^+ + xe^- \rightarrow C_6Li_{1+x}$
Cathode	Positive terminal	$LiCoO_2 \rightarrow Li_{1-x}CoO_2 + xLi^+ + xe^-$
Net		$C_6Li + LiCoO_2 \rightarrow Li_{1-x}CoO_2 + C_6Li_{1+x}$

Most lithium-ion batteries for portable applications are cobalt-based. The cell system consists of a cobalt oxide positive electrode (cathode) and a graphite carbon as the negative electrode (anode). Figure 23.38a illustrates the layered crystalline structure of cobalt oxide. An advantage of the cobalt-based battery is its high energy density, typically 140Ah/kg and energy of 500Wh/kg. A long run-time makes this chemistry attractive for cell phones, laptops, and cameras.
The widely used cobalt-based lithium-ion has drawbacks; it offers a relatively low discharge current. A high load would overheat the pack, jeopardizing its safety. The safety circuit of the cobalt-based battery typically limits the charge and discharge rate to about 1C. Another drawback is the internal resistance increase that occurs with cycling and ageing. After 2 to 3 years of use, the battery often becomes unserviceable due to a large voltage drop under load caused by a high internal resistance.
The use of cobalt is unfortunately associated with environmental and toxic hazards.

ii. The ***Lithium-Manganese oxide*** Cell - $LiMn_2O_4$
Lithium manganese oxide, termed spinel, used as a cathode material, produces a 4V cell voltage with an energy density of better than 100Ah/kg and energy of 580Wh/kg. This substance forms a three-dimensional spinel structure that improves the lithium ion flow between the electrodes, as shown in figure 23.38b. High ion flow lowers the internal resistance and increases current capability. The resistance remains low with cycling, however, the cell does age and the overall service life is similar to that of lithium-cobalt. Spinel has an inherently high thermal stability and needs simpler cell safety circuitry than a lithium-cobalt battery. The low internal cell resistance characteristic benefits fast-

charging and high-current discharging. A spinel-based lithium-ion cell has ten times the current capability of the equivalent volume lithium-cobalt cell, with marginal heat build-up. One-second current pulses of twice the specified current are permissible. Some heat build-up cannot be prevented and the cell temperature should not exceed 80°C.

$$C_6Li + LiMn_2O_4 \rightleftarrows Li_{1-x}Mn_2O_4 + C_6Li_{1+x}$$

One significant drawback of the spinel cell is the lower energy capacity, approximately half, compared to the cobalt-based cell. But spinel still provides an energy density that is about 50% higher than that of a nickel-based equivalent.
Manganese, unlike Cobalt, is a safe and more environmentally benign cathode material.

iii. The ***Lithium-nickel-cobalt-manganese*** Cell – $Li(NiCoMn)O_2$
The tri-element cathode incorporates nickel, cobalt and manganese (NCM) in the crystal structure that forms a multi-metal oxide material to which lithium is added. It is not possible to achieve a high energy density and high load capability in the same package; there is a compromise between the two. A NCM cell charges to 4.10V/cell, 100mV less than cobalt and spinel. Charging to 4.20V/cell provides higher capacities but the cycle life is more than halved, from 800 cycles to about 300.

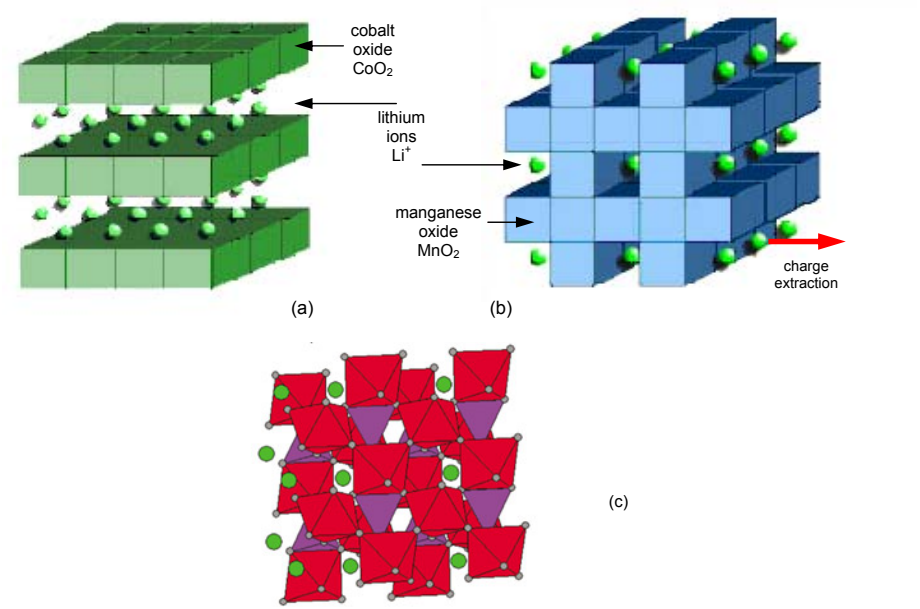


Figure 23.38. *Cathode (positive electrodes) crystalline structures where during discharge, the lithium ions (shown in green) extracted from the cathode and are inserted in the anode: (a) layered cobalt oxide structure with the lithium ions shown bound to the cobalt oxide, (b) three-dimensional framework lithium manganese oxide structure. This spinel structure, which is usually composed of diamond shapes connected into a lattice, appears after initial formation; giving high conductivity but lower energy density. The lithium-ion flow reverses on cell charge, and (c) olivine structure of $LiFePO_4$.*

iv. The ***Lithium-Phosphate*** Cell - $LiFePO_4$
Nano-phosphate materials are added to the cathode (shown in figure 23.38c), resulting in the highest power density in W/kg and energy density of 170Ah/kg, of any lithium-ion battery. The cell can be continuously discharged to 100% depth-of-discharge at 35C and can endure discharge pulses as high as 100C. The phosphate-based system has a nominal voltage of about 3.3V/cell and the peak charge voltage is 3.6V. This is lower than the cobalt-based lithium-ion, hence requires a dedicated charger.
Phosphate based technology possesses superior thermal and chemical stability which provides better safety characteristics than those of lithium-ion cathode technologies. Lithium phosphate cells are incombustible during charge or discharge, more stable under overcharge or short circuit conditions, and

withstand high temperatures without decomposing. When abused, the phosphate based cathode material does not burn and is not prone to thermal runaway. Phosphate chemistry also offers a longer cycle life.

v. The ***Lithium-Polymer*** Cell
The most economical lithium-ion battery in terms of cost-to-energy ratio is cylindrical. Such cells are used for mobile computing and other applications that do not demand an ultra-thin prismatic (rectangular) geometry. If a slim pack is required, the prismatic lithium-polymer cell or alternatively the lithium-ion polymer cell, are the best choices but come at a higher cost in terms of stored energy.
The lithium-polymer differentiates itself from conventional battery systems in the type of electrolyte used, as shown in figure 23.39. The lithium-polymer electrochemistry uses active materials such as $LiCoO_2$, $LiNiO_2$ and its Co doped derivatives, but generally not $LiCoO_2$ chemistry. The dry solid polymer electrolyte resembles a plastic-like film that does not conduct electricity but allows ion exchange (electrically charged atoms or groups of atoms). The polymer electrolyte replaces the traditional porous separator, soaked with electrolyte. The dry solid polymer cell depends on heat to enable sufficient ion flow. This requires that the battery core be kept at an operation temperature above room temperature, 60°C to 100°C.

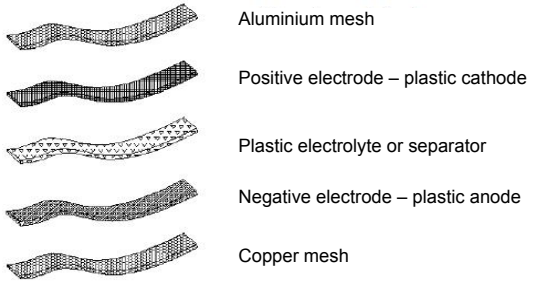


Figure 23.39. *Lithium polymer cell construction layers.*

When lithium polymer cells are first charged, lithium ions are transferred from the layers of the lithium cobaltite to the carbon material that forms the anode, according to:

$$LiCoO_2 + 6C \rightarrow Li_{1-x}CoO_2 + LixC_6$$

Subsequent discharge and charge reactions are based on the motion of lithium ions Li+ between the insertion anode and cathode electrodes.

$$Li_{1-x}CoO_2 + Li_xC \leftrightarrow Li_{1-dx}CoO_2 + Li_{x-dx}C$$

The dry polymer design offers simplifications with respect to fabrication, ruggedness, long life, safety and thin-profile geometry (flexible form factor), but it is expensive.
Unfortunately, the dry lithium-polymer suffers from poor conductivity. To compromise, some gelled electrolyte is added. Cells have a separator/electrolyte membrane with the porous polyethylene or polypropylene separator filled with a polymer, which gels (plasticizes) upon filling with the liquid electrolyte. Thus, the commercial lithium-ion polymer cells are similar in chemistry and materials to their liquid electrolyte counter parts.

With a cell thickness measuring less than a millimetre, virtually no restrictions exist in terms of cell form, shape, and size. Lithium-ion-polymer finds its market niche in lightweight, wafer-thin, flexible form-factor geometries, such as batteries for credit cards and other such applications. It is more stable and resistant to over-charge, but is expensive, has a lower energy density, discharge rate and cycle count than lithium-ion, but longer storage life.

Advantages and Limitations of Li-ion Polymer Batteries

Advantages

- Low profile - batteries that resemble the profile of a credit card are feasible.
- Flexible form factor - manufacturers are not bound by standard cell formats.
- With high volume, any reasonable size can be produced economically.
- Light weight - gelled rather than liquid electrolytes enable simplified packaging, in some cases eliminating the metal shell.
- Improved safety - more resistant to overcharge; less chance for electrolyte leakage.

Limitations

- Lower energy density and decreased cycle count compared to Li-ion - potential for improvements exist.
- Expensive to manufacture - once mass-produced, the Li-ion polymer has the potential for lower cost. Reduced control circuit offsets higher manufacturing costs.

Table 23.18 highlights a number of other alternative lithium-ion cell cathode types.

23.7.2 General Lithium-ion Cell characteristics

Lithium-ion load characteristics are similarly to nickel-cadmium in terms of discharge. The high cell voltage of 3.6V allows battery pack designs with only one cell. Most mobile phones use a single cell.
Lithium-ion is a low maintenance battery, an advantage that most other chemistries cannot claim. There is no memory and no scheduled cycling is required to prolong battery life. In addition, the self-discharge is less than half that of nickel-cadmium.
Lithium-ion has drawbacks, specifically it is fragile and requires a protection circuit to maintain safe operation, as considered in section 23.7.4.

Table 23.18: Lithium-ion chemistries

Chemistry	Cathode	Electrolyte	Nominal voltage	Open-circuit voltage	Wh/kg	Wh/dm³
$Li\text{-}MnO_2$ (Li-Mn)	Heat-treated manganese dioxide	Lithium perchlorate in propylene carbonate and dimethoxyethane, M Wt 106.4	3V	3.7V	280	580
	About 80% of the lithium battery market. Uses inexpensive materials. Suitable for low-drain, long-life, low-cost applications. High energy density per mass and volume, 650Wh/l. Can deliver high pulse currents. Wide temperature range -40°C to 60°C, max. With discharge the internal impedance rises and the terminal voltage decreases.					
$Li\text{-}SOC\ell_2$	Thionyl chloride	Lithium aluminium chloride in thionyl chloride	3.5V	3.65V	290	670
	Liquid cathode. Can operate down to -55°C, where it retains over 50% of its rated capacity. Negligible gas generated in nominal use, limited amount under abuse. Relatively high internal impedance and limited short-circuit current. High energy density, about 500Wh/kg. Toxic. Electrolyte reacts with water. After storage can form anode passivation layer, which leads to temporary voltage delay when first put into service. High cost and safety concerns. Can explode when shorted. Hazardous waste.					
$Li\text{-}SOC\ell_2$, $BrC\ell$, Li-BCX	Thionyl chloride with bromine chloride	Lithium aluminium chloride in thionyl chloride	3.75V	3.9V	350	770
	Liquid cathode. A variant of the thionyl chloride battery, with 300 mV higher voltage, which quickly drops back to 3.5V, as the bromine chloride is consumed during the first 10-20% of discharge. Added bromine chloride improve safety when abused. Wide temperature range -55°C to 80°C. 1420Wh/l.					
$Li\text{-}SO_2C\ell_2$	Sulphuryl chloride		3.7V	3.95V	330	720-1400
	Liquid cathode. Similar to thionyl chloride. Discharge does not result in build-up of elementa sulphur, which is involved in some hazardous reactions, therefore is safer. The electrolyte tends to corrode the lithium anodes, reducing the shelf life. Chlorine is added to some cells t make them more resistant to abuse. Sulphuryl chloride cells give less maximum current tha thionyl chloride, due to polarization of the carbon cathode. Sulphuryl chloride reacts violentl with water, releasing hydrogen chloride and sulphuric acid.					
$Li\text{-}SO_2$	Sulphur dioxide on teflon-bonded carbon	Lithium bromide in sulphur dioxide with small amount of acetonitrile	2.85V	3.0V	250	400
	Liquid cathode. Can operate down to -55°C and up to +70°C. Contains liquid SO_2 at high pressure. Requires safety vent, can explode in some conditions. High energy density 440Wh/l. High cost. At low temperatures and high currents performs better than $Li\text{-}MnO_2$. Toxic. Acetonitrile forms lithium cyanide, and can form hydrogen cyanide in high temperatures. Addition of bromine monochloride can boost the voltage to 3.9V and increase energy density.					
$Li\text{-}(CF)_x$	Carbon monofluoride	Lithium tetrafluoroborate in propylene carbonate, dimethoxyethane, and/or gamma-butyrolactone	2.8V	3.1V	360	680
	Cathode material formed by high-temperature intercalation of fluorine gas into graphite powder. High energy density (250Wh/kg), 7 year shelf life. Used for low to moderate current applications, e.g. memory and clock backup batteries. Maximum temperature 85°C. Low self-discharge (<0.5%/year at 60 °C, <1%/yr at 85°C).					
$Li\text{-}Ag_2CrO_4$	Silver chromate	Lithium perchlorate solution	3.1/2.6 V	3.45 V		
	High reliability. Has a 2.6V plateau after reaching certain percentage of discharge, provides early warning of impending discharge.					
$Li\text{-}FeS_2$	Iron disulphide	Propylene carbonate, dioxolane, dimethoxyethane	1.5V	1.8V	297	320
	'Lithium-iron', 'Li/Fe'. Called 'voltage-compatible' lithiums. 2.5 times higher lifetime for high current discharge regime than alkaline batteries, no advantage for low-current applications. Low self-discharge, 10 years storage time. FeS_2 is cheap. Some types rechargeable. Cathode often designed as a paste of iron sulphide powder mixed with powdered graphite. 370Wh/l. Variant is $Li\text{-}CuFeS_2$. $2Li+FeS_2 \rightarrow Li_2FeS_2$ then $2Li+Li_2FeS_2 \rightarrow Fe+2Li_2S$					
$Li\text{-}V_2O_5$	Vanadium pentoxide		3.3/2.4V	3.4V	120/260	300/660
	Two discharge plateaus. Low-pressure. Rechargeable. Used in reserve batteries. LiAl anode					

i. Discharge: Lithium-ion should function within the discharge temperature limits of -20°C to 60°C. Cell performance is temperature based, meaning that the rate capability at or below -20°C is reduced due to increased electrolyte impedance. Discharging at low temperatures does not harm the battery. Lithium-ion may be used down to -30°C with acceptable results.
A high discharge rate combined with elevated temperatures above 60°C can cause self-heating, an effect that could permanently damage the separator and electrodes of the cells.
Cells perform better on a pulse rather than dc load. The dc resistance of a cylindrical cell is approximately 110mΩ. At 1 kHz ac, the impedance reduces to about 36mΩ. As the pulse frequency increases, the cell's effective impedance reduces. This results in better performance and lower heat build-up, consequently increasing lithium-ion cell life.
The internal resistance of cobalt-based lithium-ion increases with age. A manganese-based cell maintains its resistance at a low level throughout its service life. Although the cobalt-based lithium-ion cell has a higher energy density, manganese is better suited for pulse load applications.

ii. Li ion cell Recharging: A moderate charge and discharge puts less stress on the battery, resulting in a longer cycle life, as shown in figure 23.40.

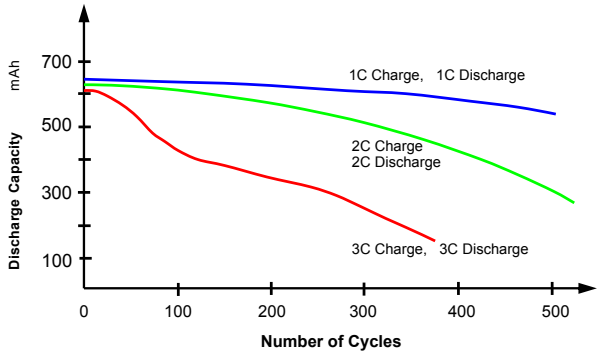


Figure 23.40. *Cycle/Lifetime of a lithium-ion cell as a function of charge and discharge rates.*

Li-ion cell charging is voltage-limited similar to lead acid battery charging. The difference lies in a higher voltage per cell, tighter voltage tolerance, and the absence of trickle or float charge when full charge is reached.
While the lead acid battery offers some flexibility in terms of voltage cut-off, Li-ion cells are strict on the correct charging cut-off voltage. Although higher cut-off voltages deliver increased energy densities, cell oxidation severely limits service life. Aided by chemical additives, Li-ion cells are charged to 4.20V, with a tolerance of ±50mV/cell. For maximum cycle life, an end-of-charge voltage threshold of about 3.90V/cell is used.
The charge time of all Li-ion batteries, when charged at a 1C initial current, is about 3 hours with about a 70% charge level being reached, during which time the battery remains cool. This charge state is often referred to as fast charging. Full charge is attained after the voltage has reached the upper voltage threshold and the current has dropped and levelled off to about 3 percent of the nominal charge current. Although the voltage peak is reached quicker with higher current, the topping charge takes twice as long as the initial charge. Figure 23.41 shows the voltage and current charging profile as the Li-ion cell passes through stages one and two.
No trickle charge is applied because Li-ion is unable to absorb overcharge. Trickle charge causes plating of metallic lithium, rendering the cell unstable. Instead, a brief topping charge is applied to compensate for the small amount of self-discharge the battery and its protective circuit consume.
A topping charge may be implemented once every 500 hours or 20 days. Typically, when the open terminal voltage drops to 4.05V/cell and ceases when 4.20V/cell is reached again.

On a charge voltage above 4.30V, Li-ion cells become increasingly unstable with lithium metal plated on the anode, and, the cathode material becomes an oxidizing agent, loses stability, generates heat, and releases oxygen.
Li-ion battery packs contain a protection circuit that prevents the cell voltage from going too high while charging. The typical safety threshold is set to 4.30V/cell. In addition, temperature sensing disconnects the charge source if the internal temperature approaches 90°C. Most cells feature a mechanical pressure switch that permanently interrupts the current path if a safe pressure threshold is exceeded. Internal voltage control circuits cut off the battery at low and high voltage levels.

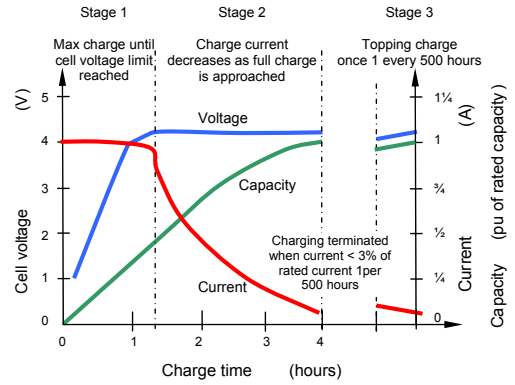


Figure 23.41. *Charge stages of a Li-ion battery.*
Increasing the charge current does not significantly shorten the charge time. Although the voltage peak is reached quicker with higher current, the topping charge takes longer.

Exceptions are made on spinel (manganese) batteries. On overcharge, this chemistry produces minimal lithium plating on the anode because most metallic lithium has been removed from the cathode during normal charging. The cathode material remains stable and does not generate oxygen unless the cell gets extremely hot.

iii. Lithium Polymer cell charging
The charge process of a Li-Polymer cell is similar to that of the Li-ion cell. Li-Polymer uses dry electrolyte and takes 3 to 5 hours to charge. Li-ion polymer with gelled electrolyte, on the other hand, is almost identical to that of Li-ion, such that the same charge algorithm can be applied. With most chargers, the user does not need to know whether the battery being charged is Li-ion or Li-ion polymer.
Most 'Polymer' batteries are a variety of the Li-ion polymer using some sort of gelled electrolyte. Low-cost dry polymer batteries for ambient temperature operation are evolving.

The charge cycle called a CC/CV two-stage charge cycle is necessary to fully charge the battery.

- The first stage of the charge cycle is a Constant Current charge until the battery voltage reaches 4.1 to 4.2V.
- Upon reaching this peak voltage, a Constant Voltage charge is initiated until the charge current reduces to 3% of the rated current. Upon charge completion, a top off charge may be used to insure to counteract the self-discharge of the battery and protective circuit. This top-off charge may be initiated when the open circuit voltage of the battery reaches less than 4.05 volts and terminates upon reaching the full charge voltage of 4.1 to 4.2V. Depending on the battery, this top off charge may be repeated once every 20 days.

iv. Battery Float Voltage
The main determining factor of a battery's float voltage is the electrochemical potential of the active materials used in the battery's cathode, which for lithium is approximately 4V. The addition of other compounds can change this voltage. A second factor is a trade off between cell capacity, cycle life, battery life and safety. The curve shown in Figure 23.42a shows the relationship between cell capacity and cycle life.
The best compromise between capacity and cycle life of a standard Li-ion cell is a 4.2V float voltage. Using 4.2V as the constant voltage limit (float voltage), a battery can typically deliver about 500 charge/discharge cycles before the battery capacity drops to 80%. A lower float voltage for Li-ion phosphate batteries results in a much higher number of charge/discharge cycles. One charge cycle consists of a full charge to a full discharge. Multiple shallow discharges add up to one full charge cycle.
Although charging to a capacity less than 100% using either a reduced float voltage or minimum charge current termination results in an initial reduced battery capacity, as the number of cycles increases beyond 500, the battery capacity with the lower float voltage can exceed that with a higher float voltage. Figure 23.42b illustrates this; how the recommended float voltage compares with a reduced float voltage in regard to capacity and the number of charge cycles.
Because of the different Li-ion battery chemistries (as in figure 23.42c for example) and other conditions that can affect battery life, the curves shown only estimate of the number of charge cycles and battery capacity levels, due to differences in battery materials and construction methods.

Battery manufacturers specify a charge method and a float voltage to meet the battery specifications for capacity, cycle life and safety. Many batteries include a battery pack protection circuit, which temporarily opens the battery connection if the maximum battery voltage is exceeded. Once opened, connecting the battery pack to the charger normally resets the pack protection. The printed battery voltage, such as 3.6V for a single cell battery, is not the float voltage, but rather the average battery voltage when discharging.

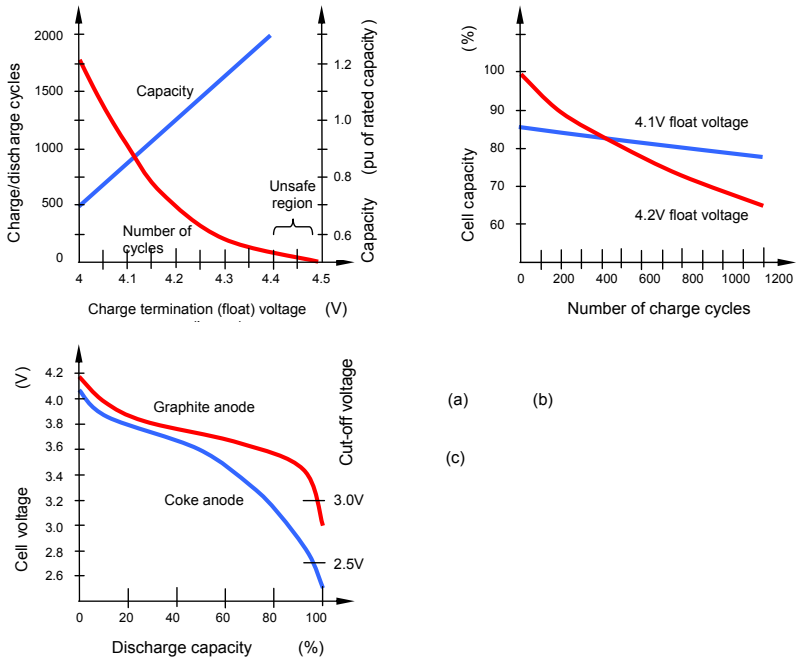


Figure 23.42. *Charge float voltage characteristics.*

v. Aging: Some capacity deterioration is noticeable after one year, whether the Lithium-ion battery is in use or not. Over two to three years, the battery frequently fails. Other chemistries also have age-related degenerative effects. This is especially true for the NiMH if exposed to high ambient temperatures.
Storage in a cool place slows the ageing process of all chemistries. For lithium-ion, manufacturers recommend storage temperatures of 15°C and a 40% partial charge.
The permanent capacity loss is not recoverable by charging. This loss is linked to battery life because when the permanent capacity loss drops to approximately 80%, the battery is considered at the end of its life. Permanent capacity loss is mainly due to the number of full charge/discharge cycles, the battery voltage and battery temperature. The longer the battery remains near 4.2V or 100% charge level (lower voltage for Li-ion Phosphate) the faster the capacity loss occurs. This is true whether the battery is being charged or floating in a fully charged condition with the voltage near 4.2V. Maintaining a Li-ion battery in a fully charged condition shortens its lifetime. The chemical changes that shorten the battery lifetime, are accelerated by high float voltage and high temperature. Permanent capacity loss is unavoidable, but may be minimised by using the following techniques.

- Use partial discharge cycles. Using only 20% or 30% of the battery capacity before recharging extends cycle life considerably. Generally, 5 to 10 shallow discharge cycles are equal to one full discharge cycle. But keeping the battery in a fully charged state also has an effect on shortening battery life. Full discharge cycles should be avoided.
- Avoid charging to 100% capacity, by using a lower float voltage. Reducing the float voltage increases cycle life and service life at the expense of reduced battery capacity. A 100mV to 300mV drop in float voltage can increase cycle life more than five times. Li-ion cobalt chemistries are more sensitive to a higher float voltage than other chemistries. Li-ion phosphate cells have a lower float voltage than the more common Li-ion batteries.

- Select the correct charge termination method. Selecting a charger that uses minimum charge current termination ($\frac{1}{10}$C) can also extend battery life by not charging to 100% capacity. For example, ending a charge cycle when the current drops to $\frac{1}{5}$C is similar to reducing the float voltage to 4.1V. In both instances, the battery is charged to approximately 85% of capacity, which can significantly increase overall battery life.
- Limit battery temperature. High temperatures accelerate chemical changes within the battery, which shorten battery life, while charging below 0°C promotes metal plating at the battery anode, which can develop into an internal short, producing heat and making the battery unstable and unsafe. Many battery chargers have provisions for measuring battery temperature to assure charging does not occur at temperature extremes.
- Avoid high charge and discharge currents as they reduce cycle life. High currents place excessive stress on the battery. Some chemistries are more suited for higher currents such as Li-ion manganese and Li-ion phosphate.
- Avoid deep discharges below 2V or 2.5V, which quickly and permanently damage a Li-ion battery. Internal metal plating occurs causing a short circuit making the battery unusable and unsafe. Most Li-ion batteries have electronic circuitry within the battery pack that opens the battery connection if the battery voltage is less than 2.5V, exceeds 4.3V or if the battery current when charging or discharging exceeds a predefined threshold.

vi. Disposal: Lithium-ion cells cause little harm when disposed. Heat treatment, pyrolysis, is used to recover the primary metals.

23.7.3 General Lithium-ion Cell properties

General battery properties for the lithium-ion types are summarised in Table 23.19 and as follow.

Advantages

- High, ever increasing energy density and high cell voltages, 3.6V.
- Does not need prolonged priming when new. Only one regular charge is needed.
- Relatively low self-discharge, that is less than half that of nickel-based batteries.
- Low maintenance since no periodic discharge is needed
- There is no memory, so no need for fully discharging before charging.
- Specialty cells can provide high current, 40C rates.
- High number of charge/discharge cycles, with fast charging.
- Light since a light metal is used.
- Float or trickle charging not necessary.

Limitations

- Needs protection circuit to maintain voltage, current, and temperature within safe limits.
- Higher internal impedance than other competing cell technologies.
- Subject to ageing, even in storage, 15°C storage at 40% charge minimises ageing.
- Transportation restrictions, subject to regulatory control.
- Expensive to manufacture being about 40% higher in cost than nickel-cadmium.
- Metals and chemicals are continuing changing.
- Rapid degradation with temperature increase.
- Cannot be recharged at low temperatures (below 0 to -10°C)

Table 23.19: Lithium-ion battery characteristics

Cathode material	advantages	disadvantages
Lithium cobalt oxide	High capacity	Lower charge and discharge rates Higher cost
Lithium manganese oxide	Lower ESR Higher charge and discharge rates Higher temperature operation Inherently safer	Lower capacity Lower life cycle Shorter lifetime
Lithium phosphate	Very low ESR Very high charge and discharge rates High temperature operation Inherently safer	Lower discharge voltage Lower float voltage Lower capacity

23.7.4 Cell protection circuits

Built into each battery pack, the protection circuit limits the peak voltage of each cell during charge and prevents the cell voltage from dropping too low on discharge. In addition, the cell temperature is monitored to prevent operation at temperature extremes. The maximum charge and discharge current is limited to between 1C and 2C. With these precautions in place, the possibility of metallic lithium plating occurring due to overcharge is virtually eliminated.
As with fuel cell and double layer capacitor systems, the Li-ion battery cell safety circuit consists of four main sections:

i. Over-charge protection: The controller that monitors each cell (or paralleled cells) voltage and prevents cells overcharge accordingly controlling the cut-off MOSFET switches. Also the voltage across the switches is monitored in order to prevent charge over-current. Stops charging when the cell voltage exceeds the specified maximum in order to prevent the battery from overheating or exploding due to overcharging.

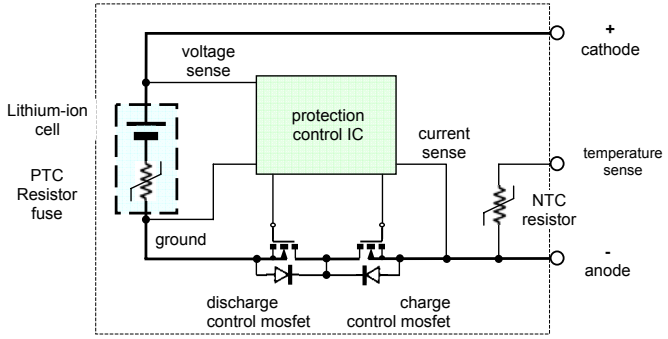


Figure 23.43. *Protection circuitry for series connected double layer cells and Li-ion cells.*

ii. Over-discharge protection: The control switches that are usually MOSFET structures that cut-off the discharge depending on the control signals from the controller. Also the switch voltages are monitored in order to prevent discharge over-current. Promptly stops discharging when a large current flows due to external shorting of the battery pack. Stops discharge when the voltage falls below the specified minimum in order to prevent battery degradation due to over-discharging.

iii. Internal controller temperature protection: The temperature fuse to cut-off the current if the control MOSFET switches experience abnormal heating. This fuse is not resettable.

iv. Over-temperature protection: The negative temperature co-efficient thermistor (usually resettable NTC) that measures the battery temperature inside the pack. Its terminals are connected to the charger so it can sense the temperature of the pack and control the charge current until the battery its full charged.

23.8 Battery Thermodynamics

From chapter 21.14, a negative value for the Gibbs free energy indicates that a reaction is spontaneous in the forward direction. Conversely, a positive value indicates that a reaction is spontaneous in the reverse direction. Similarly, the standard electrochemical potential indicates whether the reaction goes forward ($E° < 0$) or in the reverse direction ($E° > 0$) under standard conditions. The connection between $\Delta G°$ and $E°$ is given by the relationship (equation 21.14)

$$\Delta G° = -nFE° \qquad (23.9)$$

where n is the number of moles of electrons transferred in the reaction and F is the Faraday constant, which corresponds to the electric charge on one mole of electrons: 96,487 C/mol.

Electrons are involved in an electrochemical reaction. Moving one Coulomb of charge, 1/96,487 moles of electrons, between two electrodes that differ by one volt requires one Joule of energy. Conversely, spontaneous movement of 1/96,487 moles of electrons between two electrodes with a potential difference of one volt releases one joule of energy to do useful work.

i. Equilibrium and the Dead Battery

Electrochemical reaction ceases when one of the reactant is depleted, usually because the anode is consumed such that electrons are no longer liberated. Another frequently encountered example is the magnesium stake used to protect a utility pole. When the stake is consumed, it no longer protects the pole from corrosion. Hence the magnesium stakes need to be regularly replaced.

Another reason for a dead battery is that the reaction reaches equilibrium, which means there is no net change and electrons no longer flow. Electron transfer occurs as part of the tendency toward equilibrium. At equilibrium, the forward and reverse reactions occur at the same rate. One consequence of the equal rates is that the number of electrons transferred from reactants to products is equal to the number transferred from products to reactants. The net flow of electrons, like the net flow of atoms, is therefore zero. With no net electron flow, the current is zero and so is the voltage. Often the voltage produced by a battery differs from the standard value, and the most common reason is that the solution-phase ions are not in their standard state: the ideal 1 *M* concentration.

The relationship between the standard voltage, $E°$, and the voltage, E, at temperature T, is determined by the relationship between the standard Gibbs free energy $G°$ and the Gibbs free energy:

$$\Delta G = \Delta G° + RT\,\ell nN$$

where R is the gas constant, 8.314 J/K mol, and N is the ratio of the effective concentrations of the products and reactants in the reaction.
Substituting equation (23.9), this equation becomes

$$-nFE = -nFE° + RT\,\ell nN$$

or

$$E = E° - \frac{RT}{nF}\ell nN \qquad (23.10)$$

Equation (23.10) is referred to as the Nernst equation. At room temperature, the coefficient RT/F = 0.025680 V. Multiplication by 2.303 converts equation (21.27) into

$$E = E° - \frac{0.0592\text{V}}{n}\ell nN$$

At equilibrium, E is zero and

$$E° = \frac{RT}{nF}\ell nK$$

These electrochemical equation can be applied to the lead acid battery as follows.
The Nernst equation for the lead-acid cell can be written by adding the two half-cell reactions given in Table 23.4, where the overall reaction is:

$$PbO_{2(s)} + Pb_{(s)} + 2H_2SO_{4\,(aq)} \rightarrow 2PbSO_{4(s)} + 2H_2O \qquad (23.11)$$

The affect of sulphuric acid concentration on the electrode potential, is clearly seen in equation (23.11). Using equation (23.10), the Nernst equation for the lead acid cell is

$$E = E_o - 2.303\frac{RT}{nF}\times\log\frac{[\,a^2_{PbSO_4}\times a^2_{H_2O}\,]}{[a_{PbO_2}\times a_{Pb}\times a^2_{H_2SO_4}]}$$

where a is the activities of the reactants and the products of the cell, defined as an effective concentration. It is related to the actual concentration of the species, via, $a = \gamma C$. C, the molar concentration, is an activity coefficient. $a < C$, except in very dilute solutions where $\gamma < 1$ and $a > C$.
R = 8.314 J / K.mol, is the gas and T is the absolute temperature (K)

Since the activity of a pure solid and water is 1, specifically $a_{PbSO_4} = 1$, $a_{H_2O} = 1$, $a_{PbO_2} = 1$, and $a_{Pb} = 1$.

$$\begin{aligned} E &= E^o - 2.303\times\frac{RT}{nF}\times\log\frac{1}{a^2_{H_2SO_4}} \\ &= E^o + 2\times 2.303\times\frac{RT}{nF}\times\log a_{H_2SO_4} \end{aligned} \qquad (23.12)$$

where $n = 2$ since n is the number of moles of electrons involved in the oxidation-reduction reactions in equations, in Table 23.4.

Equation (23.12) shows the effect of temperature and the activity (effective concentration) of H_2SO_4, on the cell potential.

The activity of sulphuric acid is related to the actual electrolyte concentration as given by:

- $a_{H_2SO_4} = 4 \times m^3 \times C^3$ where C is the MOLAL concentration or molarity and m is defined as a mean activity coefficient.
- Typically, $a < C$, except in very dilute solutions (< 0.001 M) when $\gamma_m \approx 1$ and $a < C$. The activity approaches unity for acid concentration in the range 3.5 to 4.0 M.

The activity coefficient is generally temperature and concentration dependent, and is experimentally determined.

Example 23.5: ***Electrochemistry – battery thermodynamics***

Standard cell potentials can be used to determine the free energy change for a chemical reaction. Consider the copper–zinc galvanic cell:

$$Zn \mid Zn^{2+} \mid\mid Cu^{2-} \mid Cu$$

i. Determine the potential of this cell and the free energy for the reaction.
ii. At room temperature, what is the relationship between the zinc and copper ion concentrations at equilibrium?
iii. If the cell produces 0.67 V, determine the ratio $[Zn^{2+}]/[Cu^{2+}]$.

Solution

$$Zn(s) + Cu^{2+}(aq) \rightleftharpoons Zn^{2+}(aq) + Cu(s)$$

Since Cu and Zn are both solids, they do not appear in the equilibrium constant, N.

i. The half-reactions are

$$Zn(s) \rightarrow Zn^{2+}(aq) + 2e^- \quad E^o = 0.7628\text{V}$$
$$Cu^{2+}(aq) + 2e^- \rightarrow Cu(s) \quad E^o = 0.3402\text{V}$$

Cell potential $E = 0.7628\text{ V} + 0.3402\text{ V} = 1.103$ V.
Since two electrons are transferred from copper to zinc, $n = 2$, $F = 96{,}485$ C/mol, and $E° = 1.103$ V.

$$\Delta G° = -nFE°$$
$$\Delta G° = -2 \text{ mol } e^- \times 96485 \text{ C/mol } e^- \times 1.103 \text{ V} \times 1 \text{ J/C.V}$$
$$= -212.8 \text{ kJ/mol}$$

Since the Gibbs free energy value is negative, the reaction is spontaneous for reactants and products in their standard states. For the solids, Zn and Cu, the solid is the standard state. For the ions in solution, Cu^{2+} and Zn^{2+}, the standard state is a 1 M solution. In contact with 1 M solutions, then, the reduction of copper by zinc is spontaneous.

ii. In

$$\Delta G° = RT\,\ell nK$$

The equilibrium constant K is the ratio of the two ions, that is $K = [Zn^{2+}]/[Cu^{2+}]$. Thus

$$K = e^{\frac{-\Delta G^o}{RT}} = e^{\frac{-(-212.8 kJ/mol)}{8.3145 J/k.mol \times 298 K} \times \frac{1000 J}{kJ}}$$
$$= 1.99 \times 10^{37}$$

The zinc ion concentration is about 37 orders of magnitude larger than the copper ion concentration at equilibrium. If the zinc ion is present in lower concentration, $N < K$, the reaction proceeds as written, generating a cell potential.

iii. Rearranging equation (21.26), $E° = 1.103$ V and $E = 0.67$ V gives

$$N = e^{-\left(E-E^o\right)\frac{nF}{RT}}$$
$$= e^{-(0.67-1.103)\frac{2 \times 96485 C/mol}{8.3145 J/mol.K \times 298 K}} = 4.42 \times 10^{14}$$

As more Zn^{2+} ion is produced, the cell potential decreases even more, ultimately falling to zero. At that point, the Zn^{2+} ion concentration has increased to 1.99 ×10^{37} times the Cu^{2+} ion concentration.

♣

Nuclear batteries function by converting the heat produced by a nuclear source, to create a current using the Seebeck-effect. A second type of nuclear cell uses beta-radiation impinging on a semiconductor junction to create an electron-hole pair that migrate to the electrodes of the junction creating a current, much in the same way that a solar cell creates energy from light. Currently these primary batteries are low power and are used to trickle charge existing batteries to give a longer lifetime or to power remote transducers.

23.9 Summary of key primary and secondary cell technologies

Table 23.20: Common commercial cells chemistries, batteries

cell name	nominal voltage	Capacity Wh/kg	anode	cathode	electrolyte
Primary cell	V				
carbon-zinc (Leclanche)	1.26	65	zinc foil	MnO_2	Aq Zn-Cℓ_2-$NH_4C\ell$
zinc chloride (carbon-zinc)	1.5		zinc foil	Electrolytic MnO_2	Aq $ZnC\ell_2$
alkaline	1.5	95	zinc powder	Electrolytic MnO_2 powder	aq KOH
zinc-air	1.65	290	zinc powder	Carbon (air, O_2)	aq KOH
silver-zinc	1.6	130	zinc powder	Ag_2O	aq KOH
lithium-manganese dioxide	3.0	200	lithium foil	MnO_2	$LiCF_3SO_3$ or $LiC\ell O_4$
lithium-carbon monofluoride	3.0		lithium foil	CF_x	$LiCF_3SO_3$ or $LiC\ell O_4$
lithium-iron sulphide	1.6		lithium foil	FeS_2	$LiCF_3SO_3$ and/or $LiC\ell O_4$
Secondary - rechargeable	V				
lead acid	2.0		lead	PbO_2	Aq H_2SO_4
nickel-cadmium	1.2		cadmium	NiOOH	aq KOH
nickel-metal hydride	1.2		MH	NiOOH	aq KOH
lithium ion	4.0		Lithium (C)	$LiCoO_2$	$LiPF_6$ non-aq solvent
nickel-hydrogen	1.2		H_2 (Pt)	NiOOH	aq KOH
lithium-iodine	2.7		lithium	I_2	LiI
lithium-silver-vanadium oxide	3.2		lithium	$Ag_2V_4O_{11}$	LiAsF
lithium-sulphur dioxide	2.8		lithium	SO_2 (C)	SO_2-LiBr
lithium-thionyl chloride	3.6		lithium	$SOC\ell_2$ (C)	$SOC\ell_2$-$LiA\ell C\ell_4$
lithium-iron sulphide	1.6		lithium	FeS_2	LiCℓ-LiBr-LiF
magnesium-silver chloride	1.6		magnesium	AgCℓ	Sea water

Table 23.21: Permissible temperature limits for various batteries
Older battery technologies are more tolerant to charging at extreme temperatures than newer, more advanced chemistries.

	Slow Charge ($\frac{1}{10}$C)	Fast Charge (½C to 1C)
Nickel Cadmium	0°C to 45°C	5°C to 45°C
Nickel-Metal Hydride	0°C to 45°C	10C° to 45°C
Lead Acid	0°C to 45°C	5C° to 45°C
Lithium Ion	0°C to 45°C	5C° to 45°C

Table 23.22: Charging of different chemistries

	Lead-acid (Sealed or flooded)	Nickel-based (NiCd and NiMH)	Lithium-ion (Li-ion)
Deplete all battery energy before charging?	It is better to recharge more often; avoid frequent full discharges. Deep cycles deteriorate the battery. Use a larger battery if full cycles are required.	Fully discharge once every 1-3 months to prevent memory. It is not necessary to deplete the battery before each charge. Over cycling wears down NiMH.	It is better to recharge more often; avoid frequent full discharges. For batteries with a fuel gauge, allow a full discharge once a month to enable reset.
Partially or fully charge battery?	Does not matter. Charging in stages is acceptable. Full charge termination occurs by reading the voltage level and charge current. Charging a full battery is safe and does not cause harm.	Allow full charge without interruptions. Repeated partial charge can cause heat buildup. (Many chargers terminate charge by heat. A fully charged battery will re-heat, causing overcharge.)	Does not matter. Charging in stages is acceptable. Full charge termination occurs by reading the voltage level and charge current. Charging a full battery is safe and does not cause harm.
Battery charged state when not in use?	Always keep battery fully charged. A discharged battery causes sulphation (insulating layer in the cell). This condition is often irreversible.	Not critical. A 40% charge for long storage. (Open terminal voltage cannot determine state-of-charge.) Store in a cool place. Battery can be fully depleted and recharged. Priming may be needed.	Best to store at 40% charge or 3.75-3.80V/cell open terminal. Cool storage is more important than state-of-charge. Do not fully deplete battery because Li-ion may turn off its protection circuit.
Battery heat up during charge?	The battery should remain cool or lukewarm to the touch. The battery must remain cold on maintenance charge.	Heating towards full charge. The battery must cool down before use. Discontinue using a charger that keeps the battery warm on standby.	Little heating is generated during charge. Do not allow the battery to heat during charge.
Allowable charging temperatures?	Rechargeable batteries can be used under a wide temperature range, but charging is usually not permitted at the extremes. The maximum allowable charge temperatures are shown below:		
Slow charge (0.1) Fast charge (½-1C)	0ºC - 45ºC 5ºC- 45ºC Warm temperature lowers the battery voltage. Serious overcharge occurs if the cut-off voltage is not reached.	0ºC - 45ºC 5ºC - 45ºC Charging a hot battery decreases the charge time. The battery may not fully charge.	0ºC - 45ºC, 5ºC- 45ºC Temperature sensor may prevent charge or cut off the charge prematurely.
Charger requirements?	Multi-level charges shorten charge time. Charge must be fully saturated. Failing to do so will gradually decrease the capacity. Fastest full-charge time: 8 to 14 hours.	Best results are achieved with a fast-charger that terminates the charge by other than temperature alone. Fastest full-charge time: Slightly over 1 hour.	Charger should apply full charge. Fastest full-charge time: 2 to 3 hours.

23.10 The Electrochemical Double Layer Capacitor (EDLC) - supercapacitor or ultracapacitor

Unlike batteries, which store energy chemically, capacitors store energy as an electrostatic field. Typically, a battery stores energy and little power; a capacitor provides large instantaneous power, but relatively low energy. A capacitor is made of two conducting plates and a separating electrical insulator called the dielectric, which conducts ionically, but not electrically. In a capacitor

$$W = qE = \tfrac{1}{2}CE^2 \quad (\text{J}) \qquad (23.13)$$

where the capacitance, C, is directly proportional to the overlapping surface area A of the plates and inversely proportional to the distance between them, d.

$$C = \varepsilon_o \varepsilon_r \frac{A}{d} \quad \left(= \frac{Q}{E}\right) \quad (\text{F}) \qquad (23.14)$$

As the plate surface area increases and the separation between the plates decreases, the capacitance increases and storable energy increases. Aluminium electrolytic capacitors attain capacitance volumetric densities of about 50F/m^3. But for useful energy storage as in a battery, several orders more capacitance per cubic metre is needed, as seen in figure 23.2. The double layer capacitor stores energy electrostatically by polarising the electrolytic solution and attains these necessary capacitance density levels by increasing the surface area to about 2,000m^2/g, in a small volume. This is equivalent to 100F/g of electrode active mass.

The supercapacitor increases the electrode area just like a polymer foam cleaning sponge with millions of crenulations. But in the supercapacitor this is achieved with carbon – activated charcoal particles (or sintered metal powders), which form a non-reactive, conductive, random, rough, porous, carbon fibre surface with an extraordinarily large surface area in a small volume, with a small charge separation distance, of less than 10 Angstroms.

The second non-reactive porous electrode is activated when the molecule size voids are impregnated with a conductive liquid (for example, an aqueous acid or salt solution). Thus the supercapacitor can be viewed as two non-reactive porous plates suspended within an electrolyte.

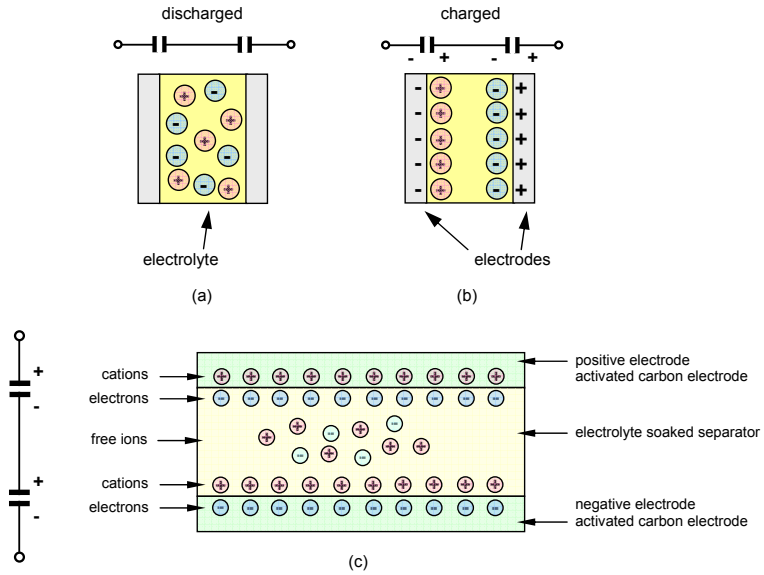


Figure 23.44. *Supercapacitor (EDLC) showing formation of the double charge boundary layers: (a) discharged state; (b) fully charged state; and (c) partially discharged.*

The dielectric is based on an electrical double layer, each layer conductive, but the interface is non-conductive. The electrodes have a physical barrier made from activated carbon that, when an electrical charge is applied to the material, a double electric field is created which acts like a dielectric, as shown in figure 23.44. This ultra-thin non-electrically conducting insulator layer on the carbon produces high capacitance.

Although the supercapacitor is an electrochemical device, no chemical reactions are involved in the energy storage mechanism. Rather the electrolytic solution is polarised by the electrode potentials, forming an ion absorption layer on the activated carbon electrodes, in order to store energy electro-statically by microscopic charge separation at the electrochemical interface. The double layer interface potential distributions are shown in figure 23.45. The mechanism is readily reversible, as is the charge polarity since the anode and cathode are generally composed of the same material.

Due to the thin (of the order of nanometres) dielectric layer thickness, supercapacitors have low voltage ratings of typically 2V to 5V, which restricts energy storage magnitudes, since the stored energy is proportional to the square of the voltage, yet the breakdown electric field is proportional to dielectric thickness. Since conduction through an ionic liquid is relatively slow, discharged limits are lower than with standard capacitors, but can be discharged (and charged) quickly compared to electro-chemical batteries. Applying a voltage differential between the positive and negative plates charges the supercapacitor. If the electrode voltage difference is increased to the decomposition voltage, current begins to flow due to electrolyte break down.

Whereas a dry capacitor consists of conductive foils and a dry plastic separator, like the electrolytic capacitor, the supercapacitor crosses into battery technology by using special electrodes and an electrolyte.

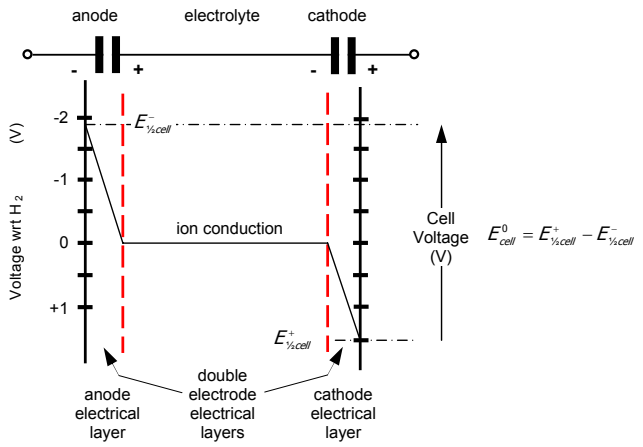


Figure 23.45. *Voltage potential profile across the battery cell, fuel cell or electrochemical capacitor, where the structure and composition of the electrical double layers differ at the anode and cathode.*

i. Electrodes

As with Li-ion technology, higher useable surface areas are attained with regular lattice structures. In the case of the supercapacitor, this is offered by carbon nanotubes, activated polypyrrole, and nanotube impregnated papers. Higher voltage properties are offered by improved insulators based on the piezio-electric material barium titanate coated with aluminium oxide and glass.

- Carbon
  - Activated – graphite Basal plane, hexagonal structure
  - graphite
  - black
  - nanotube
    - Carbon nanotubes have excellent nanoporosity properties, allowing tiny spaces for the polymer to sit in the tube and act as a dielectric.
  - carbon aerogel
    - Supercapacitors are also made of carbon aerogel. The electrodes of aerogel supercapacitors are usually of non-woven paper made from carbon fibres and coated with organic aerogel, which then undergoes pyrolysis. The paper is a composite material where the carbon fibres provide structural integrity and the aerogel provides the required large surface.
  - carbon fibre
- metal oxide, see 23.10.8.
  - RuO_2
  - $TiNO_x$
  - FeN_x
- Some conducting polymers, for example, polyacenes, have a redox (reduction-oxidation) storage mechanism along with a high surface area, see 23.10.8.

In the case of carbon aerogels the desirable properties are:

- Solid substances similar to gels but where the internal liquid is replaced with air.
- Porous and lightweight. They are typically 50 to 99.5% air, yet can theoretically hold 500 to 4,000 times their weight in applied force. They can have surface areas ranging from 250 to 3,000 m²/g, equating to 250F/g.
- One of the best insulators and lowest density solids.
- This technology will improve ultracapacitors by swapping in carbon nanotubes. This greatly increases the surface area of the electrodes and the ability to store energy since the amount of energy super-capacitors can retain is related to the surface area and conductivity of electrodes.
- They can be produced as thin films, powders, monoliths, or micro-spheres. The main problem is the cost.

The high surface area electrode activated carbon material is the cheapest of the possibilities to manufacture and is the most common. Since it stores the energy in the double layer formed at the carbon electrode surface, such a structure is termed a Double Layer Capacitor.

ii. Electrolyte

The electrolyte may be aqueous (H_2SO_4 or KOH) or organic, or a mixture of volatile acetonitrite and ionic salts. The aqueous electrolytes offer low internal resistance but limit the voltage to 1.23V. An organic electrolyte allows a dissociation voltage of greater than 2.5V (about 4 times the energy density), but the internal resistance is higher. Due to the high conductivity of the electrolyte, in spite of the small distance between the electrodes, a series equivalent resistance in the mΩ range is achieved with the entire electrolyte being absorbed in the activated carbon layer. However, with both electrolyte types, the internal resistance is less than that of an electrochemical battery.

iii. Separator

The separator, about 25μm thick, (organic electrolytes use polymer or paper while aqueous electrolytes use glass fibre or ceramic), has three functions to fulfil in supercapacitor technology:

- provide electrical insulation between the electrodes of opposite polarization,
- support the ionic conduction from one electrode to the other, and
- must not be electrically conducting.

The separator series resistance is proportional to the separator thickness which is in the range of 10 to 100 micrometre and inversely proportional to its area. The conductivity through the separator is proportional to its porosity (40% to 70%) and is given by

$$\sigma = \sigma_o \times \rho^{\alpha} \qquad 1.5 < \alpha < 2 \tag{23.15}$$

where σ_o is the electrolyte conductance, ρ is the porosity, and α is a power factor.

Most separators are a cellulose based material or porous polypropylene. They are produced in the form of self-standing films. Some other special materials (ceramic, PTFE, etc.) are intended to overcome the temperature limitation of the traditional separators. The film thickness should be as thin as possible. Its minimum thickness is limited by the electrical shorting failure risk due to free carbon particles which may create a contact between the two electrodes, and by the reduced mechanical tension which may lead to tearing of the film during the winding process.

23.10.1 Double layer capacitor model

C_{DLa} and C_{DLc} are the double layer capacitances of the anode and cathode respectively, R_i is the internal electrolyte resistance, and Z_{DL} represent the electrode impedance and dielectric leakages of the cell, as shown in figure 23.46.

For capacitors in series

$$\frac{1}{C_{DL}} = \frac{1}{C_{DLa}} + \frac{1}{C_{DLc}} \tag{23.16}$$

where for a supercapacitor $C_{DLa} = C_{DLc}$ thus

$$C_{DL} = ½C_{DLc} \tag{23.17}$$

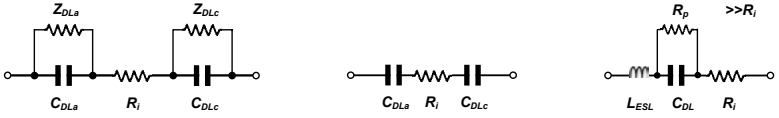


Figure 23.46. *Derivation of the electrical equivalent circuit of the double-layer super-capacitor: (a) electrode-solution interface capacitance and impedances; (b) double-layer capacitances and electrolyte resistance; and (c) R-C series equivalent circuit model, with equivalent series inductance.*

Ultracapacitors exhibit non-ideal behaviour, which results primarily from the porous material used to form the electrodes that causes the resistance and capacitance to be distributed such that the electrical response mimics transmission line behaviour. Therefore, in some modelling situations, it may be more useful to model the ultracapacitor as a distributed RC transmission line.

As shown in figure 23.47, an ultracapacitor's voltage profile, under a constant current discharge, has the capacitive component $\Delta V_C = I\Delta t/C$ and resistive component $\Delta V_{ESR} = IR$, represented by

$$\Delta V = \Delta V_{ESR} + \Delta V_C = I \times (R + \frac{\Delta t}{C}) \tag{23.18}$$

where, $\Delta V = V_w - V_{min}$ is allowable change in voltage [V], I is current [A], R is ESR [Ω], Δt is charge or discharge $t_{discharge}$ time [s], and C is capacitance [F].
The charge profile is similar, but voltages increase rather than decrease.

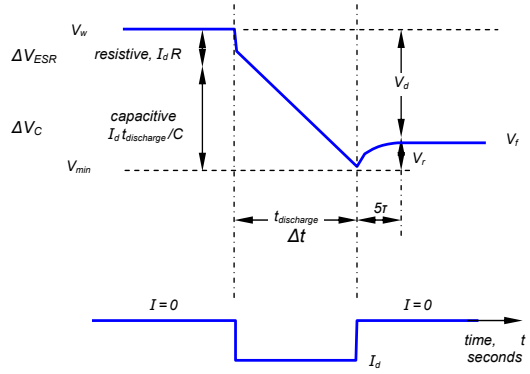


Figure 23.47. *Voltage-time discharge profile of a supercapacitor.*

Minimum allowable operating voltage, power/current requirements, and discharge time all determine cell size, and number of cells in parallel (and series). The number of cells in series, N_s tends to be specified by the voltage requirement, while the current requirements are specified by the number of cells in parallel, N_p. Since identical cell types must be used (to ensure proper sharing), the equivalent capacitance and series resistance are given by

$$C_\Sigma = C_{cell} \times \frac{N_p}{N_s} \qquad R_\Sigma = R_{cell} \times \frac{N_s}{N_p} \tag{23.19}$$

The basic information needed to determine the required ultracapacitor size, for a given application is:
- maximum operating voltage, V_w
- minimum allowable operating voltage, V_{min}
- power or current
- discharge duration (power required in each pulse) and allowable voltage change
- repetition rate
- temperature and required life

Example 23.6: ***Ultracapacitor module design using a given cell***

A 10kW un-interruptible power supply, UPS, application requires energy back up for 5 seconds. The operating voltage is 56V, and must operate down to half that voltage. The nominal tolerance on the input is better than 10%.
The specifications for the supercapacitor cells are as follows (at T_{cell} = 25°C):

R_{max} = 0.47 mΩ
C_{max} = 1500 F
V_w = 2.5 V

Will a module based of these cells fulfil the necessary back up function?

Solution

The average current requirement over 5 seconds, in order to maintain 10kW as the module voltage drops, is

$$\bar{I} = \frac{power}{\bar{V}} = \frac{power}{½(V_{max} + V_{min})}$$
$$= \frac{10{,}000}{½(56+28)} = 238\text{A}$$

The number of series cells N_s for maximum voltage is

$$N_s = \frac{V_{max}}{V_w} = \frac{60\text{V}}{2.5\text{V}} = 24 \text{ cells in series}$$

Assuming no parallel connection is necessary, N_p = 1, the module capacitance and resistance are

$$C_\Sigma = C_{cell} \times \frac{N_p}{N_s} = 1500\text{F} \times \frac{1}{24} = 62.5\text{F}$$
$$R_\Sigma = R_{cell} \times \frac{N_s}{N_p} = 0.47\text{m}\Omega \times \frac{24}{1} = 11.3\text{m}\Omega$$

Equation (23.18) is used to estimate the voltage drop during the 5-second period, that is

$$\Delta V = I \times (R + \frac{\Delta t}{C})$$
$$= 238\text{A} \times (11.3\text{m}\Omega + \frac{5\text{s}}{62.5\text{F}}) = 2.7\text{V} + 19\text{V} = 21.7\text{V}$$

Since the specification allowed a voltage decrease of 28V, (56V - ½x56V), the expected 21.7V is within the specification limit. The necessary capacitance per cell can be decreased to 1500Fx21.7/28V=1162F.

♣

23.10.2 Cell parameter specification and measurement methods

i. Charging and discharging specification
The time required for constant current and constant resistance discharging are represented by equations (23.20) and (23.21) respectively. It is assumed the internal resistance is negligible.

Charge/discharging time of constant current charge/discharge

$$t = C \times \frac{|\Delta V_o|}{i} = C \times \frac{|V_o - V_f|}{i} \tag{23.20}$$

Charge/discharging time of constant resistance R charge/discharge

$$t = CR \times \ell n \frac{V_{Xsource} - V_0}{V_{Xsource} - V_f} \tag{23.21}$$

where t = charge/discharging time (s), $V_{Xsource}$ = external voltage source (V), V_o = initial voltage (V), V_f = final voltage (V), i = capacitor current (A).
The ultracapacitor terminal voltage is measured during charge/discharge, that is, under load.

ii. Capacitance specification
Initially the capacitor is slowly charging over a 30-minute period to rated voltage. The capacitor is then discharged at a constant rate of 1mA/F, producing a linear decease in terminal voltage. The time is measured for the terminal voltage to drop from 70% to 30% of rated voltage and the capacitance is given by

$$C = i \times \frac{t_{0.7V_R} - t_{0.3V_R}}{V_{0.7V_R} - V_{0.3V_R}} = i \times \frac{t_{0.7V_R} - t_{0.3V_R}}{0.7 \times V_R - 0.3 \times V_R} \quad \text{(F)} \tag{23.22}$$

where the voltages and times correspond to 70% and 30% of the capacitor rated voltage, V_R.
Alternatively the calculation is performed for the period when the rated voltage decreases to 50% of rated voltage.

iii. Equivalent Series Resistance, AC and DC ESR, specification
AC ESR is measured with a 4-probe impedance analyzer at an AC amplitude of 5mV and a frequency of 100Hz or 1kHz. Because of the ultracapacitor time constant, typically one-second, operation at these frequencies is inefficient. The ac resistance is used to define the volumetric and gravimetric power densities as follows.

$$P_v = \frac{V_{Rated}^2}{4R_{ac}} \times \frac{1}{Volume\ dm^3}$$
$$P_g = \frac{V_{Rated}^2}{4R_{ac}} \times \frac{1}{Weight\ kg} \tag{23.23}$$

The DC ESR is measured after the capacitor has been slowly charge to rated voltage. The ultracapacitor is then discharged at a constant current of 1mA/F. The internal resistance voltage drop of ΔV is measured (the final voltage is taken 5 seconds after the constant current loading is removed) and the internal resistance, expressed in MΩ, is calculated from

$$R = \frac{\Delta V}{i} \qquad (\text{m}\Omega) \tag{23.24}$$

The total internal dc resistance is compromised of resistive components attributed to contact, electrode, electrolyte, and other material resistances.

iv. Leakage current specification
The capacitor is charged with the rated voltage at 25°C for at least 12hrs, typical 72 hours. Then the voltage source charging current, leakage current, is measured.

23.10.3 Cell characteristics

i. Charge/discharge
Typically, the supercapacitor has an RC time constant of about one second, whence the charge time of a supercapacitor is better than 10 seconds. Both charge and discharge rates tend to be limited by the current related internal heating of the electrodes, while peak current is limited by the internal resistance. The internal resistance (measured at a low frequency, $\frac{1}{5}$ Hertz), increases significantly below 0°C, doubling at -40°C, as seen in figure 23.48. The charge characteristics are similar to those of an electrochemical lead-acid battery. Full charge occurs when a set voltage limit is reached. When the charger open circuit voltage is reached, charging stops, without danger of overcharge or *memory* effect.
The initial charge is rapid, with current limiting required; the topping charge takes extra time. Over charging (higher voltages), reduces lifetime, whilst uncontrolled charging can damage cell shunt protection circuitry.
The supercapacitor can be rapidly charged and discharged virtually an unlimited number of times. Unlike the electrochemical battery, there is little deterioration or detrimental stress induced by cycling and age does not significantly affect the supercapacitor. In normal use, a supercapacitor deteriorates to about 80% charge capability after 10 years. Lifetime is usually defined as a 20% reduction in capacitance and/or a 200% increase in resistance (see estimated life duration).
Because of the typical one-second internal RC time constant, short circuiting need not be fatal and provided thermal limits are observed, consequential open circuit failure can be avoided.
The internal self-discharge of the supercapacitor is substantially higher than that of an electro-chemical battery. Organic electrolyte supercapacitors discharge most. In 30 to 40 days, the capacity decreases from full charge to 50%, with 1.2% per day charge decrease being typical. In comparison, a nickel-based battery discharges about 10 percent during the same period.
Figure 23.48 show the per unit increase in internal resistance at lower temperatures, whilst in comparison, capacitance is virtually temperature stable.

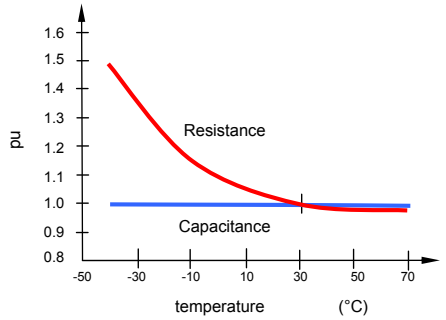


Figure 23.48. *Supercapacitor temperature dependence of capacitance and internal resistance.*

ii. Frequency Response
Ultracapacitors have a typical time constant of approximately one second, for both step voltage source charging and discharging.
Therefore, it is not possible to expose ultracapacitors to a continuous ripple current as internal overheating may result. The ultracapacitor can respond to short pulse power demands, but due to the time constant the efficiency or available energy is reduced. Figure 23.49 illustrates the performance of the ultracapacitor at various frequencies. The drop off in capacitance is associated with the response time necessary for the charged ions within the pores of the electrodes to transport between positive and negative electrodes during charge and discharge. The drop in resistance is representative of the response time of the different resistive elements within the ultracapacitor. At low frequency, all resistive elements are present while at high frequency, only quick response elements such as contact resistance are present.

The continuous DC current density recommended for maximum long life reliable operation of the ultracapacitor cells is 50mA/F. The cells will operate effectively at higher current densities however, potential cell damage can occur depending on the magnitude and duration/duty cycle of the high current pulses.

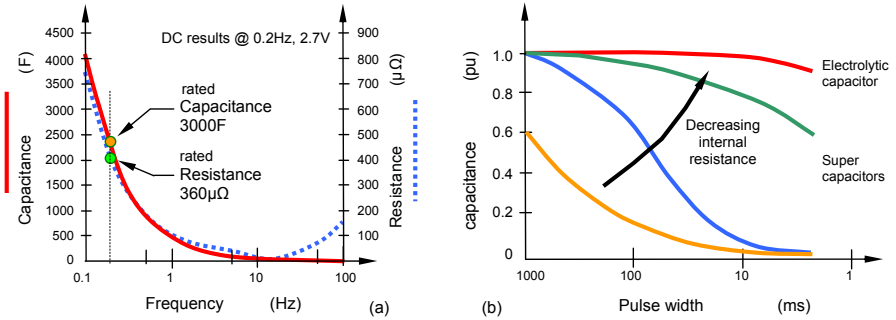


Figure 23.49. *Supercapacitor frequency dependence of capacitance and internal resistance.*

Two factors are critical in determining the voltage drop when a capacitor delivers a short current pulse; namely the ESR and 'available' capacitance as shown in Figure 23.49b.
The instant voltage drop ΔV_{ESR} is caused by and is directly proportional to the capacitor's ESR. The continuing voltage drop with time ΔV_C, is a function of the available charge, that is, capacitance. From Figures 23.47, it is apparent that, for short current pulses, for example, in the millisecond region, the combination of voltage drops in a conventional supercapacitor caused by, first the high ESR and second, the lack of available capacitance, causes a total voltage drop, unacceptable for many applications. Low ESR, (in mΩ), minimizes the instantaneous voltage drop, while the high retained capacitance drastically reduces the severity of the charge related drop.
Ultracapacitors used in electric drive-trains to load-level the battery, experience large-steady (transient) dc, much like the battery, rather than small amplitude ac signals. The dc charge or discharge time (t_{disch}) of the capacitor is related to the fundamental characteristic frequency (f_{AC} in Hz) of the ac voltage on the capacitor by $t_{disch} \gg \frac{1}{4}f_{AC}$. Hence, for backup applications, the ac signals are less than 10Hz.

In testing ultracapacitors, they are conveniently modelled as a simple series RC circuit when inductive effects are not important. Then $Q = CV$, $W = \frac{1}{2}CV^2$ and $V_o - V = iR + (Q_o - Q)/C$, where Q is charge on the capacitor, V is voltage on capacitor, W is energy stored in the inductor, and V_o and Q_o are voltage and charge at $t = 0$, respectively. Based on these assumptions, figure 23.50 parts a and b show capacitor voltage variation during constant current and constant power discharge.

23.10.4 Thermal Properties

Many ultracapacitors applications utilize the components under high duty cycles. One of the factors attributing to ultracapacitor performance reduction is temperature. For minimum performance influence over the equipment life, it is necessary to maintain the ultracapacitor core temperature within its rated temperature range. The lower the maintained core temperature, the longer the life.
If cells are provided with an electrically insulating shrink sleeving around the capacitor body and since all current passes through the capacitor terminals, cooling at the capacitor ends or terminals is the most efficient means of capacitor cooling.

Depending on the duty cycle of the application, cooling can be accomplished via heat sinks (conduction), air-flow (convection) or a combination of the two. Consideration should be made for the duty cycle and resulting capacitor temperature as well as the anticipated ambient temperature surrounding the capacitor. The combination of the two should not exceed the operating temperature for the ultracapacitor.

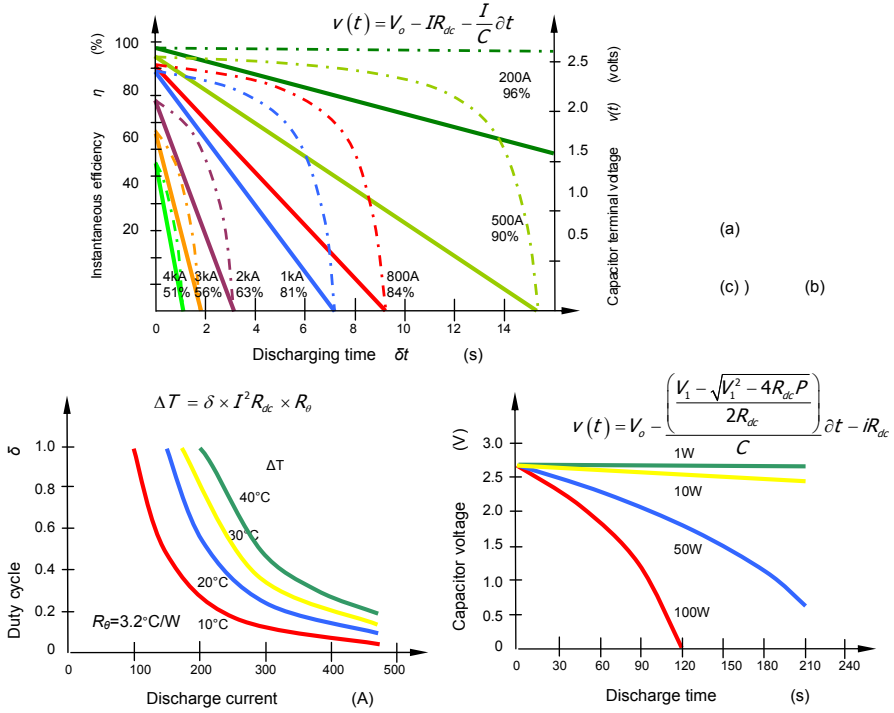


Figure 23.50. *Supercapacitor 3000F, 2.7V, 0.29mΩ cell performance characteristic profiles: (a) constant current discharge, (b) constant power discharge, and (c) thermal dissipation with free convection in a 25°C ambient.*

The thermal resistance, or R_θ (°C/W), is based on free convective cooling, and is improved by active cooling. Utilizing R_θ, an anticipated temperature rise can be predicted based on the application current and the duty cycle. Typical core temperature rise (above ambient) dependence on current and duty cycle are shown in figure 23.50c. This temperature rise is then added to the ambient temperature to determine the maximum current and duty cycle to be maintained the capacitor within its operational specification and the application life requirements. The core temperature rise of the capacitor is predicted by:

$$\Delta T = \delta \times I^2 R_{dc} \times R_\theta$$

where δ = duty cycle (0 to 1), I = current, R_{dc} = dc resistance or low frequency based on constant discharge (non pulsing).

Alternatively, for ac currents the high frequency resistance should be utilized with pulsing currents.

23.10.5 Estimated life duration

During use, capacitance decreases and internal resistance rises, as shown in figure 23.52. The lifetime of an ultracapacitor is affected by ambient temperature, applied voltage and operating current. Capacitor lifetime is increased by reducing these factors.

i. Operating Temperature Dependence

Capacitor life is affected by operating temperature. In general, lowering ambient temperature by 10°C doubles the life of a capacitor. Operation above the maximum specified temperature not only shortens capacitor life, but can also cause damage such as electrolyte leakage. The operating temperature of the capacitor not only involves the ambient temperature and internal self-heating generated temperature, due to dc and ripple current, but also radiation from other nearby heat generating elements.

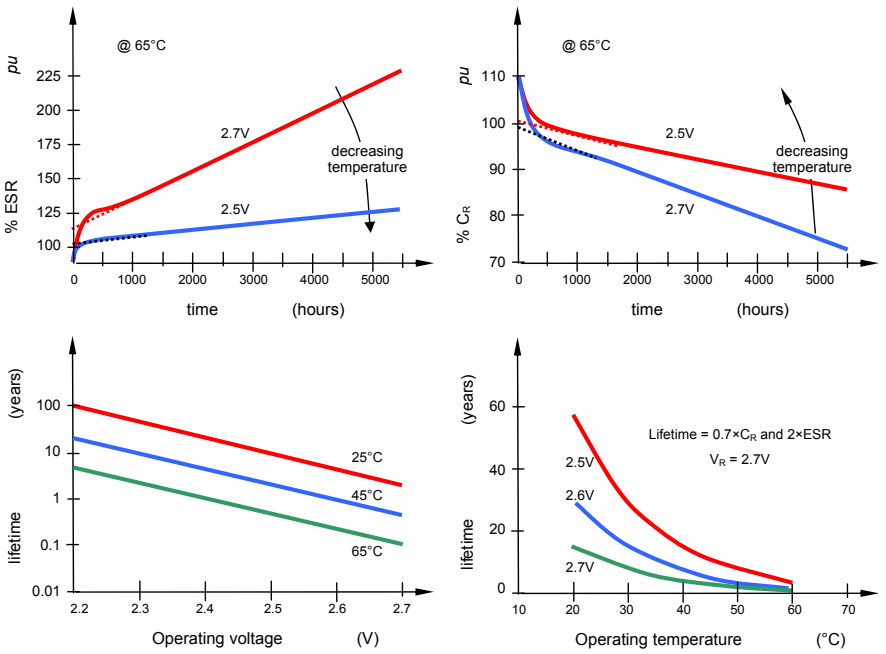


Figure 23.52. *Supercapacitor parameter variation (lifetime) over time (cycling) and with temperature and voltage (a) equivalent internal series resistance, (b) rated capacitance, (c) operating voltage, and (c) operating temperature.*

ii. Voltage Dependence

If an ultracapacitor is used at a voltage exceeding its rated voltage, not only is its lifetime shortened, but depending on the actual voltage, gas generated by electrochemical reactions inside the capacitor may cause leakage or rupture. Reverse voltage polarity is allowed but shortens the life expectancy.
Figure 23.52c show the deterioration of lifetime related parameters, over time, depending on operating voltage and operating temperature.

High-temperature load lifetime is used to measure the lifetime of ultracapacitor. For example, 1000 hours under full charging conditions at a temperature of 70°C is equivalent to 7.3 years at room temperature, 25°C, under normal use.

The lifetime criteria are:

A reduction in capacitance of 30%, C_{limit}=80%$C_{nominal}$; and
An increase in ESR (equivalent serial resistance) of 100%.

These electrical parameter operational limits are referred to as endurance capacitance and endurance resistance. Typical capacitance decays to the -20% level before the ESR doubles. Therefore, capacitance is generally the first parameter to experience end of life based on the arbitrary criteria.
The progression of capacitance in time is separated into three distinct sequential periods, as shown in figure 23.53:

Exponential decrease during the first hours/cycles (burn-in)
Linear decrease during the main part of the life
Slow exponential decrease due to supercapacitor natural aging (burn-out)

The model applies only when the capacitor is operating in the middle period in which there is a linear decrease of capacitance (and a linear increase in ESR) over time.
From figure 23.53, estimated life duration, ELD, for a 30% decrease in capacitance, is defined by:

$$\frac{C_1 - C_2}{\Delta t} = \frac{C_1 - C_{limit}}{EDL} \quad \text{that is} \quad EDL = \Delta t \frac{C_1 - C_{limit}}{C_1 - C_2}$$

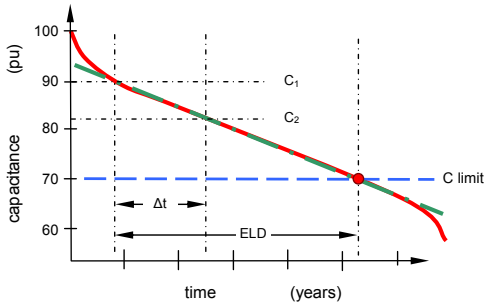


Figure 23.53. *Supercapacitor lifetime determination.*

A phenomenon termed *recovery* is observed, which is to be avoided during lifetime measurement periods. If the capacitor loading is interrupted after a long period of continuous usage (DC or voltage cycles), the measurement of capacitance and series resistance show a recovery as a function of the time without use. The capacitance recovery is explicit in the voltage recovery V_r, shown in figure 23.47, after the load is removed.

23.10.6 Cell Voltage Equalization in a Series Stack of Ultracapacitors

Many applications require that capacitors be series and/or parallel connected, to form a 'bank' or 'module' with a specific voltage and capacitance rating. Because sustained overvoltage can cause an ultracapacitor to fail, the voltage across each cell in a series stack must not exceed the maximum continuous working voltage rating of individual cells in the stack. The charging electronics must either reduce the rate of charge being delivered to a cell, or stop charging a cell whose voltage approaches its surge voltage rating.

i. Cell Balancing in Low Duty Cycle Applications

The voltage distribution in a series stack of ultracapacitors initially and during transients is a function of capacitance. After the stack has been held at voltage for a period-of-time, the cell voltage distribution becomes a function of leakage current, modelled by current through internal parallel resistance.

Initially, the cells with greater capacitance will be charged to lower voltages while the cells with smaller capacitance will be charged to higher voltages. This is because each series cell conducts the same current, and voltage is a function of current and capacitance. The average cell voltage will be the total voltage divided by the number of capacitor cells.

Any cell supporting voltage discharges through internal parallel resistance. The current through this parallel resistance is referred to as leakage current. The leakage current has the effect of self-discharging the cell.

After a period-of-time, the voltages on the individual cells will vary based on the differences in leakage current, rather than on the differences in capacitance. The cells with higher leakage retain lower cell voltages, and vice versa. This is because the higher leakage current discharges the cell, lowering its voltage. If the series string is held on a constant voltage source, this voltage, via the leakage current, is redistributed onto other cells in the series arrangement.

One technique to compensate for variations in leakage current is to place a bypass resistor in parallel with each cell, sized to dominate the total cell leakage current. This effectively reduces the variation of equivalent parallel resistance between the cells. For instance, if the cells have an average leakage current of 10uA ±3uA, a resistor with 1% tolerance which bypasses 100uA ± 1uA is appropriate. The average leakage current becomes 110uA, ±4uA. Introduction of this resistor decreases the variation in leakage current from 30% to 3.6%.

Since the parallel resistances are the same, the cells with higher voltages will discharge through the parallel resistance at a higher rate than the cells with lower voltages. This distributes the total stack voltage uniformly across the entire series of capacitors. See series voltage sharing in Chapter 10.

A typical trade-off is based on time to balance versus leakage current; the faster the balancing circuit responds, the greater the leakage. A 10:1 ratio of bypass resistor leakage current to cell leakage current may take days to balance a severely unbalanced bank, while a 100:1 ratio may balance in a few hours. Once the system is balanced, response time is less of an issue unless the bank is severely cycled.

For high duty cycle applications, more efficient, active balancing is normally used, as follows.

ii. Active Voltage Management Methodology

The active balancing circuit has an active switching device, like a bipolar transistor or MOSFET, connected in series with each bypass element ladder. The switches are controlled by voltage-detection circuits that only turn a switch 'on' when the voltage across that particular cell approaches a value slightly below the continuous working-voltage rating of the cell. This is called the bypass threshold voltage. Figure 23.54 depicts a typical circuit diagram of an active charging-current diversion circuit.

Key balancing circuit features:

- Elimination of module-to-module balancing requirements
- Improved efficiency with reduction of parasitic losses
- Improved operation over full range of temperatures
- Improved filtering and operation with noisy input power
- Increased balancing power for quicker balancing

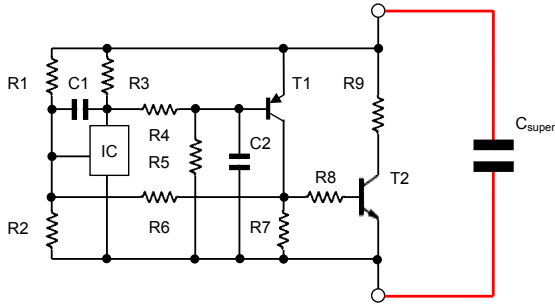


Figure 23.54. *Supercapacitor voltage balancing circuit.*

Balancing electronics is a current sinking topology capable of sinking current at 300mA to 400mA from each individual cell, whenever the specific cell voltage is at least 2.73V. The on-off trigger of the voltage management circuits are optimized for performance in terms of temperature and other influences. Each cell voltage is individually monitored against a nominal reference voltage. When the cell voltage reference is exceeded, the voltage management circuit discharges the cell voltage to below the trip level. Then the voltage management circuit for that cell enters a quiescent state with respect to balancing activity. Maintaining the cell voltage below this 'trigger' level assures cell long lifetime and, therefore, module long lifetime.

The circuitry features tolerance to a full charge reversal, without damage to the circuit electronics or capacitors. If an ultracapacitor module is allowed to approach complete self-discharge, it is possible that some capacitors in the module experience a negative voltage. This negative state and subsequent recharging of the module will not damage the balancing electronics.

Circuit temperature compensation enables identical performance from the circuit over the range of allowable operating temperatures. This is accomplished through a stable voltage reference subsystem for the entire temperature range ensuring accuracy of the balancing scheme while in operation. This also prevents voltage management circuit instability during cell fast transient voltage conditions (discharge-charge) or if the voltage management circuit is supplied with the current from the capacitor.

Integrated monitoring indicates an extreme stop-charge condition (when the cell exceeds maximum allowed operating voltage). The circuit 'triggers' at 2.73V and rapidly reduces the cell voltage to below the threshold of 2.70V, at which point a quiescent state reoccurs. Typical balance times are measured in minutes and the smaller the cell capacitance the more rapidly the voltage management circuit reduces the cell voltage.

The circuit quiescent current is approximately 30uA at room temperature. As a result, modules (including the heavy-duty transportation modules) featuring this circuit experiences an extremely low parasitic discharge. In addition, the design tolerates a wide variation in input power making it robust relative to any 'dirty power' requirement.

High-voltage modules have a single opto-isolated output for over voltage indication when any cell in the string approaches an over voltage condition, 2.80V, ± 0.6V. This signal monitors that the state of charge of each cell in the module is in the safe operating range. Active voltage balancing occurs prior to the threshold for output of the over voltage signal, hence can be used as a stop charge signal. Charging can resume when the signal is no longer present. When the over voltage signal is triggered, failure to stop the charging will result in charging the cell to even higher voltages, damaging the cell or shortening the cell lifetime depending upon the charge rate. Monitoring the over voltage signal when operating the bank near its maximum voltage can be used to maximize bank lifetime. An ultracapacitor can operate at its maximum voltage, surge voltage, for a few seconds without irreversible damage or cell open circuiting.

23.10.7 Supercapacitor general properties

i. Uses: Rather than operating as a main battery, supercapacitors are more commonly used as power and memory backup during short power interruptions. Another application is improving the power density, the current handling, of a battery. The supercapacitor is placed in parallel to the battery and provides current boost on high load demands. The supercapacitor can also fulfil the same role for portable fuel cells that have a lower power density than the battery. Because of its ability to rapidly charge - high power density - large supercapacitors can used to store regenerative braking energy from electric/hybrid vehicles. Hundreds of supercapacitors are series and parallel connected to attain the required voltage, hence energy storage capacity. Some applications and range of ratings for super-capacitors are shown in Table 23.23.

Table 23.23: Super-capacitor ratings and applications

Energy	Voltage	Capacitance	Mass	Height @ dia 226 mm	Application
kJ	**V**	**F**	**kg**	**mm**	
9	14	100	11	125	4/6-Cylinder Cars
28	28	55	20	280	Off-Road Vehicles
40	96	8.5	31	390	Rail-Road Engines
40	300	0.85	32	410	Power Supplies
44	140	4.5	22	380	Pulse-Power
6	300	0.13	10	300	Electro-magnetic Switches

Table 23.24 gives a summary of some critical properties of different energy source technologies. Because there are so many types with widely different properties, battery values are shown as a range.

Supercapacitors have several advantages over batteries:

- indefinite number of charging and discharging cycles;
- maintenance free and safe;
- light weight;
- high-rate discharge – high power density;
- higher efficiency, with lower internal resistance;
- lower internal heating and much wider operating temperature range -50°C to 85°C;
- recharge quickly, since no slow chemical reaction involved; and
- no negative environmental impact.

Supercapacitors are more expensive than batteries and have a lower volumetric efficiency.

Table 23.24: Storage cell comparison

Property	units	Supercapacitors	MP Capacitors	Micro-Fuel Cells	Batteries
Charge/Discharge Time	s	milliseconds to seconds	picoseconds to milliseconds	Typically 10 to 300 hrs. Instant charge (refuel).	1 to 10 hrs
Operating Temperature	°C	-40 to +85	-20 to +100	+25 to +90	-20 to +65
Operating Voltage	V	2.3-2.75 V/cell	6 to >800	0.6 V/cell	1.25 - 4.2V/cell
Capacitance	F	100 mF to > 2F	10 pF - 2.2 mF	N/A	N/A
Efficiency	%	90 to 95	99	50	90 to 95 (Li-ion)
Life/Cycles	hrs	>30,000 hrs ave 10^5 - 10^6	>10^5 cycles	1500 to 10^4 hrs	150 to 1500 cycles
Weight	kg	1 - 2 g	1 g - 10 kg	20 g to > 5 kg	1 g to > 10 kg
Gravimetric Power Density	kW/kg	10 to 100	0.25 to 10,000	0.001 to 0.1	0.005 to 0.4
Gravimetric Energy Density	Wh/kg	1 to 5	0.01 to 0.05	300 to 3000	8 to 600
Volumetric Energy Density	Wh/l	0.05-10 Wh/l			50-250 Wh/l
Pulse Load	A	< 100	< 1000	< 150 mA / cm^2	< 5
Pollution		none	none	none	Cd/Pb

Supercapacitors have a higher energy density than the conventional capacitor and a higher power density than the electrochemical battery. The gravimetric energy density of the supercapacitor is 1Wh/kg to 10Wh/kg. Although this energy density is high in comparison to an electrolytic capacitor, as shown in figure 23.2, it represents only one-tenth that of the nickel-metal-hydride battery. Whereas the electro-chemical battery delivers a near constant voltage in the usable energy spectrum, the charge-voltage of the supercapacitor is linear and drops uniformly from full voltage to zero volts during a constant discharge current. Because of this, the supercapacitor is unable to deliver the full charge.
If, for example, a 12V lead-acid battery discharges to 9V before the equipment cuts off, the supercapacitor reaches the same voltage within the first quarter of the discharge cycle, with over half the energy remaining in the capacitor, and that being difficult to exploited.
To operate at higher voltages, supercapacitors, like Li-ion cells, are connected in series. Also, like Li-ion batteries, for the series connection of more than three capacitors, voltage balancing is required to prevent any cell from experiencing and over-voltage.
Supercapacitors are relatively expensive in terms of cost per Watt, hence it may be better to use a larger battery with extra cells. But the supercapacitor and electro-chemical battery are not necessarily alternatives, rather they are complementary.

General features of super-capacitors are highlighted in the following points.

Advantages

- High power density
- Wide operating temperature range, -50°C to over 85°C:
  - capacity increases with decreased temperature,
  - lifetime doubles with every 10°C decrease in operating temperature, and
  - lifetime increases with decreased operating voltage.
- Virtually unlimited cycle life - can be cycled millions of times.
- High rates of charge and discharge.
- Chargeable and operable at any voltage up to its voltage limit.
- Reversibility, but asymmetrical charging and discharging rates.
- Low impedance - enhances load handling when paralleled with a battery.
- Rapid load discharging, without thermal heat.
- Charge - discharge cycling efficiency of over 95%, but reduce with high power pulses.
- Rapid charging - charged in seconds.
- Simple charge methods - no full-charge detection is needed; no danger of overcharge, not affected by deep discharges. State of Charge only a function of voltage.
- Do not involve hazardous substances. Aluminium case, etc, (25% to 30% by weight) can be recycled.
- Can be stored, transport, etc., safely retaining zero charge, since hermetically sealed.
- Cost-effective energy storage - lower energy density is compensated by a high cycle count.

Limitations

- Linear discharge voltage prevents use of the full energy spectrum. Stored charge highly dependant on operating voltage level.
- Low energy density - typically stores one-tenth the energy of an electrochemical battery.
- Cells are low voltage, 2V to 3V, - serial connection is needed to obtain higher voltages.
- Internal short-circuiting rate increases as voltage rating increases.
- Voltage balancing is required if more than three capacitors are series connected.
- High self-discharge - the rate is considerably higher than an electrochemical battery.
- Life expectance is decreased with increased humidity, temperature, and voltage.
- Possible susceptibility to resonant vibration modes.
- Generally restrict to cylindrical in shape. Typical shape form factor results in reduced volumetric densities.
- The enigma of high cost.

23.10.8 Pseudocapacitors

In contrast to EDLCs, which store charge electrostatically, pseudocapacitors store charge Faradaically through the transfer of charge between the electrode and electrolyte. This is accomplished through electrosorption, reduction-oxidation reactions, and intercalation processes. These Faradaic processes allow pseudocapacitors to achieve greater capacitances and energy densities than EDLCs.
There are two electrode materials used to store charge in pseudocapacitors, namely, conducting polymers and metal oxides, RuO_2 (rather than an activated carbon electrode material as in supercapacitors).

Conducting polymers have a relatively high capacitance and conductivity, plus a relatively low ESR and cost compared to carbon-based electrode materials. The n/p-type polymer configuration, with one negatively charged (n-doped) and one positively charged (p-doped) conducting polymer electrode, has greater potential energy and power densities; however, a lack of efficient, n-doped conducting polymer materials prevents pseudocapacitors from reaching their potential. The mechanical stress on conducting polymers during reduction-oxidation reactions limits the stability of pseudocapacitors through many charge-discharge cycles, thus hindering conducting polymer pseudocapacitor development.
Because of their high conductivity, *metal oxides*, particularly ruthenium oxide, is an electrode material used for pseudocapacitors. Other metal oxides attain lower capacitance. The capacitance of ruthenium oxide is achieved through the insertion and removal, termed intercalation, of protons into its amorphous structure. In its hydrous form, the capacitance exceeds that of carbon-based and conducting polymer materials. The ESR of hydrous ruthenium oxide is lower than that of other electrode materials. As a result, ruthenium oxide pseudocapacitors can achieve higher energy and power densities than similar EDLCs and conducting polymer pseudocapacitors, but at a prohibitive cost. The *RC* time constant is about 5ms.

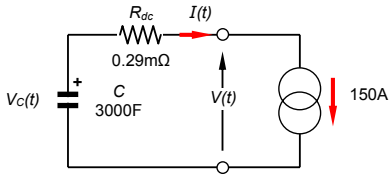
Example 23.7: ***Ultracapacitor constant current characteristics***

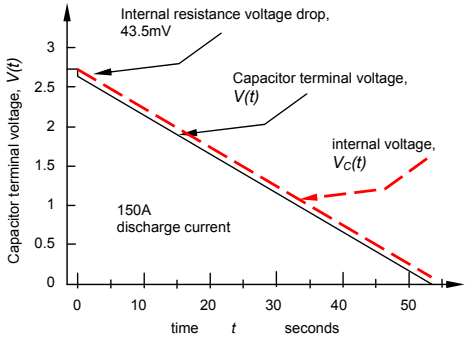
A 3kF supercapacitor with an internal dc resistance of 0.29mOhms is charged to 2.7Vdc. Based on a series RC model for the supercapacitor:

(a) if it is discharged from 2.7V at a constant 150A rate, determine:
- *i.* the initial stored energy;
- *ii.* the instantaneous initial output voltage and internal power losses;
- *iii.* an expression in time for the output voltage and ideal capacitor voltage, hence power
- *iv.* the final capacitor voltage, the time to reach this voltage, and the residual stored energy;
- *v.* the energy dissipated in the internal resistance during the constant current discharge period; and
- *vi.* energy delivered to the constant current load, hence the overall efficiency of energy transfer.

(b) if it is charged at a constant 150A rate from a residual voltage level of 1V to 2.7V, determine:
- *i.* the terminal voltage range to charge the capacitor;
- *ii.* the time to charge to 2.7V; and
- *iii.* the internal losses.

Solution





(a)
i. The initial stored energy in the 3000F supercapacitor is

$$W_o = ½CV_{Co}^2$$
$$= ½ \times 3{,}000\text{F} \times 2.7^2 = 10.935\text{kJ}$$

ii. The output voltage is the idea voltage minus the internal resistor voltage drop, that is

$$V(t=0) = V_{Co} - IR_{dc}$$
$$= 2.7\text{V} - 150\text{A} \times 0.29\text{m}\Omega$$
$$= 2.7\text{V} - 43.5\text{mV} = 2.657\text{V}$$

Since the load current is constant, the internal resistor losses are constant, assuming the internal resistance does not increase due to temperature heating effects.

$$P_{Rdc} = I^2R_{dc}$$
$$= 150^2 \times 0.29\text{m}\Omega = 6.525\text{W}$$

iii. By Kirchhoff's voltage law, the capacitor terminal voltage is

$$V(t) = V_{Co} - \frac{1}{C}\int Idt - IR_{dc}$$
$$= V_{Co} - \frac{1}{C}I \times t - IR_{dc}$$

Since the discharge current is constant the ideal capacitor component discharges at a constant rate, being the first two terms on the right hand side of the above equation. The load absorbed power is

$$P(t) = I \times V(t) = I \times V_{Co} - \frac{1}{C}I^2 \times t - I^2R_{dc}$$

iv. The constant current discharge ceases when the output voltage reaches zero, but the idea capacitor voltage is equal to the voltage drop across the internal resistance, that is

$$V(t_{dis}) = V_{Co} - \frac{1}{C}I \times t_{dis} - IR_{dc} = 0$$
$$t_{dis} = \frac{V_{Co}C}{I} - R_{dc}C = \frac{Q_{Co}}{I} - \tau$$
$$= \frac{2.7\text{V} \times 3000\text{F}}{150\text{A}} - 0.29\text{m}\Omega \times 3000\text{F}$$
$$= 53.13\text{s}$$

such that

$$V_C(t_{dis}) = IR_{dc}$$
$$= 150\text{A} \times 0.29\text{m}\Omega = 43.5\text{mV}$$

The residual stored energy remaining in the ideal supercapacitor is

$$W(t_{dis}) = ½CV_C^2 = ½ \times C \times (IR_{dc})^2$$
$$= ½ \times 3{,}000\text{F} \times 0.0435^2 = 2.838\text{J}$$

v. The energy dissipated as heat in the internal resistance is 6.525W over 53.13s, that is, 346.67J.

vi. The energy consumed by the constant current sink is 10,935J-346.67J-2.838J=10,585.49J.

The system load efficiency is

$$\eta = \frac{W_{load}}{W_{load} + losses}$$
$$= \frac{10585.49\text{J}}{10585.49\text{J} + 346.67\text{J}} \times 100 = 96.83\%$$

The load energy transfer efficiency, the energy transferred to the load in terms of the available energy, is

$$\eta = \frac{W_{load}}{W_o}$$
$$= \frac{10{,}585.49\text{J}}{10{,}935\text{J}} \times 100 = 96.80\%$$

(b)
i. The terminal voltage at the start of the charge period is

$$V = V_C + IR_{dc}$$
$$= 1.0\text{V} + 150\text{A} \times 0.29\text{m}\Omega = 0.044\text{V}$$

The terminal voltage at the end of the charge period is

$$V = V_C + IR_{dc}$$
$$= 2.7\text{V} + 150\text{A} \times 0.29\text{m}\Omega = 2.744\text{V}$$

ii. The charge time is found by solving

$$2.7\text{V} = 1.0\text{V} + \frac{It}{C}$$
$$= 1.0\text{V} + \frac{150\text{A} \times t}{3000\text{F}}$$

which yields $t = 34\text{s}$.

iii. The internal resistive losses are the power losses over 34s, that is

$$W_{R_{dc}} = I^2 R_{dc} t$$
$$= 150^2 \times 0.29\text{m}\Omega \times 34\text{s} = 221.85\text{J}$$

♣

23.11 Thermoelectric modules

A thermoelectric module is a highly reliable, small, light, solid-state device that can operate as an electronic heat pump (a refrigerator) or as a low efficiency dc electrical power generator. When used to generate dc electricity, the module operates on the *Seebeck effect* and is called a thermoelectric generator, TEG. When used as a heat pump, the module utilizes the *Peltier effect* to remove heat and is called a thermoelectric cooler, TEC.
Dissimilar metals that are connected at two different locations (junctions) will develop a micro-voltage if the two junctions are held at different temperatures. The inverse is to apply a dc voltage, which causes a temperature difference between the junctions. This results in a compact heat pump, referred to as a thermoelectric cooler, TEC. In a thermoelectric material, there are free electrons or holes that carry both charge and heat.
A TEC can be used in applications where cooling or temperature control of an object is required. In general, a TEC can be used when an object:

i. needs to be cooled below the ambient temperature, or
ii. is required to be maintained at a precise temperature under a fluctuating ambient temperature.

A TEC is ideal for cooling small, low heat load objects. Due to the low Coefficient of Performance, *CoP*, compared with compressor cooling, a TEC looses its advantage if the cooling load is higher than 300 W.

23.11.1 Background

The Seebeck coefficient is a material property that determines the performance of thermocouples and Peltier elements.
The *Seebeck effect* (or thermoelectric effect) is the direct conversion of heat energy into electrical energy, due to an emf developed due to a temperature difference between two junctions of dissimilar conductors in the same circuit. The Seebeck coefficient (or thermal emf coefficient) is related to the fact that electrons are carriers of both electricity and heat. When an electrical loop is made up with two different metals (materials A and B), two junctions exist (1 and 2). If the junction temperatures (1 and 2) are different, a net EMF is produced and a current flows. s_A and s_B are the Seebeck coefficients of the materials (the equation in figure 23.55a is simplified as Seebeck coefficients are temperature dependant). The magnitude and sign of the Seebeck coefficient are related to an asymmetry of the electron distribution around the Fermi level. Typical Seebeck coefficient values for different materials are shown in Table 23.25.
The *Peltier effect* is manifested as heat pumping (moving heat energy from a cold place to a hotter one) by using electrical energy. Specifically the effect describes the liberation or absorption of heat at a contact where current passes from one material to another, whereby the contact becomes heated or cooled. If a voltage source is inserted and a current forced into the two-metal loop, one junction cools down while the other heats up. A similar equation applies in figure 23.55b and the cooling power P_e is proportional to the difference of Peltier coefficients of the two metals (π_A and π_B) and the current I. Typical Peltier factor values for different materials are shown in Table 23.26.
Peltier cooling/heating is the phenomenon of absorption/dissipation of heat by a junction between two dissimilar materials when electrical current flows through a junction. The Peltier coefficient is a measure of heat carried by electrons or holes.
The heat absorbed/dissipated P_p by a junction is

$$P_p = \pi \times I \qquad (23.25)$$

where π is the temperature dependant Peltier coefficient, in V, corresponding to a specific pair of materials. The effect may be reversed wherein a change in the direction of electric current flow will reverse the direction of heat flow. Joule heating $I^2 \times R$, having an emf magnitude of $I \times R$ (where R is the electrical resistance of the semiconductors), also occurs in the conductors because of current flow. This Joule heating effect acts in opposition to the Peltier effect and causes a net reduction of the available cooling. Heat conducted from the hot junction, back to the cold junction also opposes the Peltier produced effect.
Seebeck power generation is a process by which heating/cooling of a junction of two dissimilar materials generates an electrical junction potential due to charge carrier diffusion and phonon drag:

$$\pi = s \times T \qquad (23.26)$$

which relates the Seebeck and Peltier coefficients, where s is the Seebeck coefficient, V/K, and T is absolute temperature, K.

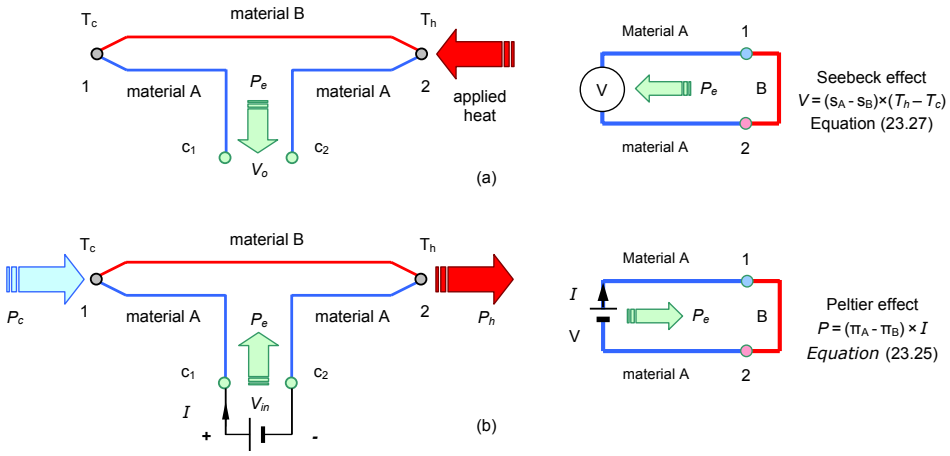


Figure 23.55. *The (a) Seebeck and (b) Peltier effects.*

The emf potential difference for the two junctions, with Seebeck coefficients s_n and s_p for n-type and p-type materials respectively, is

$$\Delta V = \int_{T_e}^{T_a} (s_p - s_n)\, dt = s \times (T_a - T_e) = s \times \Delta T \qquad (23.27)$$

where $T_{a/e}$ is the temperature of the absorbing/emitting junctions. The Seebeck coefficient is sometimes called the thermal emf coefficient or thermoelectric power.

When an electric current I is passed through a conductor having a temperature gradient over its length dT/dx, heat will be either absorbed by or dissipated from the conductor. Whether heat is absorbed or dissipated depends upon the direction of both the electric current and temperature gradient. This phenomenon is known as the Thomson effect, τ, (which is insignificant in TE modules) and is described by

$$\frac{dP}{dx} = \tau I \frac{dT}{dx}$$

where dP/dx is the rate of the heating per unit length, W/m.

Table 23.25: Seebeck coefficients for Standard Thermocouples. The absolute value for Pt (specifically, 5 ÂµV/K) should be subtracted from each value listed.

type	couples	Seebeck coefficient, *s*
		µV/K
E	Chromel - Constantan	60
J	Iron - Constantan	51
T	Copper - Constantan	40
K	Chromel -A	40
N	Nicrosil - Nisil	38
S	Pt (10% Rh) - Pt	11
B	Pt (30% Rh) – Pt (6% Rh)	8
R	Pt (13% Rh) - Pt	12

Table 23.26: Peltier factors, π, for different pairs of metals

Fe-constantan		Cu-Ni		Pb-constantan	
T	π	T	π	T	π
K	mV	K	mV	K	mV
273	13.0	292	8.0	293	8.7
299	15.0	328	9.0	383	11.8
403	19.0	478	10.3	508	16.0
513	26.0	563	8.6	578	18.7
593	34.0	613	8.0	633	20.6
833	52.0	718	10.0	713	23.4

The irreversible thermodynamic theory relating the three coefficients π, s, and τ is described by

$$\pi = sT$$

$$\tau = T\frac{\delta s}{\delta T} \qquad \left(\mathrm{V/K^2}\right) \tag{23.28}$$

The Peltier, Seebeck, and Thomson effects are reversible (in fact, the Peltier and Seebeck effects are reversals of one another); Joule heating and heat conduction are not reversible in complying with the laws of thermodynamics.

Thermoelectric modules for power semiconductor cooling are solid-state heat pumps that operate on the Peltier effect. A thermoelectric module consists of an array of p and n type narrow band gap, semiconductor element pairs that are heavily doped with electrical carriers. The elements are arranged into an array that is electrically connected in series but thermally connected in parallel. This array is then rigidly sandwiched between two thermally conducting and electrical insulating ceramic substrates, one on each side of the elements as seen in figures 23.56a and 23.57.

When current passes through the junction of two different types of conductors, a temperature change results at the junction. The practical application of this concept requires semiconductors that are good conductors of electricity but poor conductors of heat. Anisotropic orientated polycrystalline bismuth telluride is mainly used as the semiconductor material, heavily doped to create either an excess (n-type) or a deficiency (p-type) of electrons, as shown in figure 23.57.

When connected to a DC power source, current causes heat to move from one side of the TEC to the other. This creates a hot side and a cold side on the TEC. A typical cooling application exposes the cold side of the TEC to the object or substance to be cooled and the hot side to a heatsink which dissipates the heat to the environment. A heat exchanger with forced air or liquid is usually required. A thermoelectric cooler does not absorb heat, it only transfers or moves it.

If the current is reversed, the heat is moved in the opposite direction. In other words, what was the hot face will become the cold face and vice-versa.

Unlike a heat pipe, Peltier elements consume energy or produces heat. It is therefore possible to cool a device below ambient temperature, unlike a heat pipe. A heat exchanger with forced air or liquid may be required for both cooling mechanisms.

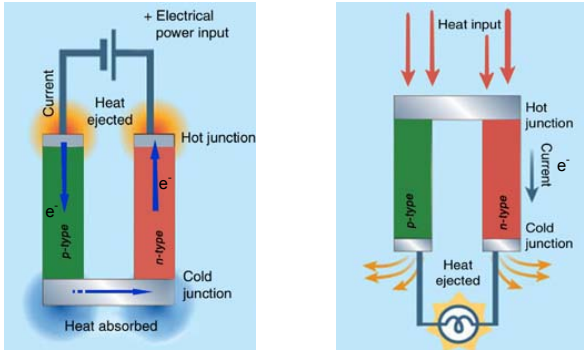


Figure 23.56. *Schematic representations of a pn-couple used as: (a) a thermoelectric cooler TEC based on the Peltier effect or (b) a thermoelectric generator TEG based on the Seebeck effect.*

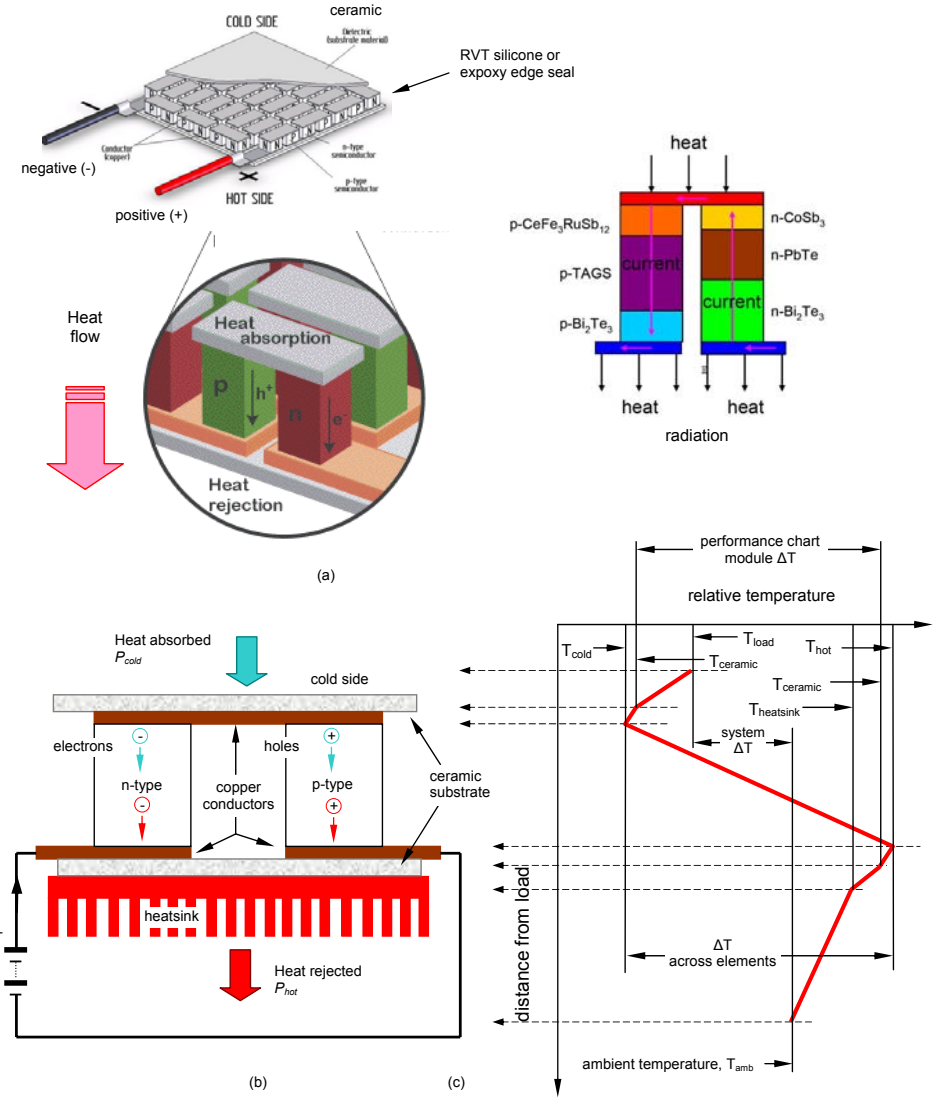


Figure 23.57. *The thermoelectric cooler: (a) module; (b) Peltier principle; and (c) temperature profile.*

The heat transfer occurs as electrons flow through one pair of p and n type elements (often referred to as a *couple*) within the thermoelectric module:

The p-type semiconductor is doped with specific atoms that have fewer electrons than necessary to complete the atomic bonds within the crystal lattice. When a voltage is applied, there is a tendency for conduction electrons to complete the atomic bonds. When conduction electrons do this, they leave 'holes' which essentially are atoms within the crystal lattice that now have local positive charges. Electrons are continually filling and being released from the holes and move on to the next available hole. In effect, the holes act as the electrical positive carriers.

Electrons move much more easily in the copper interconnect conductors but not so easily in the semiconductors. When electrons leave the p-type and enter the copper on the cold-side, holes are created in the p-type as the electrons jump to a higher energy level to match the energy level of the electrons already moving in the copper. The extra energy to create these holes comes by absorbing heat. Meanwhile, the newly created holes travel to the copper on the hot side. Electrons from the hot-side copper move into the p-type and fill the holes, releasing excess energy in the form of heat.

The n-type semiconductor is doped with atoms that provide more electrons than necessary to complete the atomic bonds within the crystal lattice. When a voltage is applied, these excess electrons are easily moved into the conduction band. However, additional energy is required to enable the n-type electrons to match the energy level of the incoming electrons from the cold-side copper. The needed energy is gained by absorbing heat. Finally, when the electrons leave the hot-side of the n-type, they once again move freely in the copper. They fall to a lower energy level, and release heat in the process.
The key fact is that heat is always absorbed at the cold side of the n and p type elements, and heat is always released at the hot side of thermoelectric element. The heat pumping capacity of a module is proportional to the current (equation (23.25)) and is dependent on the element geometry, number of couples, and material properties.

When a Peltier cell cooling system with starts up, both cold and hot sides are at the same temperature, so ΔT is zero and the cell develops its maximum power. As the temperature difference across the cell increases, the cooling power decreases and temperatures stabilize. Targeted applications mainly operate at around room temperature.
Since Peltier elements are active heat pumps, they can be used to cool components below ambient temperature - which is not possible using conventional cooling, or even heat pipes.

23.11.2 Thermoelectric materials

The thermoelectric semiconductor material most often used in TE coolers is an alloy of Bismuth Telluride that has been suitably doped to provide individual blocks or elements having distinct n and p characteristics.

i. n-type: bismuth-telluride-selenium (Bi_2Te_3 / Sb_2Se_3) compound
Seebeck's coefficient at 300K: $s \geq (140 \text{ to } 180) \times 10^{-6}$ V/K
Specific conductivity at 300K: $\sigma = 1/\rho = (1300 \text{ to } 1700)$ $Ohm^{-1}.cm^{-1}$

ii. p-type: bismuth-telluride-antimony (Bi_2Te_3 / Sb_2Te_3) compound
Seebeck's coefficient at 300K: $s \geq (130 \text{ to } 170) \times 10^{-6}$ V/K
Specific conductivity at 300K: $\sigma = 1/\rho = (1400 \text{ to } 1800)$ $Ohm^{-1}.cm^{-1}$

Seebeck coefficient values for different TE semiconductor materials are shown in Table 23.27.

Bismuth Telluride is a thermoelectric semiconductor material that exhibits optimum performance in a 'room temperature' range. Thermoelectric materials most often are fabricated by either directional crystallization from a melt or pressed powder metallurgy. Each manufacturing method has its own particular advantage, but directionally grown materials are most common.
In addition to Bismuth Telluride (Bi_2Te_3) and Lead Telluride (PbTe), there are other thermoelectric materials including, Silicon Germanium (SiGe), and Bismuth-Antimony (Bi-Sb) alloys that may be used in specific situations. Bi-Sb alloys perform better than Bi_2Te_3 at low temperatures, <200K, while Lead Telluride (PbTe) based thermoelectric materials are typically employed for operational temperature around 500K to 800K. Te-Ag-Ge-Sb materials have been developed to replace p-type PbTe. Silicon Germanium (SiGe) alloys, with boron and phosphorous (or arsenic) as p-type and n-type doping agent can be used up to 1300K, with a Figure of Merit Z approaching unity.

Figure 23.58 illustrates the relative performance or Figure-of-Merit of various materials over a range of temperatures. The performance of Bismuth Telluride peaks within a temperature range that is best suited for most room temperature cooling applications.

Ceramic TE components are made of beryllium oxide, Be0, which is typically used in multi-stage coolers due to its higher thermal conductivity. The advantages of this material are that it enhances the thermal performance of the TE Cooler because of the high heat conductance as well as makes it easier to assemble. Disadvantages are that it is more expensive and can be toxic when its processing dust is inhaled. Aluminium oxide Al_2O_3 and particularly, aluminium nitride AlN, are viable ceramic alternatives, with properties as shown in Table 23.28.

Table 23.27: Seebeck coefficients for some metals and alloys, compared to Platinum

metals	Seebeck coefficient	semiconductors	Seebeck coefficient
	s		*s*
	μV / K		μV / K
antimony	47	Se	900
nichrome	25	Te	500
molybdenum	10	Si	440
cadmium	7.5	Ge	300
tungsten	7.5	n-type Bi_2Te_3	-230
gold	6.5	p-type $Bi_{2-x}Sb_xTe_3$	300
silver	6.5	p-type Sb_2Te_3	185
copper	6.5	PbTe	-180
lead	4.0	$Pb_{03}Ge_{39}Se_{58}$	1670
aluminium	3.5	$Pb_{06}Ge_{36}Se_{58}$	1410
carbon	3.0	$Pb_{13}Ge_{29}Se_{58}$	-1710
mercury	0.6	$Pb_{15}Ge_{37}Se_{58}$	-1990
platinum	0	$SnSb_4Te_7$	25
sodium	-2.0	$SnBi_4Te_7$	120
potassium	-9.0	$SnBi_3Sb_1Te_7$	151
nickel	-15	$SnBi_{2.5}Sb_{1.5}Te_7$	110
constantan	-35	$SnBi_2Sb_2Te_7$	90
bismuth	-72	$PbBi_4Te_7$	-53

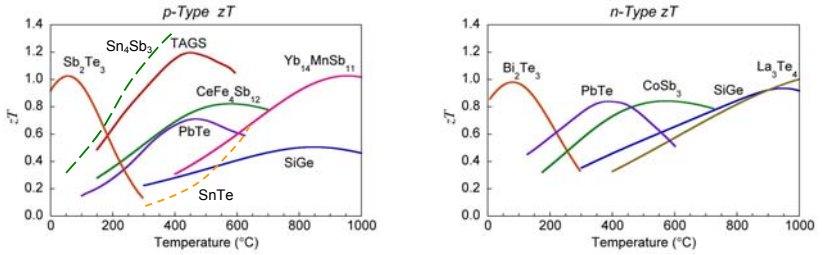


Figure 23.58. *Performance of thermoelectric materials at various temperatures*. **TAGS (Te-Ag-Ge-Sb)**

The most common TEC shape is a square or a rectangular substrate device. The practical size of a single stage TEC ranges from 3mm x 3mm up to 60mm x 60mm. A size limitation of 60mm x 60mm is due to the thermal stress from thermal expansion deformations between the cold and the hot junctions of the TEC. To obtain a larger temperature difference (but not higher cooling powers), a multistage cascaded TEC can be built, where 6 stages is the maximum practical limit.

Table 23.28: TEC/TEG substrate properties (also see tables 5.27 and 5.30)

Materials	Density	Thermal Conductivity	Coeff. of thermal expansion	Specific heat	Electric Resistivity 25°C	Dielectric strength ac	Dielectric loss 25°C@1MHz	Dielectric Constant	Poisson's Ratio
	g/cm^3	W/m.K	10^{-6}/K	J/kg.K	Ohm-m	kV/mm	x10^{-4}	@ 1 MHz	
Alumina 99.5% $A\ell_2 0_3$	3.89	35	8.4	880	> 10^{12}	8.7	1	9.8	0.22
Beryllia Be0	2.85	248	7.2	1260	10^{15}			6.5	0.38
Aluminium Nitride AℓN	3.26	140 - 180	4.5	740	> 10^{12}	15	1	9	0.24

23.11.3 Mathematical equation for a thermoelectric module

A thermoelectric couple is shown in figure 23.59, on which the following physical and electrical parameters are defined.

L = element height, m
A = cross-sectional area, m^2
G = Area / Length (inverse of aspect ratio) of TE Element, m
P_c = heat load on the cold side (into and pumped by the TEC), W
P_h = heat released from the hot side (pumped from the TEC), W
P_{tec} = electrical input power (heat load) to the thermoelectric, W
T_c = cold-side temperature, K
T_h = hot-side temperature, K
$\Delta T = T_h - T_c$, K
$T_{ave} = ½\times(T_c + T_h)$, K
I_{te} = applied current, A

Additionally
s = Seebeck coefficient, V / K
ρ = electrical resistivity, Ωm
κ = thermal conductivity, W/m K
V = input dc voltage, V
N = number of thermocouples, element pairs

CoP	Coefficient of Performance ($P_c / I V$)
Z	Figure of Merit ($s^2 / \rho\times\kappa$), 1/K
S	Device Seebeck Voltage ($2N\times s$), V / K
R	Device Electrical Resistance ($2N\times\rho / G = 2N\times\rho\times L/A$), Ohms
K	Device Thermal Conductance ($2N\times\kappa\times G = 2N\times\rho\times A/L$), W / K

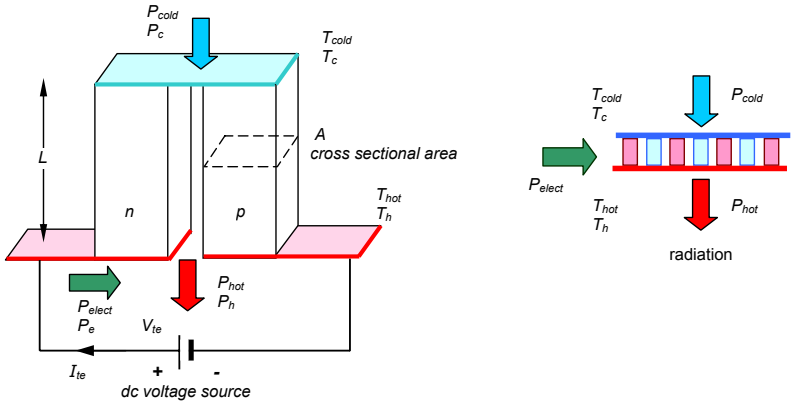


Figure 23.59. *The thermoelectric couple based on the Peltier effect.*

The heat pumped at the cold surface P_c is

$$P_c = 2\times N\times\left[sIT_c - ½I^2\frac{\rho}{G} - \kappa G\Delta T\right] \quad (\text{W}) \qquad (23.29)$$

- The first P_c term, $s\times I\times T_c$, is the Peltier cooling effect.
- The second term, $½\times I^2\times\rho\times L/A$, represents the Joule heating effect ($I^2\times R$) associated with passing an electrical current through a resistance. The Joule heat is distributed throughout the element, so that half the heat goes towards the cold side, and the remaining half of the heat goes towards the hot side.
- The last term, $\kappa\times A/L\times(T_h - T_c)$, represents the Fourier effect in which heat conducts from a higher temperature to a lower temperature.

Peltier cooling is reduced by the losses associated with electrical resistance and thermal conductance.
The heat dissipated from the hot side P_h is

$$P_h = 2\times N\times\left[sIT_h + ½I^2\frac{\rho}{G} - \kappa G\Delta T\right] \quad (\text{W}) \qquad (23.30)$$

The electrical input power applied to the TEC, P_e, is the difference between the hot and cold powers, that is

$$P_e = P_h - P_c = 2\times N\times\left[sIT_h + I^2\frac{\rho}{G}\right] = V\times I \quad (\text{W}) \qquad (23.31)$$

The input terminal voltage to the module is therefore P_e / I

$$V = 2\times N\times\left[s\times\Delta T + I\times\frac{\rho}{G}\right] \quad (\text{V}) \qquad (23.32)$$

In the voltage equation, the first term, $s\times(T_h - T_c)$ represents the Seebeck voltage. The second term, $I\times\rho\times L/A$ represents the electrical voltage related by Ohm's law.

The heat rejected by the module P_{hot} is

$$P_{hot} = P_{cold} + P_{tec} \quad (\text{W}) \qquad (23.33)$$

The *Coefficient of Performance*, *CoP*, is useful in selecting a module (the larger *CoP* the better), and is defined by

$$CoP = \frac{\text{heat absorbed at the cold junction}}{\text{electrical dc input power}} = \frac{P_{cold}}{V_{te}\times I_{te}} = \frac{P_{cold}}{P_{tec}} \qquad (23.34)$$

where I_{te} is the current drawn by the TE module
V_{te} is the voltage applied to the TE module
A maximum *CoP* represents the minimum input dc power P_{tec}, therefore minimum total heat to be rejected by the heat exchanger on the hot side P_{hot}, that is $P_{hot} = P_{cold} + P_e$.

A physical properties materials figure of merit is

$$Z = \frac{s^2}{\rho\times\kappa} \quad (\text{K}^{-1}) \qquad (23.35)$$

The figure of merit Z is directly related to TEC module ability to pump heat. This equation shows why it is difficult to obtain good thermoelectric materials. A good thermoelectric material must achieve low thermal conductivity κ (to prevent heat losses through heat conduction between the hot and cold side) and a high electrical conductivity $\sigma = 1/\rho$ (to minimize Joule heating).

Equations (23.29), (23.32), and (23.35) can be simplified by substituting $S = 2N\times s$, $R = 2N\times\rho / G$ and $K = 2N\times\kappa\times G$, as appropriate.

$$P_c = SIT_c - ½I^2R - K\times\Delta T \quad (\text{W}) \qquad (23.36)$$

$$V = S\times\Delta T + I\times R \quad (\text{V}) \qquad (23.37)$$

$$Z = \frac{S^2}{R\times K} \quad (\text{K}^{-1}) \qquad (23.38)$$

S, R and K are not usually given by manufacturers. Rather they are derived from the given parameters I_{opt}, V_{opt}, P_{max}, and T_{max}, at a specified T_h.

Expressions for I_{opt}, V_{opt}, P_{max}, and T_{max}
The maximum heat pumping rate is derived from the differential of equation (23.36) with respect to I at zero P_c, which gives an optimum current of

$$I_{opt} = \frac{S}{R}\times T_c = \frac{S}{R}\times(T_h - \Delta T_{max}) \qquad (23.39)$$

Substituting I_{opt} in to equation (23.36) gives the maximum heat pumping rate, $P_{c,max}$

$$P_{c,\max} = ½\frac{S^2 \times T_c^2}{R} - K\Delta T \qquad (\mathrm{W}) \qquad (23.40)$$

The peak maximum is when $\Delta T = 0$, that is

$$\hat{P}_{c,\max} = ½\frac{S^2 \times T_c^2}{R} \qquad (23.41)$$

The corresponding *CoP* is

$$CoP_{opt} = \frac{½Z T_c^2 - \Delta T}{Z T_h T_c} \qquad (23.42)$$

and from equation (23.38)

$$ZT = \frac{S^2 T}{R \times K} \qquad (23.43)$$

From equation (23.40) the largest temperature difference occurs when $P_{c,max}$ is zero, whence the maximum temperature difference is

$$\Delta T_{\max} = ½\frac{S^2}{K \times R} \times T_c^2 = ½\, Z \times T_c^2 = ½\, Z \times \left(T_h - \Delta T_{\max}\right)^2 \qquad (23.44)$$

or in terms of only the hot side temperature, ΔT_{max} from equation (23.44) and I_{opt} from equation (23.39) become

$$\Delta T_{\max} = T_h - \frac{\left(1 + 2 \times Z \times T_h\right)^{½} - 1}{Z} \qquad (23.45)$$

$$I_{opt} = \frac{\kappa \times G}{s} \times \left[\left(1 + 2 \times Z \times T_h\right)^{½} - 1\right] = \frac{K}{S} \times \left[\left(1 + 2 \times Z \times T_h\right)^{½} - 1\right] \qquad (23.46)$$

From equation (23.37) the optimum voltage for a maximum temperature differential is therefore

$$V_{opt} = S \times \Delta T_{\max} + I_{opt} \times R \qquad (\mathrm{V}) \qquad (23.47)$$

Calculation of *Z*, *S*, *K,* and *R* from data sheet information:

$$\begin{aligned} Z &= \frac{2\Delta T_{\max}}{\left(T_h - \Delta T_{\max}\right)^2} \\ S &= \frac{2 \times V_{opt}}{T_h} = \frac{2 \times P_{\max}}{\left(T_h - \Delta T_{\max}\right) \times I_{opt}} \\ K &= \frac{\left(T_h - \Delta T_{\max}\right) V_{opt} I_{opt}}{2 T_h \Delta T_{\max}} = \frac{\left(T_h - \Delta T_{\max}\right) P_{\max}}{\left(T_h + \Delta T_{\max}\right) \Delta T_{\max}} \\ R &= \frac{\left(T_h - \Delta T_{\max}\right) \times V_{opt}}{T_h \times I_{opt}} = \frac{S^2}{K \times Z} \end{aligned} \qquad (23.48)$$

If *N* and *G* are known, *s*, *ρ* and *κ* can be calculated.

From equations (23.36) and (23.37) the *CoP* , equation (23.34), is now given by

$$CoP = \frac{P_c}{P_e} = \frac{SIT_c - K\Delta T - ½I^2 R}{SI\Delta T + I^2 R} \qquad (23.49)$$

Differentiation of the *CoP* equation with respect to the current term, *dCoP/dI* = 0, gives:
Optimum Current:

$$I_{CoP,opt} = \frac{S\Delta T}{R\left(1 + Z \times T_{ave}\right)^{½} - 1} \qquad \frac{\kappa \times G}{s} \times \frac{\Delta T}{T_{ave}} \times \left[\left(1 + 2 \times Z \times T_h\right)^{½} + 1\right] \qquad (23.50)$$

The corresponding maximum *COP*, (calculated at I_{opt}):

$$COP_{\max} = \frac{P_c}{VI} = \frac{T_c}{\Delta T} \times \frac{\left(1 + Z \times T_{ave}\right)^{½} - \frac{T_h}{T_c}}{\left(1 + Z \times T_{ave}\right)^{½} + 1} \qquad \frac{T_{ave}}{\Delta T} \times \frac{\left(1 + Z \times T_{ave}\right)^{½} - 1}{\left(1 + Z \times T_{ave}\right)^{½} + 1} - ½ \qquad (23.51)$$

For a given thermoelectric material, equation (23.51) can be used to plot the maximum *CoP* as a function of the thermoelectric element ΔT (T_h - T_c). This is illustrated in figure 5.59a. for Bismuth Telluride (Bi_2Te_3) thermoelectric materials. COPs in excess of one are possible for ΔT's less than about 30°C. The *COP* increases significantly for systems designed to run optimally at even lower ΔT's.

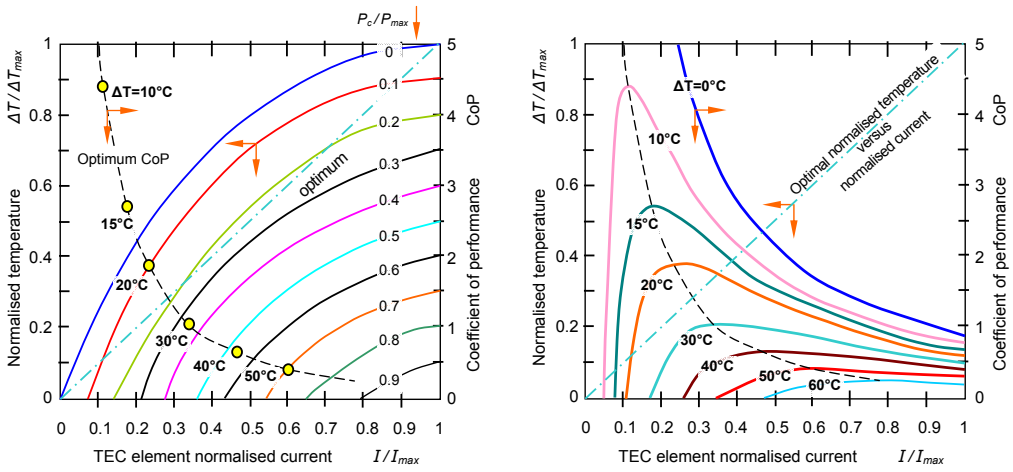


Figure 23.60. *TEC:*
(a) generic performance curve and (b) theoretical optimum (maximum) COP as a function of current.

Operation at near rated maximum current potentially results in significantly degraded performance as illustrated in figure 23.60b, which shows a plot of COP as a function of the percentage of maximum rated current (I/I_{max}).

The peak COP for each of the curves coincides with the single curve of figure 5.59a. Each curve represents a TEC optimized for that particular ΔT and then operated at various different currents from 0 to I_{max}. For a given curve, the CoP decreases significantly as the TEC is operated 'off-optimum', especially for ΔT's less than 30°C.
Figure 23.60b shows that the current required to produce high CoPs is a low percentage of the TEC rated maximum current (10 to 15% of I_{max}). In other words, the TEC must pump the desired heat while 'idling' at a low percentage of I_{max} in order to achieve the high CoPs.

The *CoP* of an n-stage, $n > 1$, thermoelectric module can be expressed as

$$CoP_n = \frac{1}{\left(1 + \frac{1}{n \times \left(CoP_1 + ½\right) - ½}\right)^n - 1} \qquad (23.52)$$

assuming that each stage operates over a temperature difference of $\Delta T/n$ and CoP_1 is applicable to a single stage TEC that operates over ΔT. For example, the *CoP* of a two-identical-stage module is

$$CoP_2 = CoP_1 + \frac{1}{8 \times \left(2 \times CoP_1 + 1\right)}$$

The presented equations are simplified but show the basic idea behind the calculations involved. The actual differential equations do not have a closed-form solution because *S*, *R*, and *K* are temperature dependent. Assuming constant properties can lead to significant errors.

23.11.4 Features of Thermoelectric Cooling - Peltier elements

The use of thermoelectric modules often provides solutions, and in some cases the only solution, to many difficult thermal management problems where a low to moderate amount of heat must be handled. While no one cooling method is ideal in all respects and the use of thermoelectric modules will not be suitable for every application, TE coolers will often provide substantial advantages over alternative technologies. TE Coolers typically have a *CoP* of approximately 2, which is lower than the *CoP* of 3 to 5 of vapour compression refrigerators.
The lowest practically achievable temperature is about -100 ºC, since the efficiency of thermoelectric modules decreases considerably at very low temperatures. The highest practical temperature limit is about 80ºC, which is imposed mainly by the manufacturing techniques used to assemble thermoelectric modules.

Some of the more significant features of thermoelectric modules include:
Compared with standard designs using refrigeration cycles with compressors and cooling mediums (such as CFC's), thermoelectric cooling possesses the following traits:

- There is no environmental damage since no cooling medium such as CFCs or any gas is used. Environmentally friendly and safe. No coolant gas, corrosive gas or fluid leakage, easy maintenance.
- Small compact size - dimensions (small form factor) and lightweight, giving high cooling density.
- Wide choice of configuration – flexible form.
- By simply changing the applied dc current polarity, heating is possible in addition to cooling with the same module (including temperature cycling). Easy switching from cooling to heating mode. Since cooling and heating are both possible, it is also possible to regulate temperatures close to room temperature. Precise temperature control to within ±0.1°C, with smooth and fine adjustment of cooling capacity and temperature.
- DC operation, with high power efficiency.
- Good responsiveness to heat. (Quickly heats or cools.) Quick cooling to below ambient - economically. Wide operating temperature range, sub-ambient cooling, cooling to low temperatures, below ambient, multistage cascades to below -100°C, wide range of operating temperatures.
- Solid-state device, therefore no moving parts; there is no vibration or noise. Acoustically silent and electrically 'quiet'.
- Since there are no fatiguing or breakable machine parts it is the most long-lasting, highly reliable method of cooling. Maintenance-free, >200,000hr.
- With only a power cord, it is easy to handle.
- Spot Cooling: It is possible to cool one specific component or area only, thereby often making it unnecessary to cool an entire package or enclosure.
- Operation in any orientation or any spatial position, zero gravity and high G levels, resistance to high mechanical loads, shocks and vibration.

Problems related to Peltier cooling

- High power usage and high power dissipation are the biggest problems related to Peltier cooling. Large power dissipation requires a large heatsink, powerful (and thus loud) fans, and a dc supply.
- Limited to power dissipation of the order of 600W.
- Low temperatures may cause moisture condensation, leading to short circuits between the elements. Moisture condensation depends on the temperature inside the system block, the temperature of the cooled device and air moisture. The warmer the air and the higher the moisture, the more probable condensation occurs.
- Low efficiency, or low *CoP* compared with compressor-based cooling.
- Failure modes are thermal cycle fatigue of solder to chip, copper corrosion, copper migration, and crystal inclined cleavage planed defects.

23.11.5 TE cooling design

TEC design involves the initial specification of three parameters, the hot and cold side temperatures, T_{hot} and T_{cold}, (or T_h and T_c) hence the temperature gradient or difference $\Delta T = T_h - T_c$, ($\Delta T > 0$) and the amount of heat, in Watts, to be absorbed at the cold surface of the TEC, P_{cold}.
The cold surface temperature, T_c, is the desired temperature of the object to be cooled, directly in contact with the TEC.

The hot surface temperature, T_h, is defined by two major parameters:

- The temperature of the ambient environment to which the heat is being rejected.
- The efficiency of the heat exchanger that is between the hot surface of the TEC and the ambient.

The third parameter required is the amount of heat, the thermal load, P_{cold}, to be removed or absorbed by the cold surface of the TEC. The thermal load includes the active I^2R type losses of the device to be cooled, as well as parasitic loads such as conduction through any mechanical object in contact with both the cold surface and any warmer environment, like conduction through mounting bolts and plates (and the radiation from the plates). Figure 23.61 show the thermal resistance components and system model. Performance characteristic charts, as in figure 23.62, are usually provided. These allowing the terminal dc voltage and dc current requirements to be determined from the temperature difference ΔT and heat to be absorbed on the cold side, P_{cold}. The maximum ΔT is about 67°C for a single TEC, higher than this requires cascading (stacking) of TECs. The negative quadratic shape in the lower plot on figure 5.57, represents the optimal operating curve. Further TE technical details can be found in Chapter 5.10.

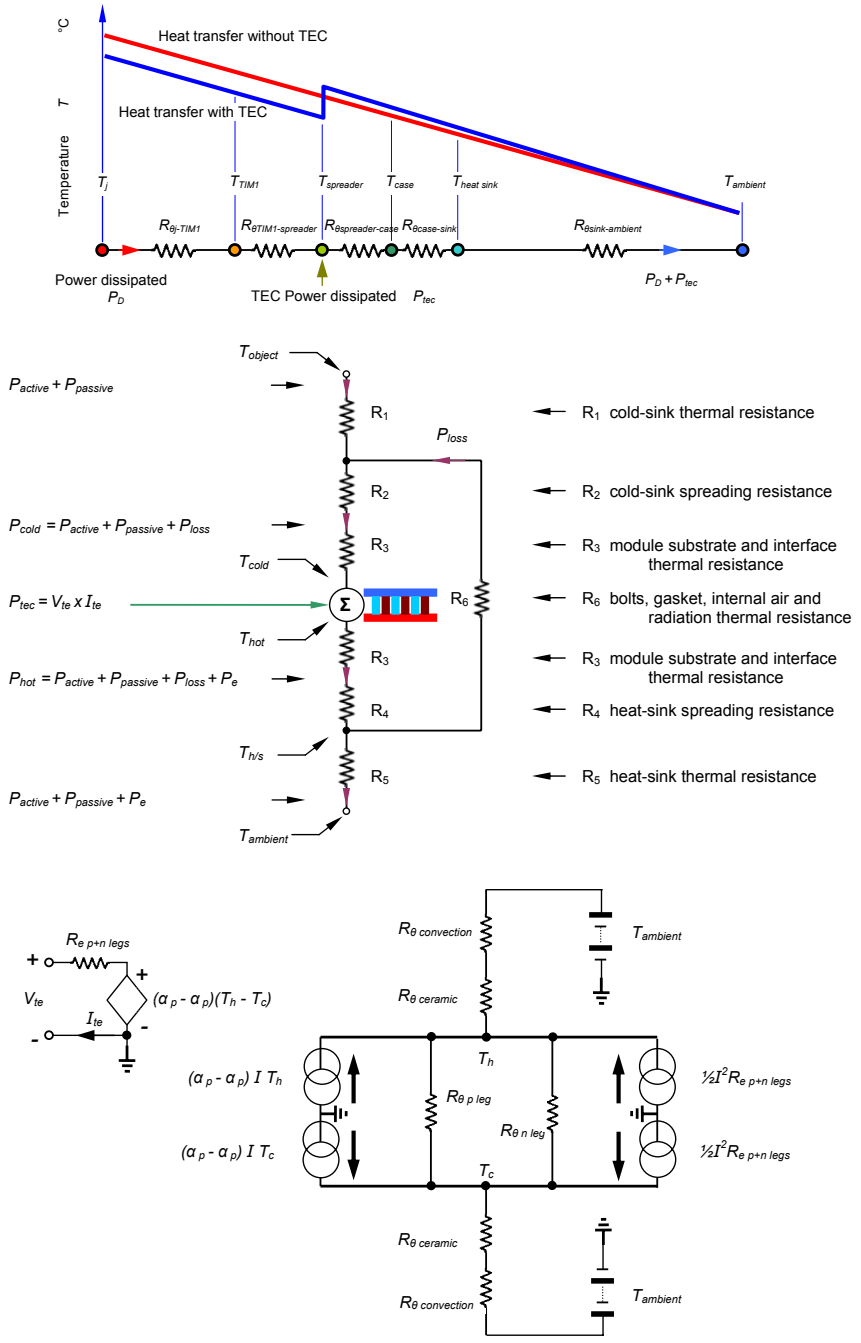


Figure 23.61. *TEC cooling models:*
(a) thermal resistance model; (b) equivalent electrical circuit model; and (c) PSpice electrical model.

Example 23.8: ***Thermoelectric cooler design***

An application has an estimated heat load of 22W, a forced convection type heat sink with a thermal resistance of 0.15°C/W, an ambient temperature of 25°C, and an object that needs to be cooled to 5°C. The cold side of the thermoelectric cooler is in direct contact with the object to be cooled.
The specifications for the thermoelectric module in figure 23.62 are as follows (at T_h = 25°C):

I_{max} = 6.0A
Q_{max} = 51.4W
V_{max} = 15.4V
ΔT_{max} = 67°C

Solution

To determine if the thermoelectric cooler is appropriate for this application, it must be shown that the parameters ΔT and Q_c are within the appropriate boundaries of the performance curves.
The parameter ΔT follows directly from T_h and T_c. Since the cold side of the thermoelectric is in direct contact with the object being cooled, T_c is estimated to be 5°C. Assuming a 10°C rise above ambient for the forced convection type heat sink, T_h is estimated to be 35°C. Without knowing the power into the thermoelectric cooler, an exact value of T_h cannot be found initially. The temperature difference across the thermoelectric cooler is:

$$\Delta T = T_h - T_c = 35°C - 5°C = 30°C$$

Figure 23.62 shows performance curves for the TEC at a hot side temperature of 35°C. Referring to figure 23.62b, the intersection of P_c = 22W and ΔT = 30°C show that this thermoelectric can pump 22W of heat at a ΔT of 30°C with an input current of 3.6A.

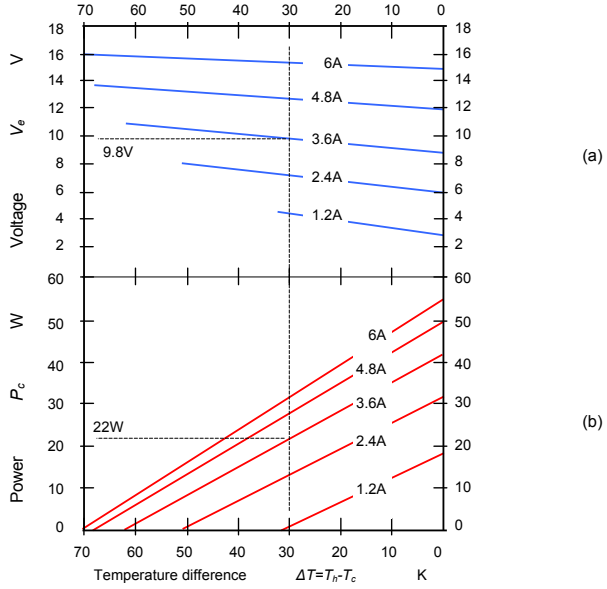


Figure 23.62. *TEC performance curves:*
(a) upper plot ΔT versus Voltage and (b) lower plot, ΔT versus P_c.

These values are based on the estimate T_h = 35°C. Once the power into the TEC is determined, the equations to follow can be used to solve for T_h and to determine whether the original estimate of T_h was appropriate.

The input power to the thermoelectric, P_{in}, is the product of the current and the voltage. Using the 3.6A line in Figure 23.62a for the current, the input voltage corresponding to ΔT = 30°C is almost 10V.
T_h can now be calculated from:

$$P_h = P_c + P_{tec} = 22\text{ W} + 3.6\text{ A} \times 10\text{ V} = 22\text{ W} + 36\text{ W} = 58\text{ W}$$

Therefore, using

$$T_h = T_{amb} + R_\theta \times P_h \quad \text{where } T_{amb} = 25°C \text{ and } R_\theta = 0.15°C/W$$
$$T_h = 25°C + 0.15°C/W \times 58\text{ W}$$
$$= 25°C + 8.7°C = 33.7°C$$

The calculated T_h (33.7°C) is close enough to the original estimate of T_h (35°C), to conclude that the thermoelectric will work in the given application. If an exact solution is required, the process of solving for T_h mathematically can be repeated until the value of T_h does not change.
The TEC operating coefficient of performance *CoP* is

$$CoP = \frac{P_{cold}}{P_{tec}} = \frac{22\text{W}}{36\text{W}} = 0.61$$

♣

23.11.6 Thermoelectric power generation

A thermoelectric generator, TEG, is an energy conversion system that converts thermal energy to electrical energy. The fundamental physics of this type of energy conversion is summarize as, the temperature difference ΔT between the hot T_h and cold T_c sources leads to a difference in the Fermi energy ΔE_F across the thermoelectric material yielding a potential difference, which drives a current.
Bismuth Telluride-based thermoelectric modules are designed primarily for cooling or combined cooling and heating applications where electrical power creates a temperature difference across the module. By using the modules 'in reverse', however, whereby a temperature differential is applied across the faces of the module, it is possible to generate electrical power. Although power output and generation efficiency are low, useful power may be obtained where a source of heat is available.
A thermoelectric module used for power generation has certain similarities to a conventional thermocouple. A single thermoelectric couple or generator with an applied temperature difference is shown in Figure 23.63.

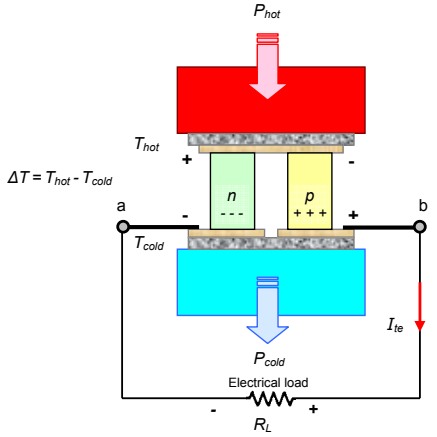


Figure 23.63. *Single thermoelectric couple where $T_{hot} > T_{cold}$.*

With no load (R_L not connected), the open circuit voltage as measured between points 'a' and 'b' is:

$$V_{o/c} = s \times \Delta T$$

where: $V_{o/c}$ is the output voltage from the couple (generator), V
s is the average Seebeck coefficient, V/K
ΔT is the temperature difference across the couple where $\Delta T = T_h - T_c$., K
When a load is connected to the thermoelectric couple, the output voltage V_o drops as a result of internal couple thermoelectric materials resistance.

$$V_o = s \times \Delta T - I_{te} \times R_c \qquad (23.53)$$

The current through the load is:

$$I_{te} = \frac{S \times \Delta T}{R_c + R_L} \tag{23.54}$$

where: I_{te} is the couple output current, A
R_c is the average internal resistance of the thermoelectric couple, Ω
R_L is the load resistance, Ω

The output power, in terms of the input heat P_T and efficiency η, is given by

$$P_{tec} = I_{te}V_o = \eta P_T$$

Maximum efficiency is reached when the load and internal resistances are equal because this is the maximum power achieved from load matching.

The total heat input to the couple, P_h, is:

$$P_h = S \times T_h \times I_{te} - ½ \times I_{te}^2 \times R_c + \kappa \times \Delta T \tag{23.55}$$

where: P_h is the heat input, W
κ is the thermal conductance of the couple, W/K
T_h is the hot side of the couple, K

The efficiency of the couple η, is:

$$\eta = \frac{V_o \times I_{te}}{P_h} \tag{23.56}$$

A complete module consists of a number of couples, it is therefore necessary to rewrite the various equations for an actual module, as follows:

$$V_{o/c} = S \times \Delta T = I_{te} \times (R + R_L) \tag{23.57}$$

where: $V_{o/c}$ is the generators open circuit output voltage, V
S is the module's average Seebeck coefficient, V/K
R is the module's average resistance, Ω

The module Seebeck coefficient, resistance, and thermal conductance properties are temperature dependent. The values of S, R, and K must be selected at the average module temperature T_{ave} where:

$$T_{ave} = ½(T_h + T_c) \tag{23.58}$$

The power output, P_o, from the module in watts is:

$$P_o = R_L \times I_{te}^2 \tag{23.59}$$

The voltage at maximum power is half the open circuit voltage $V_{o/c}$ ($V_{o/c} = SxΔT = 2V_{max}$) and the maximum power changes with temperature difference as a function of ΔT^2.

$$P_{max} = \frac{V_{max}^2}{R} = \frac{V_{o/c}^2}{4R} = \frac{S^2 \times \Delta T^2}{4R} \tag{23.60}$$

The maximum efficiency of a thermoelectric material depends on two terms. The first is the Carnot efficiency: no heat engine can exceed the Carnot efficiency. The second is a term that depends on the thermoelectric properties, Seebeck coefficient, electrical resistivity and thermal conductivity, which together form a material property called zT, which is a good approximation for the thermoelectric module Figure of Merit ZT. For small temperature difference ($T_c \approx T_h$) this efficiency is given by:

$$\eta_{max} = \frac{\Delta T}{T_h} \times \frac{\sqrt{1+ZT}-1}{\sqrt{1+ZT}+\frac{T_c}{T_h}} \approx \frac{\Delta T}{T_h} \times \frac{\sqrt{1+ZT}-1}{\sqrt{1+ZT}+1} \tag{23.61}$$

where

$$ZT = \frac{S^2 T}{RK} = \text{module figure of merit} \approx zT = \frac{s^2 T}{\rho\kappa} = \text{material figure of merit} \tag{23.62}$$

and

$$\frac{\Delta T}{T_h} = \text{Carnot efficiency} \tag{23.63}$$

ZT (involving S, K, and R) refers to the TEC module while zT (involving s, κ, and ρ) refers to the raw material property. That is, S, K, and R involve TEC dimensions and the number of couples.

Most thermoelectric generators contain a number of individual modules which may be electrically connected in either a series, parallel, or series/parallel arrangement. A typical generator configuration is illustrated in figure 23.64. This generator has a total number of modules N_T with N_s modules connected in series and N_p modules connected in parallel. The total number of modules in the system is:

$$N_T = N_S \times N_p \tag{23.64}$$

The current I in amperes passing through the load resistance R_L is:

$$I = \frac{N_s \times S \times \Delta T}{\frac{N_s}{N_p} \times R + R_L} = \frac{V_{o/c}}{R_{gen} + R_L} \tag{23.65}$$

where the open circuit output voltage $V_{o/c}$ is:

$$V_{o/c} = N_s \times S\Delta T \tag{23.66}$$

and the generator Thevenin equivalent resistance R_{gen} is:

$$R_{gen} = \frac{N_s}{N_p} \times R \tag{23.67}$$

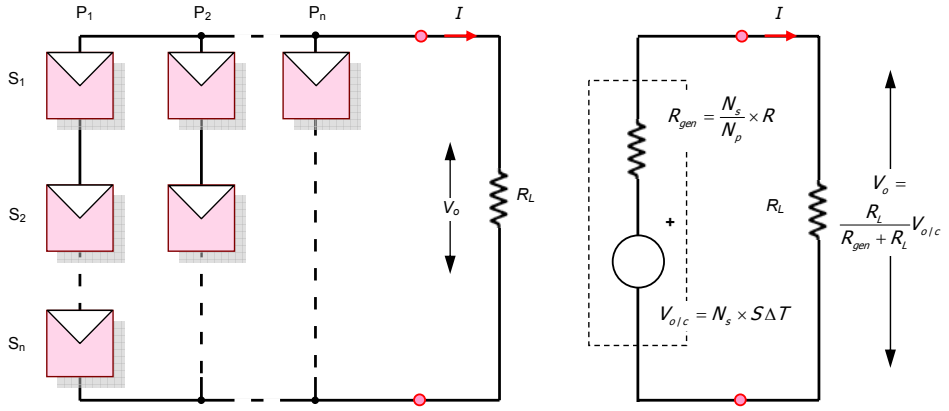


Figure 23.64. *Typical thermoelectric generator with a series-parallel arrangement of modules.*

The loaded output voltage V_o from the generator in volts is:

$$V_o = V_{o/c}\frac{R_L}{R_{gen}+R_L} = N_s \times S\Delta T \frac{R_L}{\frac{N_s}{N_p} \times R + R_L} \tag{23.68}$$

The output power P_o from the generator in watts is:

$$P_o = I^2 R_L = \left(\frac{N_s \times S \times \Delta T}{\frac{N_s}{N_p} \times R + R_L}\right)^2 R_L \tag{23.69}$$

The heat input P_h to each module in watts is:

$$P_h = S \times T_h \times \frac{I}{N_p} - ½ \times \left(\frac{I}{N_p}\right)^2 \times R + K \times \Delta T \tag{23.70}$$

The total heat input P_h to the generator in watts is:

$$\begin{aligned} P_H &= N_T \times \left(S \times T_h \times \frac{I}{N_p} - ½ \times \left(\frac{I}{N_p}\right)^2 \times R + K \times \Delta T\right) \\ &= N_s \times S \times T_h \times I - ½ \times I^2 \frac{N_s}{N_p} \times R + N_T \times K \times \Delta T \end{aligned} \tag{23.71}$$

The efficiency η of the generator is:

$$\eta = \frac{P_o}{P_H} \times 100\%$$

Maximum efficiency occurs when the generator internal resistance R_{gen} equals the load resistance R_L. That is, from equation (23.67), maximum output power occurs when

$$R_L = R_{gen} = \frac{N_s}{N_p} \times R \tag{23.72}$$

Assuming temperature independent material parameters, the maximum efficiency is approximately

$$\eta_{max} = \frac{\Delta T}{2 \times T_{ave} + ½\Delta T - \frac{4}{Z}} \tag{23.73}$$

Example 23.9: ***Thermoelectric generator design***

A 12V, 1.5A thermoelectric power generator is needed to power telemetry electronics at a remotely located oil pipeline where the hot, continuously flowing oil produces a 130°C pipe casing temperature. Flowing water (having a temperature of 10°C) also is available at the remote site, and an efficient water-cooled heat sink can maintain the TE generator cold-side at a temperature of +30°C.
For a 127-couple, 6 A module, the following values are applicable at T_{ave} = 353.2K:

S = 0.05544 V/K
R = 3.0994 Ω
K = 0.6632 W/K

Solution

The system parameters are:

T_h = +130°C ≡ 403.2K V_o = 12V
T_c = +30°C ≡ 303.2K I = 1.5A

therefore

T_{ave} = ½×(T_h+T_c) = ½×(403.2+303.2) = 353.2K
R_L = V_o / I_{te} = 12V / 1.5A = 8.0 Ω
P_o = V_o x I_{te} = 12 x 1.5 = 18 W
ΔT = T_h - T_c = 403.2 - 303.2 = 100K

It is usually desirable to select a relatively 'high power' thermoelectric module for generator applications in order to minimize the total system cost. Thus a 127 couple, 6A module is used in the design.
The required power P_o for the load is 12 x 1.5 = 18 W. The minimum number of modules needed to meet this load requirement is calculated from the maximum output power from one module, equation (23.60):

$$P_{\max} = \frac{(S \times \Delta T)^2}{4 \times R} = \frac{(0.05544 \times 100\text{K})^2}{4 \times 3.0994\Omega} = 2.48\text{W}$$

The minimum number of modules needed is:

$$N_{T\,\min} = \frac{P_o}{P_{\max}} = \frac{18}{2.48} = 7.3, \text{ use } 8$$

Because maximum generator efficiency occurs when R_{gen} = R_L = 8 Ω, it is desirable for most applications to select the series/parallel module configuration that will best approximate this resistance matching. One possible exception to the equalizing R_{gen} with R_L is when a relatively low current (in the mA range) and moderate voltage is required. In this case, the connection of all modules electrically in series may give better results. However, the maximum output voltage from the generator is obtained from a series-connected group of modules only when the resistance of the load is significantly higher than the internal resistance of the generator.
The string series-connected configuration of eight modules gives an internal resistance of:

$$R_{gen} = \frac{N_s}{N_p} \times R = \frac{8}{1} \times 3.0994\Omega = 24.8\Omega$$

This 24.8Ω generator resistance is considerably higher than the 8.0Ω load resistance, thereby indicating that a series module connection probably is not the best arrangement. For the all series case where N_s = 8 and N_p = 1, the open circuit voltage 8×0.05544×100K = 44.35V and the loaded output voltage, from equation (23.68), is:

$$V_o = N_s \times S\Delta T \times \frac{R_L}{\frac{N_s}{N_p} \times R + R_L}$$

$$= 8 \times 0.05544 \times 100\text{K} \times \frac{8\Omega}{\frac{8}{1} \times 3.0994\Omega + 8\Omega} = 10.82\text{V}$$

With a group of eight modules, a logical alternative generator connection configuration is two parallel strings of four series modules, i.e., N_s = 4 and N_p = 2. Generator resistance for this configuration is thus:

$$R_{gen} = \frac{N_s}{N_p} \times R = \frac{4}{2} \times 3.0994\Omega = 6.2\Omega$$

While R_{gen} = 6.2Ω does not exactly match the 8.0Ω load resistance, this is the closest resistance match that can be obtained with the selected module type. The voltage for this arrangement is:

$$V_o = N_s \times S\Delta T \frac{R_L}{\frac{N_s}{N_p} \times R + R_L}$$

$$= 4 \times 0.05544 \times 100\text{K} \times \frac{8\Omega}{\frac{4}{2} \times 3.0994\Omega + 8\Omega} = 12.49\text{V}$$

V_o is close to the desired 12V value. If 'fine tuning' of V_o is required, this can be accomplished either by some form of electronic voltage regulation or by externally altering the applied temperature differential ΔT. If the output voltage is significantly out of range despite trying all possible series/parallel combinations, an alternate thermoelectric module having a different current rating and/or number of couples should be used.
With V_o established, the output power P_o is:

$$P_o = \frac{V_o^2}{R_L} = \frac{12.49^2}{8\Omega} = 19.51\text{W}$$

and the load current is 12.49V/8Ω = 1.562A.
The total heat input P_H to the generator, at 1.56A, is:

$$P_H = N_T \times \left(S \times T_h \times \frac{I_{te}}{N_p} - ½ \times \left(\frac{I_{te}}{N_p} \right)^2 \times R + K \times \Delta T \right)$$

$$= 8 \times \left(0.05544 \times 403.2\text{K} \times \frac{1.562\text{A}}{2} - ½ \times \left(\frac{1.562}{2} \right)^2 \times 3.2\Omega + 0.6632 \times 100\text{K} \right) = 662.4\text{W}$$

The generator efficiency η is:

$$\eta = \frac{P_o}{P_H} \times 100\% = \frac{19.51\text{W}}{662.4\text{W}} \times 100\% = 2.95\%$$

The heat transferred to the cold-side heat sink P_o is:

$$P_c = P_H - P_o = 662.4 - 19.5 = 642.9 \text{ W}$$

The maximum allowable thermal resistance P_s of the cold-side heat sink is:

$$P_s = \frac{T_{rise}}{P_c} = \frac{30°\text{C} - 10°\text{C}}{642.9\text{W}} = 0.031°\text{C/W}$$

♣

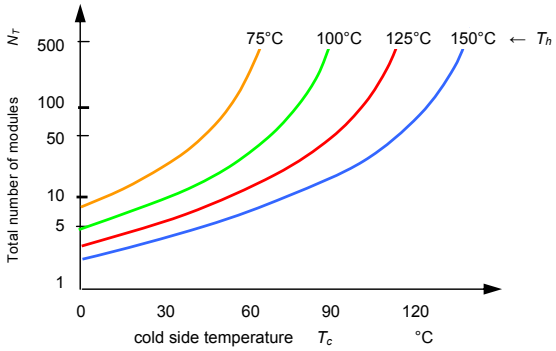


Figure 23.65. *The total number of 127-couple, 6A Modules required for a 12V, 1A thermoelectric power generator*

For any thermoelectric generator design, it is desirable to maximize the applied temperature differential in order to minimize the total number of modules in the system. This situation can be seen in figure 23.65. Module requirements for a typical 12V, 1A power generator are plotted for several values of T_h based on the use of 127-couple 6A TE modules. From this graph, a large number of modules are needed when the cold side temperature T_c is high and the temperature differential is small. Performance of the cold-side heat sink is import and its thermal resistance must be low. In many cases, cold-side heat sink design will prove to be a challenging engineering problem.

23.11.7 *Thermoelectric performance*

The best room temperature thermoelectric materials are alloys of Bi_2Te_3 with Sb_2Te_3 (giving a p-type semiconductor material) and Bi_2Te_3 with Bi_2Se_3 (giving an n-type semiconductor material). *ZT* is of the order of 1 at room temperature. This *ZT* value gives a Coefficient of Performance, *CoP*, of about 1, as shown in figure 23.66a, which compared to household refrigerators and air conditioners, with CoP's from 2 to 4), makes thermoelectric cooling generally not competitive. The same holds for power generation, as shown by the low thermoelectric efficiencies in figure 23.66b.

One of the problems with traditional Peltier elements is their limited capability of cooling heat fluxes over 5 to 10W/cm^2. Because the cooling density of a Peltier cooler is inversely proportional to its length, scaling to smaller size is desirable. The material structure resulting from conventional crystal growth techniques for producing bismuth telluride thermoelectric materials impose significant limitations on thermoelectric element dimensions. Poor manufacturing yields prevent thermoelectric elements from being made very short. New fine-grain micro-alloyed bismuth telluride materials that do not suffer element geometry limitations, offer better performance. Nanocoolers use a monolithic process with thicknesses of about 1 to 2 micrometres, have a tunable performance of 10 to 1000W/cm^2 with a single stage ΔT of 50 to 70K. Thin-film thermoelectrics offer cooling in excess of 160W/cm^2.

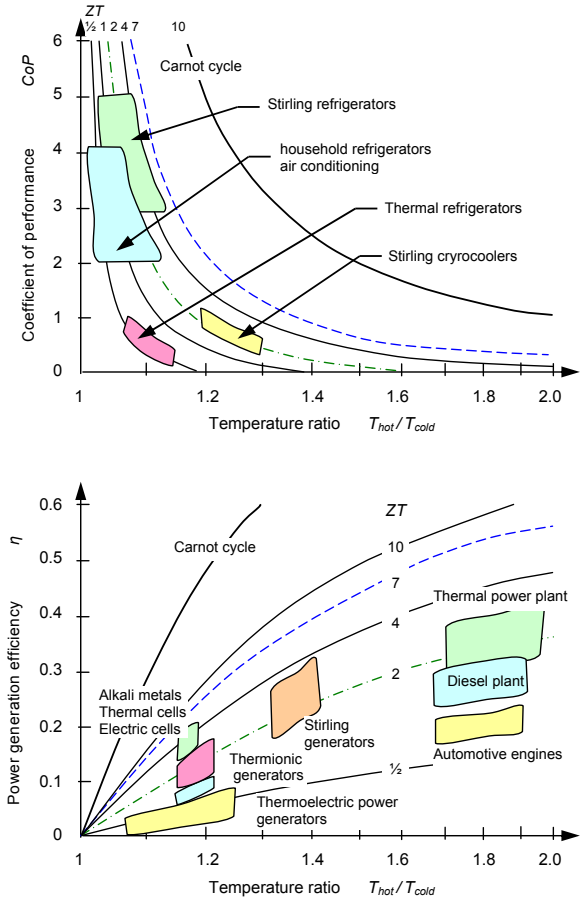


Figure 23.66. *Comparison of thermoelectric technology with other energy conversion methods for: (a) cooling and (b) power generation.*

23.12 Appendix: Primary cells

Zinc Carbon Battery (1.5V, -10°C to 55°C)
The zinc/carbon cell uses a zinc anode and a manganese dioxide cathode. Carbon is added to the cathode to increase conductivity and retain moisture. The manganese dioxide takes part in the reaction, not the carbon.

Negative (anode) reaction: $Zn \rightarrow Zn^{2+} + 2e^-$
Positive (cathode) reaction: $2NH^{4+} + 2MnO_2 + 2e^- \rightarrow Mn_2O_3 + H_2O + 2NH_3$

The cathode reaction is complicated by the fact that the ammonium ion produces 2 gaseous products:

$2NH^{4+} + 2e^- \rightarrow 2NH_3 + H_2$

These two products are absorbed by 2 mechanisms in order to prevent pressure build up.:

$ZnCl + 2NH_3 \rightarrow Zn(NH_3)_2Cl_2$
$2MnO_2 + H_2 \rightarrow Mn_2O_3 + H_2O$

Total reaction (discharge): $Zn + 2MnO_2 \rightarrow ZnO + Mn_2O_3$

Lithium Manganese Dioxide Battery (3V, -40°C to +125°C)
The coin-type lithium manganese dioxide battery uses manganese dioxide (MnO_2) as its positive active material, lithium (Li) as its negative active material, and an organic electrolyte.

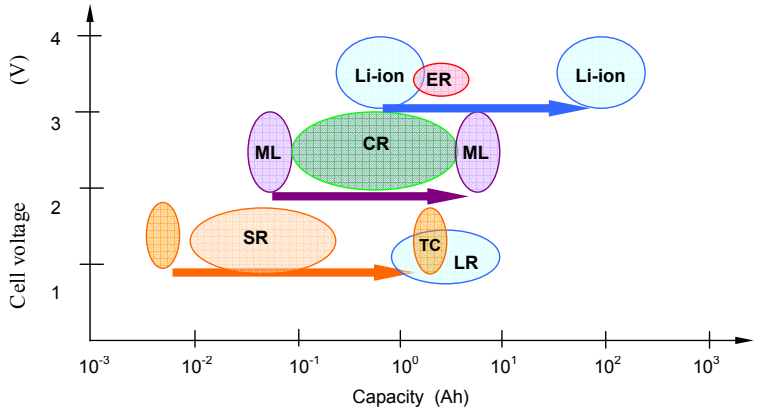
Positive reaction: $MnO_2 + Li^+ + e^- \rightarrow MnOOLi$
Negative (anode) reaction: $Li \rightarrow Li^+ + e^-$
Total reaction (discharge): $MnO_2 + Li \rightarrow MnOOLi$

Lithium Thionyl Chloride Battery (3.6V, -55°C to 85°C)
The lithium thionyl chloride battery uses liquid thionyl chloride ($SOCl_2$) as its positive active material, and lithium (Li) as its negative active material. The reactions of the battery are shown below.

Positive reaction: $2SOCl_2 + 4Li^+ + 4e^- \rightarrow 4LiCl + S + SO_2$
Negative (anode) reaction: $Li \rightarrow Li^+ + e^-$
Total reaction (discharge): $2SOCl_2 + 4Li \rightarrow 4LiCl + S + SO_2$

Silver Oxide Battery (1.55V)
The silver oxide battery uses stable monovalent silver oxide (Ag_2O) as its positive (cathode) active material and fine zinc alloy (Zn) as its negative active material. Potassium hydroxide (KOH) – for high drain or sodium hydroxide (NaOH) – for low drain, are used as an electrolyte.

Positive reaction $Ag_2O + H_2O + 2e^- \rightarrow 2Ag + 2OH^-$
Negative (anode) reaction $Zn + 2OH^- \rightarrow ZnO + H_2O + 2e^-$
Total reaction (discharge): $Ag_2O + Zn \rightarrow 2Ag + ZnO$



Primary
CR- Lithium Manganese Dioxide Battery
E- Lithium Thionyl Chloride Battery
SR- Silver Oxide Battery

Secondary
Li-ion - Lithium Ion Rechargeable Battery
ML- Lithium Manganese Dioxide Rechargeable Battery
TC- Titanium Carbon Lithium Rechargeable Battery

Alkaline Manganese Battery (1.50V)
Electrolytically produced powdered manganese dioxide (MnO_2) is used for the active cathode material, and specially processed fine zinc-alloy powder is used for the active anode material. Concentrated aqueous potassium hydroxide (with added zinc oxide to retard anode corrosion) is used as an electrolyte.

Positive reaction $2MnO_2 + {}_2H_2O + 2e^- \rightarrow 2MnOOH + 2OH^-$
Negative (anode) reaction $Zn + 2OH^- \rightarrow ZnO + H_2O + 2e^-$
Total reaction (discharge): $2MnO_2 + Zn + H_2O \rightarrow 2MnOOH + ZnO$

Zinc Air Battery (1.4V)
The cathode is catalyzed carbon that reduces oxygen from the air, with the anode a gelled mixture of amalgamated zinc powder and a highly conductive solution of KOH in water electrolyte. The electrode reactions for a zinc air battery are as follows

Positive (cathode) reaction $O_2 + 2H_2O + 4e^- \rightarrow 4OH^-$
Anode reaction $2Zn + 4OH^- \rightarrow 2ZnO + 2H_2O + 4e^-$
Overall reaction: $2Zn + O_2 \rightarrow 2ZnO$

	Battery Type	Anode	Cathode	Electrolyte	Advantages	Disadvantages
Primary Batteries	Alkaline Cell	Zn	MnO_2	KOH	High energy density, long shelf life, good leak resistance, performs well under heavy or light use.	Costlier than zinc-carbon cell but more efficient
	Aluminum/Air Cell	Al	O_2	KOH or neutral salt solution	Can operate exposed to sea water (neutral salt solution), easily replaceable electrolytes/electrodes	Anode quickly degrades, short shelf life, short operational life
	Leclanché Cell (Zinc Carbon or Dry Cell)	Zn	MnO_2	NH_4Cl or $ZnCl_2$	Cheap and common (oldest available battery type)	Poor performance under heavy or continuous use.
	Lithium Cell	Li	Various liquid or solid materials	$SOCl_2$, SO_2Cl_2, or organic solutions	Very high energy density, long shelf life, long operational life	Poor performance under heavy use, vulnerable to leaks or explosions
	Mercury Oxide Cell	Zn or Cd	HgO	KOH	Higher energy density than (Zn/MnO_2) alkaline cell	High cost and being phased out due to toxicity concerns
	Zinc/Air Cell	Zn	O_2	KOH	Environmentally benign, cheap, very high energy density, and virtually unlimited shelf life	Short operational life, low power density
Secondary (rechargeable) Batteries	Iron Nickel Cell	Fe	$Ni(OH)_2$	KOH	Long life under a variety of conditions, excellent back-up battery	Low rate-performance, slow recharge rate
	Lead/Acid Cell	Pb	PbO_2	dilute $H_2SO_{4(aq)}$	Low cost, long life cycle, operates well under a variety of conditions. Common car batteries	Minor risk of leakage
	Lithium Ion Cell	C, carbon compounds	Li_2O, intercalated into graphite	$LiPF_6$, $LiBF_4$, related compounds	Relatively cheap, high energy density, long shelf life, long operational life, long cycle life	Minor risk of leakage
	Nickel/Cadmium Cell	Cd	$Ni(OH)_2$	KOH	Good performance under heavy discharge and/or low temperature	High cost, can temporary loose cell capacity if not fully discharged before recharging (memory effect)
	Nickel/Metal Hydride (NiMH) Cell	Lanthanide or Ni alloys	$Ni(OH)_2$	KOH	High capacity and power density	High cost, some memory effect
	Nickel/Zinc Cell	Zn	NiO	KOH	Low cost, low toxicity, good for high discharge rates	Zinc on the electrolyte tends to redeposit unevenly on anode, severely reducing efficiency
	Sodium/Sulfur Cell	Molten Na	Molten S	Al_2O_3	Inexpensive materials, long cycle life, high energy and power	High operational temperature lower efficiency, some danger of explosion upon degradation

23.13 Appendix: Empirical Battery Model

The electrochemical battery can be modelled as a series resistor and a charge-dependent voltage source whose voltage as a function of charge has the following reciprocal relationship:

$$V = V_o\left[1 - \frac{\alpha(1-x)}{1-\beta(1-x)}\right]$$

where: x is the ratio of the ampere-hours left to the number of ampere-hours, AH, for which the battery is rated.
V_o is the voltage when the battery is fully charged, as defined by the nominal voltage, $V_{nominal}$ parameter.
The constants α and β are calculated to satisfy the following battery conditions:

- the battery voltage is zero when the charge is zero, that is, when $x = 0$.
- the battery voltage is V_1 (the voltage $V_1 < V_{nominal}$ when charge is AH_1 parameter value) when the charge is the charge AH_1 when no-load volts are V_1 parameter value, that is, when $x = AH_1/AH$.

The equation defines a reciprocal relationship between internal voltage and remaining charge. It is an approximation to a real battery, but it does replicate the increasing rate of voltage drop at low charge levels. It also ensures that the battery voltage becomes zero when the charge level is zero.

$V_{nominal}$ is the nominal voltage; the open circuit voltage at the output terminal when the battery is fully charged, typically 12 V. Ampere-Hour rating, AH, is the maximum battery charge in ampere-hours, typically 50 hr*A. Initial charge: the battery charge at the start, typically 50 hr*A.

R_{int} is the internal resistance, typically less than 1Ω.

Battery charge capacity models the charge capacity of the battery according to the battery voltage which decreases as charge decreases.

Voltage $V_1 < V_{nominal}$ when charge is AH_1. The battery no-load output voltage, typically 11.5 V, when the charge level is AH_1, typically 25 hr*A hr*A.

Self-discharge resistance of the battery, R_2: the resistance across the battery output terminals in the model that represents battery self-discharge, typically 2kΩ.

The state of charge SoC measures the fraction of charge remaining in the battery, and is defined as

$$SoC = 100\left(1 - \frac{Q}{C}\right) = 100\left(1 - \frac{\int_o^t idt}{C}\right) \qquad 0 \le SoC \le 1$$

where C is the battery capacity in Ah and Q is the charge already delivered in Ah.

A fully charged battery has a SoC of 100%, while a fully discharged battery has a SoC of 0%. Depth of discharge DoD measures the fraction of discharge reached, and is DoD=1-SoC.

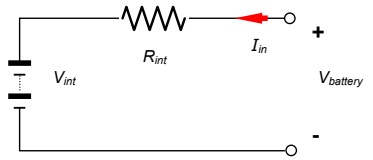


Figure 23.67. *Battery model.*

Reading list

http://www.mpoweruk.com/
http://science.howstuffworks.com/solar-cell3.htm
http://americanhistory.si.edu/fuelcells/index.htm
http://www1.eere.energy.gov
http://www.fctec.com/fctec_basics.asp

24

Capacitors

Selection of the correct type of capacitor is important in all power electronics applications. Just satisfying capacitances and voltage requirements is usually insufficient. In previous chapters, capacitors have been used to perform the following circuit functions:

- turn-off snubbering (8.3.1)
- *dv/dt* snubbering (8.1)
- RFI filtering (10.2.4, 14.7)
- transient voltage sharing of series connected devices (10.1.1)
- voltage multipliers (11.1.3iv)
- dc rail splitting for multilevel converters (15.3)
- cascaded multilevel inverters for VAr compensation (15.3)
- power *L*-*C* filters (15.6)
- switched-mode power supply output filtering and dc blocking (17)
- ac power factor correction and compensation (19.2)

as well as

- dc rail decoupling
- motors for single phase supplies

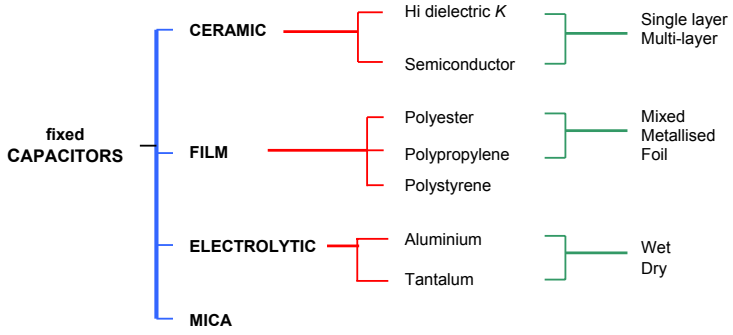
which is just to name a few uses of capacitors in electrical power applications. In each application, the capacitor is subjected to stresses, such as high temperature, *dv/dt* or high ripple current, which must be taken into account in the design and selection process. To make the correct capacitor selection it is necessary to consider various capacitor types, their construction, electrical features, and uses.

Two broad capacitor types are found extensively in power electronic circuits, namely:

- liquid and solid (wet and dry) electrolyte, oxide dielectric capacitors, for example an aluminium electrolytic capacitor
- plastic film dielectric capacitors, for example a polyester capacitor.

The first capacitor group has a metal oxide dielectric and offers large capacitance for a small volume. The second capacitor group, which uses a thin plastic film as a dielectric, offers high ac electrical stress properties.

Ceramic and mica dielectric capacitors are also considered. Ceramic capacitors are used extensively in high power, high frequency switched mode power supplies where they offer small size, low cost, and good performance. The voltage and capacitance ranges for the four main types of dielectric capacitors are shown in figure 24.1.



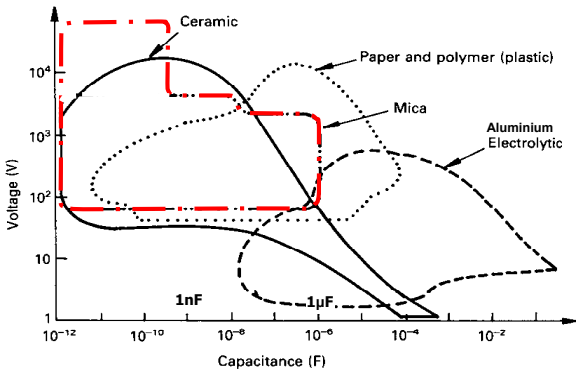


Figure 24.1. *Voltage/capacitance boundaries for the principal types of capacitors.*

24.1 General capacitor properties

The following general principles, properties, and features are common to all capacitor dielectric types.

24.1.1 Capacitance

The primary function of a capacitor is to store electrical energy in the form of a charge. The amount of electrical charge, *Q*, is given by

$$Q = CV = \int i\,dt \quad \text{(C)} \quad [\text{or} \quad i = C\,dv/dt] \tag{24.1}$$

while the stored energy and force between the plates are given by

$$W = \tfrac{1}{2}QV = \tfrac{1}{2}CV^2 \quad [\text{or} \quad W = \tfrac{1}{2}DE \times Aw] \quad \text{(J)} \qquad F = \tfrac{1}{2}CV^2/w = \tfrac{1}{2}\varepsilon_o\varepsilon_r AV^2/w^2 \quad \text{(N)} \tag{24.2}$$

The value of capacitance, *C,* is directly proportional to surface area, *A,* and inversely proportional to the thickness of the dielectric layer (plate separation distance), *w;* that is

$$C = \varepsilon_r\varepsilon_o\frac{A}{w} \quad \text{(F)} \tag{24.3}$$

The dielectric constants ε_r (or *K)* for materials in common usage, are summarised in Table 24.1.

Table 24.1: Dielectric constants for common dielectric materials

Dielectric material $\varepsilon_o = 4\pi\times10^{-7}$	*Relative dielectric constant* ε_r
Vacuum	1
Air (1 atmosphere)	1.00059
Polystyrene	2.5
Polypropylene	2.5
Polycarbonate (now obsolete)	2.8
Polyethylene-terephthalate	3
Impregnated paper	2 - 6
Mica	6.5 - 8.7
Al_2O_3	7
Glass	4 - 9.5
Ta_2O_3	10 - 25
Ceramic	20 -12,000

Dielectrics in capacitors have the property of changing the spacing effectively between two plates. This is manifested in two ways.

First, placing a dielectric between two electric charges reduces the force (Coulomb's Law) acting between them, just as if they were moved apart.

In a vacuum, the force between the two charges is:

$$F_o = \frac{q_1 q_2}{4\pi\varepsilon_o r^2}$$

With a dielectric of relative permittivity ε_r, the force between the two charges reduces to:

$$F_r = \frac{q_1 q_2}{4\pi\varepsilon_o\varepsilon_r r^2} = \frac{q_1 q_2}{4\pi\varepsilon_o\left(\sqrt{\varepsilon_r}\, r\right)^2} = \frac{F_o}{\varepsilon_r}$$

The dielectric increases the effective distance from r to $\varepsilon_r^{1/2} xr$.

Secondly, the dielectric constant of a material affects how electromagnetic fields (light, radio waves, millimetre-waves, etc.) move through the dielectric material. A high dielectric constant increases the effective spacing. This means that light travels slower. It also 'compresses' the waves to behave as if the field has a shorter wavelength.

In a vacuum, the electric field created by a charge q is:

$$E_o = \frac{q}{4\pi\varepsilon_o r^2}$$

With a dielectric of relative permittivity ε_r, the electric field reduces to:

$$E_r = \frac{q}{4\pi\varepsilon_o\varepsilon_r r^2} = \frac{q}{4\pi\varepsilon_o\left(\sqrt{\varepsilon_r}\, r\right)^2} = \frac{E_o}{\varepsilon_r}$$

The dielectric increases the effective distance from r to $\varepsilon_r^{1/2} xr$. In the case of a capacitor it is the electric field effect created by the dielectric that is relevant, $\varepsilon_r = E_o / E_r$. Since $C=Q/V$, the dielectric effectively decreases E (hence V), thus increases capacitance C.

The effective capacitance of parallel, C_p, and series, C_s, connected capacitors are

$$C_p = C_1 + C_2 + C_3 + \ldots$$

$$\frac{1}{C_s} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} + \ldots \qquad (24.4)$$

In parallel, the capacitor with the lowest voltage rating specifies the parallel combination voltage rating.

24.1.2 Volumetric efficiency

The volumetric efficiency of a capacitor is a measure of the effectiveness of a given physical construction and dielectric material. Volumetric efficiency η_v, is defined by

$$\eta_v = \frac{C_R \times V_R}{volume} \qquad \left(\mathrm{C/m^3}\right) \qquad (24.5)$$

Dimensionally, longer is better since there is less percentage dielectric wastage of the unused dielectric ends. Cylindrical is better than oval, except an oval cross-section may allow better stacking with parallel connected capacitors or may result in lower lead inductance for pcb mounted capacitors.

24.1.3 Equivalent circuit

The impedance of a capacitor can be modelled by one of the capacitor equivalent circuits shown in figure 24.2. In series with the ideal capacitor, C_R, termed *rated capacitance*, is an *equivalent series resistor* R_s (ESR) and *equivalent series inductor* L_s (ESL). R_s is determined by lead and junction resistances, while L_s is the inductance of the electrodes due to the construction and the supply lines. The value of L_s is usually given for a specific package and capacitor type, and is generally neglected at lower frequencies, below the self-resonant frequency, which is given by

$$\omega_r = \frac{1}{\sqrt{L_s C_R}} \qquad (\mathrm{rad/s}) \qquad (24.6)$$

The electrical impedance Z of a capacitor, neglecting

- R_i the leakage (insulation) resistance which is usually large,
- R_d is the dielectric loss due to dielectric absorption and molecular polarisation (significant at high frequencies), and
- C_d is the inherent dielectric absorption, only significant in electrolytic capacitors

is

$$Z = R_s + jX = ESR + j2\pi f \times ESL - j\frac{1}{2\pi f \times C_R} \qquad (\Omega) \qquad (24.7)$$

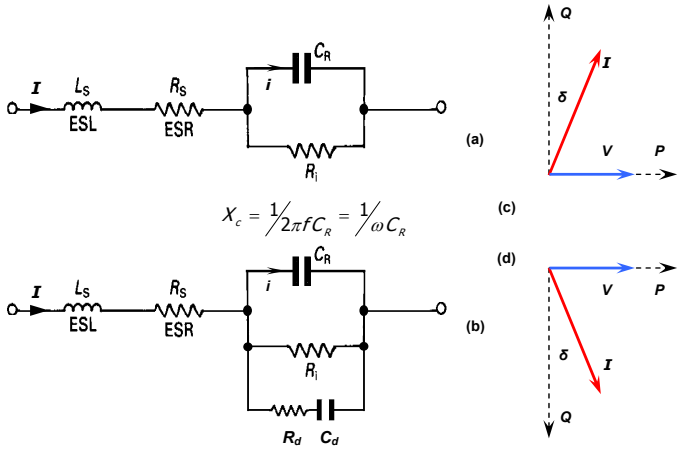
Since the ESL is neglected, at lower frequencies, when $2\pi f$ (that is ωL) is small

$$Z = R_s - \frac{j}{\omega C_R} \qquad (\Omega) \qquad (24.8)$$

and

$$\tan\delta = \omega C_R R_s = \omega C_R \times ESR = \frac{R_s}{X_c} = \frac{1}{Q} = \frac{\text{real power}}{\text{reactive power}} \qquad (24.9)$$

where δ is the loss angle and tanδ is termed the *dissipation factor, DF*, which is the reciprocal of the circuit quality factor, Q. The angle δ is that necessary to make the capacitor current lead the terminal voltage by 90° in figure 24.2c and d, as for an ideal capacitor.



$$X_c = 1/2\pi f C_R = 1/\omega C_R$$

Figure 24.2. *Capacitor: (a) equivalent circuit for a film capacitor and; (b) electrolytic capacitor; and terminal V-I phasor diagram; (c) below resonance; and (d) above resonance.*

If the insulating or dielectric dc leakage resistance, R_i *(= $\rho_i \ell/A$)*, is included, then

$$\tan\delta = \frac{1}{\omega C_R R_i} + \omega C_R R_s \qquad (24.10)$$

and at low frequency ($\omega << \omega_o$)

$$\tan\delta_l = \frac{1}{\omega C_R R_i} \qquad (24.11)$$

while at high frequency ($\omega >> \omega_o$)

$$\tan\delta_u \approx \omega C_R R_s \qquad (24.12)$$

Both R_s and X_c are dependent on temperature and frequency as shown in figure 24.3. Figure 24.3a shows that the rated capacitance illustrated has a positive temperature coefficient, the value of which also depends on capacitance and rated voltage. Also shown is the negative temperature dependence of equivalent series resistance ESR. Figure 24.3b shows that C_R and ESR both decrease with frequency. Since C_R and ESR are temperature and frequency dependent, and are related to tan δ and Z, then tan δ and Z are frequency and temperature dependent as illustrated in figures 24.3c and 24.3d. Figure 24.3c shows the typical characteristics of the impedance of an oxide dielectric capacitor versus frequency, at different temperatures. At low frequencies the negative slope of Z is due to the dominance of the capacitive reactance, $Z \approx X_c = 1/\omega C_R$, whereas the horizontal region, termed the resonance region, is where Z is represented by the ohmic resistance R_s, that is $Z \approx R_s$. At higher frequencies the inductive reactance begins to dominate, whence $Z \approx \omega L_s$ and tan $\delta = R_s/\omega_s L$.

Figure 24.3d shows how the dissipation factor, tanδ, increases approximately proportionally with frequency to a high value at resonance, as would be expected from equation (24.9). At lower frequencies tanδ may be considered as having a linear frequency dependence, according to tanδ = tan $\delta_o + kf$.

ESR and *tan* δ dictate internal power dissipation hence self-heating, namely, for terminal voltage V

$$P = I^2 \times ESR = 2\pi fC \times \tan\delta \times V^2 = \left(2\pi fC\right)^2 \times ESR \times V^2 \qquad (24.13)$$

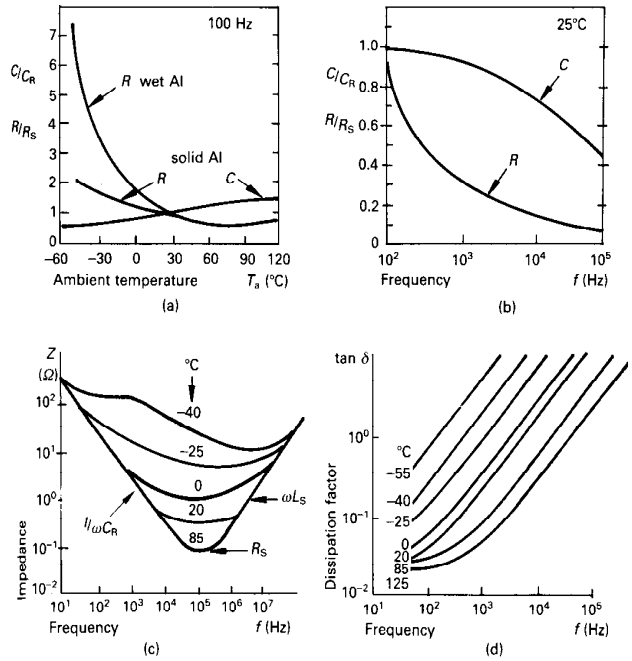


Figure 24.3. *Variation of capacitor equivalent circuit parameters with frequency and temperature for a high voltage (47 µF, 350 V) metal oxide liquid dielectric:(a) R_s and C_R as a function of temperature; (b) R_s and C_R as a function of frequency; (c) impedance Z as a function of frequency and temperature; and (d) tan δ as a function of frequency and temperature.*

24.1.4 Lifetime and failure rate

The service life of a capacitor occurs when its parameters fall outside the specification limit, termed *degradation.* Such parameters are usually the capacitance, dissipation factor, impedance, and leakage current. The service life is specified under specific operating conditions such as voltage, ambient temperature, and current, and will increase

- the lower the ambient temperature, T_a
- the lower the ripple current or voltage, I_r
- the lower the operating voltage in proportion to the rated voltage, V_{op}/V_R
- the higher the ac load frequency, *f*.

Other factors may be relevant to specific dielectrics.
Lifetime is the period until a given failure rate is reached. The failure rate, λ, is the ratio of the number of failures to the service life expected. It is usually indicated in failures per 10^9 component hours (*fit* – failure in time) and is an indicator of equipment reliability.
If, in a large number *N* of identical components, percentage Δ*N* fail in time Δ*t*, then the failure rated λ, averaged over Δ*t* is expressed as

$$\lambda = \frac{1}{N} \times \frac{\Delta N}{\Delta t} \qquad (/\text{h}) \qquad (24.14)$$

If the sample *N* is large, then the failure rate in time can be represented by a continuous 'bathtub'-shaped curve as shown in figure 24.4, such that

$$\lambda = \frac{1}{N} \frac{dN}{dt} \qquad (/\text{h}) \qquad (24.15)$$

This figure shows the three distinct failure periods, and the usual service life is specified according to the failure λ_o, which is constant.

In the case of voltage, current, and other stresses including temperature, which differ from those under which λ_o is specified, *conversion* or *acceleration factors* are used to calculate the new failure rate.
Typical conversion factors are given in Table 24.2 for ambient temperature T_a, and operating voltage V_{op}, in relation to rated voltage V_R. Alternatively conversion graphs are also used or the Arrhenius' law

$$\lambda = \lambda_o \left(\frac{V_{op}}{V_R}\right)^n e^{-\frac{E_a}{K}\left(\frac{1}{T}-\frac{1}{T_o}\right)} \qquad (24.16)$$

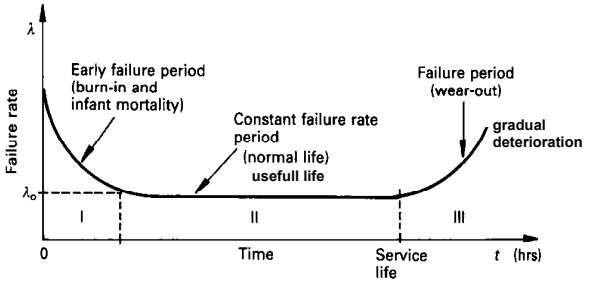


Figure 24.4. *The bathtub curve showing variation of failure rate with operating hours.*

Table 24.2: Stress conversion factors for an aluminium electrolytic capacitor

$\frac{V_{op}}{V_R}$ %	*Conversion factor*	*Temperature* T_a (°C)	*Conversion factor*
100	1	≤40	1
75	0.4	55	2
50	0.2	70	5
25	0.06	T_{jmax}	10
10	0.04		
(a)		(b)	

Example 24.1: ***Failure rate***

A component has a failure rate λ_o = 2 x 10^9/h, commonly termed 2 fit (failures in time) using 10^9/h as reference.
With reference to Table 24.2, what is the failure rate if

i. the ambient temperature, T_a, is increased to 55°C
ii. the operating voltage is halved
iii. i. and ii. occur simultaneously?

Solution

Assume λ_o applies to conditions at $T_a \le 40$°C and V_R.

i. If the ambient temperature is increased from 40°C to 55°C, then using a conversion factor of 2 from table 24.2b

$$\lambda_{55} = 2 \times \lambda_o$$
$$= 4 \quad \text{fit}$$

that is, the failure rate has doubled, from 2 fit to 4 fit.

ii. Similarly, by halving the operating voltage, a conversion factor of 0.2 is employed from table 24.2a. The new failure rate is

$$\lambda_{\frac{1}{2}V} = 0.2 \times \lambda_o$$
$$= 0.4 \quad \text{fit}$$

That is, the failure rate has decreased by a factor of 5, from 2 fit to 0.4 fit.

iii. If simultaneously both the ambient temperature is increased to 55°C and the operating voltage is halved, then (assuming independence and superposition of factors)

$$\lambda_{55\,½V} = 2 \times 0.2 \times \lambda_o$$
$$= 0.8 \quad \text{fit}$$

The conversion factors are cumulative and the failure rate decreases from 2 fit to 0.8 failures in time.

♣

If the number of units surviving decreases exponential with time, then the probability of failure *F* after a service time *t* is given by

$$F(t) = 1 - e^{-\lambda t} \tag{24.17}$$

Equipment failure rate can be calculated by summing the failure rates of the individual components, that is

$$\lambda_{total} = \lambda_1 + \lambda_2 ..... + \lambda_n \tag{24.18}$$

If the failure rate is to be constant, then the instantaneous failure rate of the number of faults per unit time divided by the number of non-failure components must yield a constant

$$\frac{1}{1-F(t)} \frac{dF(t)}{dt} = \lambda \tag{24.19}$$

For *n* components in a system, the probability of system survival is

$$1 - F(t) = (1 - F_1(t)) \times (1 - F_2(t)) \times .... (1 - F_n(t)) = e^{-\lambda_1 t} \times e^{-\lambda_2 t} \times ..... \; e^{-\lambda_n t}$$
$$= n\lambda \tag{24.20}$$

if, since the units are identical, $\lambda_1 = \lambda_2 = \ldots = \lambda_n$.

The *mean time between failure* (*mtbf*) is given by

$$mtbf = \frac{1}{\lambda_{total}} = \int_0^\infty 1 - F(t)\, dt = \int_0^\infty e^{-\lambda t}\, dt = \frac{1}{\lambda} \tag{24.21}$$

The service operating life *τ* for a specified probability of failure is therefore given by

$$\tau = \frac{1}{\lambda} \ell n \frac{1}{1-F} \tag{24.22}$$

Example 24.2: ***Capacitor reliability***

A capacitor has a failure rate λ of 200 x 10^{-9} failure/hour, 200 fit. Calculate

i. the probability of the component being serviceable after one year
ii. the service life if the probability of failure is chosen to be 1% or 0.1%
iii. the mean time between failure
iv. the mean time between failure for 10 parallel connected capacitors
v. the probability of survival for 1 year and of failure for units, if 1000 units each have 10 parallel connected capacitors.

Solution

i. The probability of the capacitor being serviceable after 8760 h (1 yr) is given by

$$1 - F(1\,yr) = e^{-\lambda t}$$
$$= e^{-200 \times 10^9 \times 8760} = 0.998 \quad (99.8\%)$$

ii. Component lifetime is given by

$$\tau = \frac{1}{\lambda} \ell n \frac{1}{1-F}$$
$$\tau(1\%) = \frac{10^9}{200} \ell n \frac{1}{1-0.01} = 50{,}000 \text{ h} = 5.7 \text{ years}$$
$$\tau(0.1\%) = \frac{10^9}{200} \ell n \frac{1}{1-0.001} = 5{,}000 \text{ h} = 0.57 \text{ years}$$

iii. The mean time between failure, given by equation (24.21) is

$$mtbf = 1/\lambda = \frac{10^9}{200} = 5 \text{ x } 10^6 \text{ h} = 570 \text{ years}$$

iv. The failure rate for 10 capacitors is 10λ = 2000 fit and the mtbf is

$$\frac{1}{10\lambda} = \frac{10^9}{2000} = 57 \text{ years}$$

v. For 1000 units, each with a failure rate of 10λ, the probability of one unit surviving 1 year is

$$1 - F(1\text{ yr}) = e^{-10\ x\ 200\ x\ 10^{-9}\ x\ 8760} = 98.2 \text{ per cent}$$

The probable number of first year failures with 1000 units is

$$F(1\text{ yr}) = 1 - e^{-200\ x\ 10^{-9}\ x\ 8760} = 0.002 \text{ pu} = 2 \text{ units}$$

♣

The reliability concepts considered are applicable to all electronic components (passive and active) and have been used to illustrate capacitor reliability.

24.1.5 Self-healing

One failure mode of a capacitor is voltage breakdown in a defective area of the dielectric. As a result of the applied voltage, the defective area (due to pores and film impurities) experiences an abnormally high electric field which may cause failure by arcing within a few tens of nanoseconds. Oxide capacitors using an electrolyte and plastic film dielectric capacitors exhibit self-healing properties, which in the case of plastic film dielectrics allow the capacitor to remain functional after voltage breakdown.

In the case of a defect in the dielectric oxide layer of an electrolytic capacitor, the maximum field strength is reached first in the defective region. This is effectively the process which occurs during the formation of the oxide layer, which results in the growth of new oxide, thereby repairing the defect. The reforming process is relatively slow compared with the healing time for non-polarised capacitors.

By contrast, the high electric field at the defect in a metallised plastic film capacitor causes a continuous high pressure plasma arc which pushes the dielectric layers apart and evaporates the metallisation in the breakdown region. Temperatures can reach 6000K and insulated areas are formed around the original failure area, which after the arc self-extinguishes, isolate the faulty dielectric within 10µs..

24.1.6 Temperature range and capacitance dependence

The operating temperature upper and lower limits are either dictated by expected service life or the allowable variation limits on the nominal capacitance. Most capacitors can be used outside their nominal temperature limits, but at reduced lifetime, hence with reduced reliability. The extremes -55°C to 125°C are common, but obviously electrolytic capacitors must be restricted to a smaller range if the electrolyte is not either to freeze or to boil.

Capacitor reversible temperature dependence can be expressed in terms of a temperature dependant capacitance co-efficient α_c, by

$$C(T) = C_{20°C}\left(1 + \alpha_c (T - 20°C)\right) \tag{24.23}$$

where the temperature co-efficient of capacitance α, with respect to reference C_{ref} at 20°C, is

$$\alpha_c = \frac{C_{T2} - C_{T1}}{C_{ref}(T_2 - T_1)} \qquad 10^{-6}/\text{K}$$

24.1.7 Dielectric absorption

After a capacitor is discharged from a voltage V_i, a small voltage V_r reappears, due to a polarisation process in the insulating material. [This phenomena could be considered to be equivalent to remanence flux in magnetic materials]. The voltage tends to be independent of capacitance and dielelctric thickness and is defined at 20°C. The dielectric absorption factor δ_A is defined by

$$\delta_A = \frac{V_r}{V_i} \times 100\% \tag{24.24}$$

Typical factor percentage values for various dielectric types are shown in the following table.

Dielectric type	polypropylene KP	polyethylene-terephthalate (polyester) KT	mixed dielectric	ceramic X7R	ceramic Z5U
δ_A %	0.05 - 0.10	0.21 - 0.25	0.12 - 0.15	0.60 - 1.00	2.00 - 2.50

24.2 Liquid (organic) and solid, metal oxide dielectric capacitors

The oxides of metals such as aluminium and tantalum are capable of blocking current flow in one direction and conducting in the other. Operation of metal oxide dielectric capacitors is based on the so-called *valve effect* of these two metals.

24.2.1 Construction

The capacitor dielectric layer consists of aluminium oxide Al_20_3 or tantalum oxide Tn_20_3 which is formed by an electrochemical oxidising process of aluminium foil (0.02 to 0.1mm thick) or sintered tantalum powder. These starting metals form the capacitor anode. The oxide layer withstands high electric field strengths, typically 8 x 10^8 V/m for Al_20_3 which represents 1.45 nm per volt, and are excellent insulators (hence result in a high capacitor loss factor). This field strength is initially maintained constant (with constant current) during the oxidising process (this electrochemical process is aided by weak phosphoric acid in the case of tantalum capacitors and chloride solution for electrolytic capacitors), then constant voltage, so that the oxide thickness is dependent and practically proportional to the *forming voltage* V_F. To avoid changing the oxide thickness during normal use, the component *operated rated voltage* V_R should always be lower than the forming voltage, as shown in figure 24.5. The difference V_F - V_R is the *over-oxidisation voltage* and substantially determines the capacitor operational reliability. For general-purpose electrolytic capacitors, the value of V_R / V_F is about 0.8, while solid capacitors are rated at 0.25.

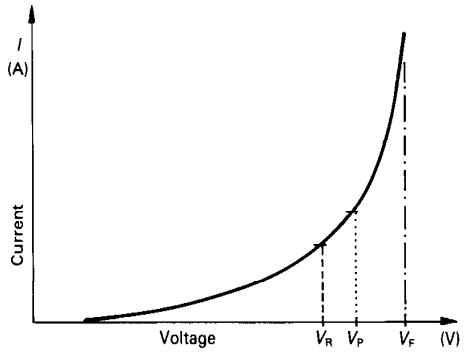


Figure 24.5. *Current (leakage) dependence on voltage of Al electrolytic capacitors.*

The oxide dielectric constant ε_r is approximately 8.5 for Al_20_3 and 25 for Ta_20_3, while in comparison paper-based dielectrics have a value of approximately 5. An oxide thickness of w = 0.7µm is sufficient for high voltage capacitors (≥ 160 V) as compared with minimum practical paper dielectric thickness of about 6µm. The metal oxide type capacitors potentially offer high capacitance per unit volume. To further improve the capacitance per unit volume, before oxidation, the aluminium anode surface area is enlarged 10-300 times (*foil gain* - depending on the voltage – 100x for low voltage and 20x for high voltage capacitors) by electrochemical deep etching processes. In the case of tantalum capacitors, the sintered tantalum sponge like lattice structure results in the same increase of area effect. In the case of tantalum capacitors, the oxide not only grows on the surface of the tantalum, about one third grows into the porous lattice. This tends to limit the maximum voltage rating of tantalum capacitors.

The capacitor is formed by the placement of the cathode on to the oxide layer. In the case of the electrolytic capacitor, a highly conductive organic acid electrolytic (based on dimethylacetamide) which is impregnated into porous paper, forms the capacitor cathode. The electrolyte largely determines the ESR hence it must have a low resistivity over a wide temperature range. It must also have a breakdown voltage well above the capacitor rated voltage at maximum operating temperature. For long life, electrolytes with a water content must be avoided. Teflon spacers are sometimes used rather than paper. The electrical contact to the cathode is a layer of etched aluminium, which has a thin oxide layer.
In the case of solid capacitors, (with lower voltages), a high conductive cathode is formed by a solid semiconductor metal oxide, such as manganese dioxide. This is achieved by the pyrolysis (continued dipping and baking at 200°C) of manganese nitrate in to manganese dioxide. In solid oxide capacitors, the manganese dioxide is dipped into graphite which is coated with silver epoxy for soldering.

The four wet/dry oxide possibilities are shown in figure 24.6. A porous paper or glass fibre is used as a space keeping agent in order to avoid short circuits and direct mechanical contact.

Long strips of the cross-sections are wound into cylindrical bodies and encased as shown in figure 24.6. Operation at high voltages causes oxide growth and the production of hydrogen. Any gas pressure relief valve provided should be orientated upwards, just as the anode terminal should be above the cathode if the capacitor is orientated horizontally.

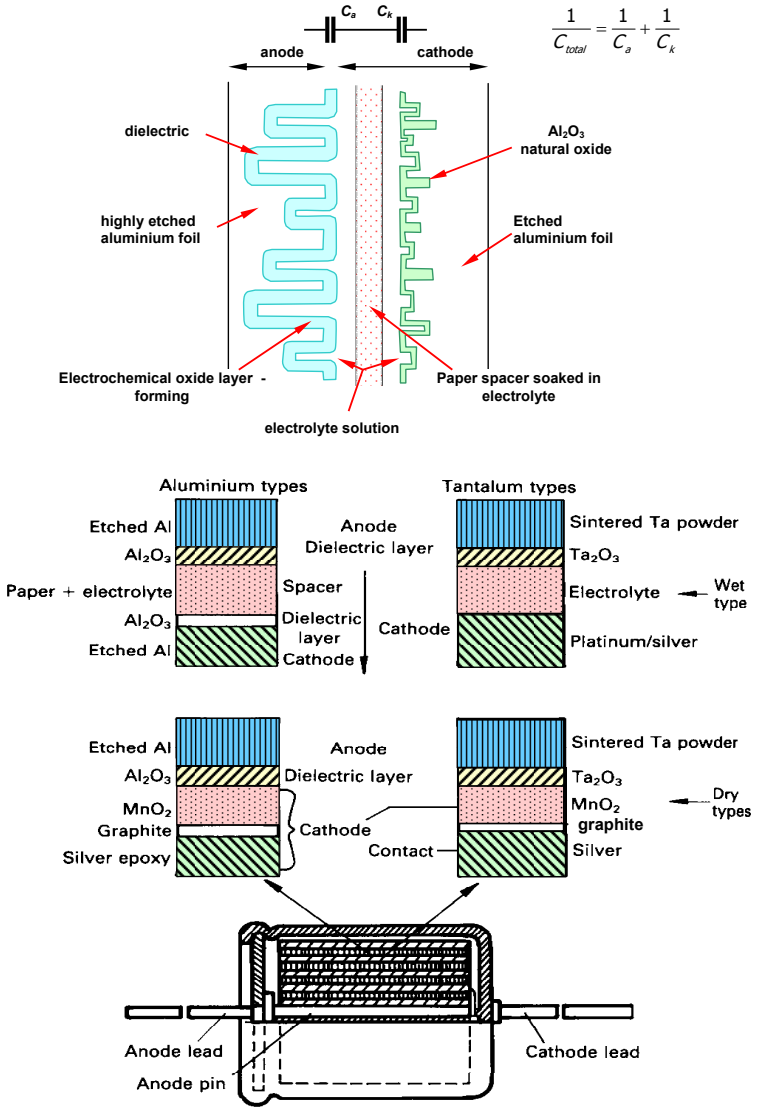


Figure 24.6. *Construction of metal oxide capacitors.*

24.2.2 Voltage ratings

Basic electrolyte (electrolytic) capacitors are suitable only for unipolar voltages, where the anode is positive with respect to the cathode. In the case of the aluminium electrolytic capacitor, the cathode

connection metal does have a thin air-oxide layer which corresponds to an anodically generated layer with a blocking voltage capability of about 2 V. Above this voltage level, an electrolytic generated dielectric oxide film would be formed on the cathode foil. The effect is to decrease the capacitance and cause high internal heating and oxidation of the cathode foil thus gas formation, gas as shown in the following formula, which can lead to failure.

$$2Al + 3H_2O - 6e^- \rightarrow Al_2O_3 + 6H^+$$

$$6H^+ + 6e^- \rightarrow 3H_{2\ gas}$$

This pressure build up may cause the safety vent to open or possibly destroy the capacitor. Deterioration is slow with a reverse voltage of a few volts.

Solid, oxide capacitors are in principle capable of supporting bipolar voltage since the cathode is a semiconductor, manganese oxide. In practice, impurities such as moisture restrict the reverse voltage limits to 5-15 per cent of V_F. The usable reverse voltage decreases with increased ambient temperature. The rated voltage V_R may be exceeded under specified intermittent conditions, resulting in a maximum or peak voltage limit V_P, as shown in figure 24.5, where

for $V_R \leq 315V$ $V_P = 1.15\ V_R$

for $V_R > 315V$ $V_P = 1.1\ V_R$

Both V_R and V_P, are derated with increasing temperature. Tantalum capacitors are linearly voltage derated above 85°C, to 66% at 125°C, which is the maximum operating voltage limit.

24.2.3 Leakage current

When a dc voltage is applied to capacitors, a low current, $I_{\ell k}$ called the *leakage current,* flows through every capacitor, as implied by the presence of R_l in the equivalent circuit model in figure 24.2. With oxide dielectric capacitors, this current is high at first and decreases with working time to a final value, as shown in figure 24.7.

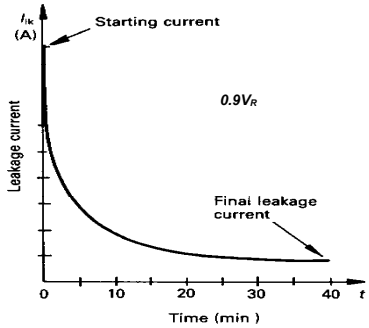


Figure 24.7. *Leakage current variation with working time for a liquid aluminium oxide capacitor.*

A low final leakage current is the criterion of a well designed dielectric, thus leakage current can be considered as a measure for the quality of the capacitor. The current is a result of the oxidising activity within the capacitor. The leakage current depends on both dc voltage and ambient temperature, as shown in figure 24.8. The purity of the anode metal, hence oxide dielectric determines the leakage current.

Liquid, oxide capacitors have the lower leakage currents at rated voltage since when a voltage is applied; anions in the electrolyte maintain the dielectric electrochemical forming process. The $Mn0_2$ in solid oxide capacitors has lower reforming capabilities.
From figure 24.8 it will be seen that leakage increases with both temperature and voltage. The increase in leakage current with temperature is lower in liquid capacitors than in the solid because, once again, the electrolyte can provide anions for the dielectric reforming process.
For an aluminium electrolyte capacitor at 85°C, an expected lifetime of 2000 hours is achieved by selecting $V_R / V_F = 0.8$. However, V_F is inversely proportional to absolute temperature so for the same leakage current at 125°C, the ratio of V_R / V_F must be decreased to

$$\frac{V_R}{V_F} = 0.8 \times \frac{273+85}{273+125} = 0.7$$

For higher temperature operation, a higher forming voltage is required. But since $V_F \times C_R$ (energy volume) is constant for any dielectric/electrode combination, C_R is decreased.

When connecting electrolytic capacitors in series, parallel sharing resistors are necessary to compensate for leakage current variation between the capacitors. The design of the sharing network is as for the steady-state voltage sharing of semiconductors presented in 10.1.1, where the sharing resistance is

$$R = \frac{nV_R - V_s}{(n-1)\Delta I_\ell} \text{ where } \Delta I_\ell = 1.5 \times 10^{-3} C_R V_R \qquad (24.25)$$

Additionally, the resistors provide a discharge path for the stored energy at power-off.
When parallel connecting capacitors, highest reliability is gained if identically rated capacitors (voltage and capacitance) are used.

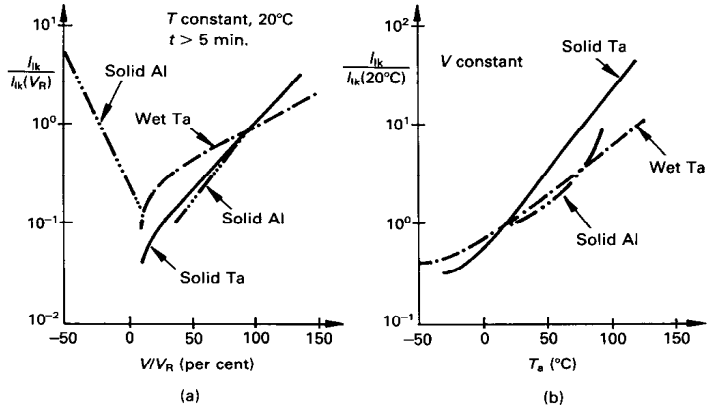


Figure 24.8. *Typical leakage current of oxide capacitors versus: (a) voltage and (b) temperature.*

24.2.4 Ripple current

The maximum superimposed alternating current, or ripple current $\hat{I}_r$ is the maximum rms value of the alternating current with which a capacitor is loaded, which produces a temperature difference of 10 K between the core and ambient. Ripple current results in power being dissipated in the ESR, according to

$$P_d = \hat{I}_r^2 R_s \qquad \text{(W)} \qquad (24.26)$$

which results in an internal temperature rise until equilibrium with ambient occurs, see equation (24.13). The maximum power dissipation $\hat{P}_d$ is dependent on the thermal dissipation properties of the capacitor, and from equation (5.4)

$$\hat{P}_d = h\,A\,\Delta T \qquad \text{(W)} \qquad (24.27)$$

where h = heat transfer coefficient, W/m^{2}K
A = capacitor outer surface area, m^2
$\Delta T = T_s - T_a$ = temperature difference between capacitor surface, T_s, and ambient, T_a ,K

Thus the maximum ripple current is given by

$$\hat{I}_r = \sqrt{\frac{\hat{P}_d}{R_s}} = \sqrt{\frac{hA\Delta\hat{T}}{R_s}} \qquad \text{(A)} \qquad (24.28)$$

The ESR, R_s, is both temperature and frequency dependent, hence rated ripple current I_{ro}, is specified at a given temperature and frequency, and at rated voltage V_R. Due to the square root in equation (24.28), conversion to other operating conditions is performed with the frequency multiplier $\sqrt{r}$ and temperature multiplier $\sqrt{k}$, such that

$$I_r = \sqrt{k}\sqrt{r}\ I_{ro} = \sqrt{k\,r}\ I_{ro} \qquad \text{(A)} \qquad (24.29)$$

Typical multiplier characteristics for aluminium oxide capacitors are shown in figure 24.9. It will be seen from figure 24.9a that electrolytic capacitors are rated at 85°C, while as seen in figure 24.9b solid types are characterised at 125°C. For each type, a reference frequency of 100 Hz is used.

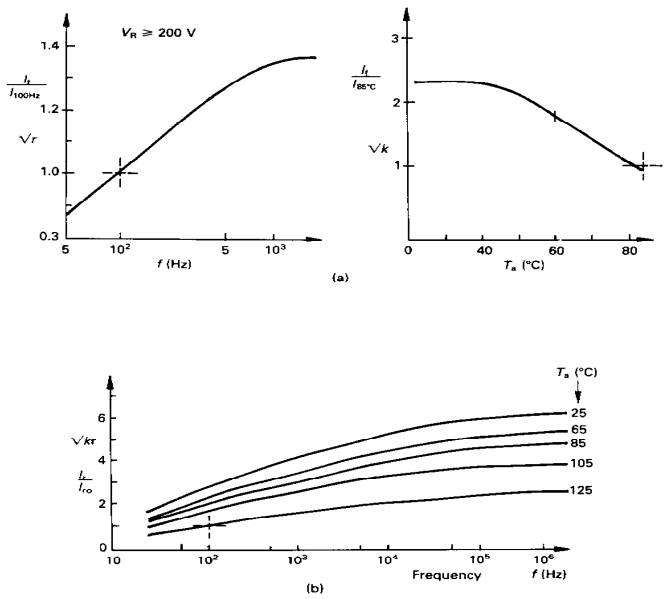


Figure 24.9. *Frequency and temperature ripple current conversion multipliers for: (a) liquid and (b) solid $A1_20_3$ capacitors.*

Alternatively, the temperature derating multiplier is expressed in terms of the ambient, core and rated temperatures by

$$\sqrt{k} = \sqrt{\frac{T_{core} - T_{amb}}{T_{core} - T_R}} \qquad (24.30)$$

No simple expression exist for the frequency derating factor, *r*, although it may be used to infer ESR frequency derating

$$ESR_{freq} = ESR_{100\,Hz} \Big/ r \qquad (24.31)$$

Electrolytic capacitors usually have a thermal time constant of minutes, which can be exploited to allow intermittent overloads.

Example 24.3: ***Capacitor ripple current rating***

A 1000 μF, 385 V liquid, aluminium oxide capacitor has an rms ripple current rating I_{ro} of 3.7 A at 100 Hz and 85°C.
Use figure 24.9a to calculate the allowable ripple current at

i. 60°C and 1 kHz
ii. lowest stress conditions.

Solution

i. Using equation (24.29)

$$I_r = \sqrt{k}\sqrt{r}\, I_{ro} = \sqrt{k\,r}\, I_{ro} \qquad \text{(A)}$$

where from figure 24.9a at 60°C, $\sqrt{k}$ = 1.85
at 1 kHz, $\sqrt{r}$ = 1.33
whence $\hat{I}_r$ = 1.33 × 1.85 × 3.7A
= 9.1 A

ii. This capacitor experiences lowest stressing at temperatures below 40°C, where $\sqrt{k}$ = 2.25 and at frequencies in excess of 2 kHz when $\sqrt{r}$ = 1.37. Under these conditions the ripple current rating is

$$\hat{I}_r = 2.25 \times 1.37 \times 3.7\text{A}$$
$$= 11.4\text{A}$$

♣

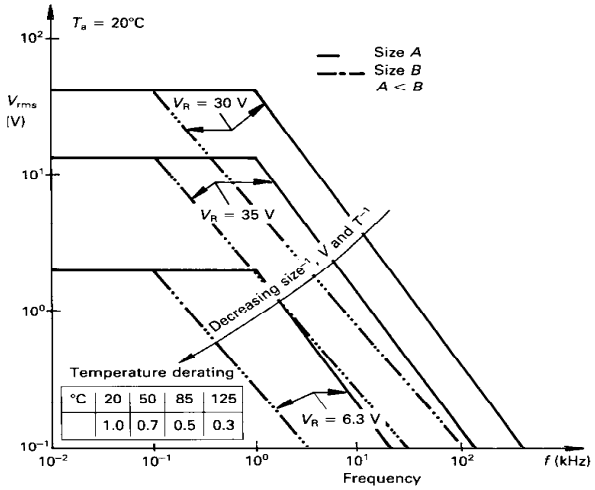


Figure 24.10. *Rms voltage limits of solid tantalum capacitors for different physical dimensions, temperature, voltage rating, and frequency.*

Non-sinusoidal ripple currents have to be analysed such that at a given temperature, the individual frequency components satisfy

$$\hat{I}_r^2 \geq \sum_n \frac{I_n^2}{r_n} \qquad (24.32)$$

where $\hat{I}_r$ is for the appropriate rated ambient and reference frequency as indicated in figure 24.9.
Liquid tantalum capacitors have a ripple current rating which is determined by the physical dimensions, independent of temperature over a wide range, and independent of frequency above 50 Hz.

Ripple current ratings may not be specifically given for some capacitor types, for example solid tantalum capacitors. In this case an indirect approach is used. In satisfying ac voltage limitations as illustrated in figure 24.10, and any series resistance requirement, allowable ripple currents can be specified for a given temperature.

24.2.5 Service lifetime and reliability

24.2.5i - Liquid, oxide capacitors

As considered in 24.1.3, the reliability and lifetime of a capacitor can be significantly improved by decreasing the thermal and electrical stresses it experiences. Stress reduction is of extreme importance in the case of liquid aluminium oxide capacitors since it is probably the least reliable and most inappropriately used common electronic component.
The reliability and service lifetime of an aluminium oxide electrolytic capacitor are dominated by its ripple current, operating temperature, and operating voltage. Figure 24.11 in conjunction with figure 24.9a, can be used to determine service life.

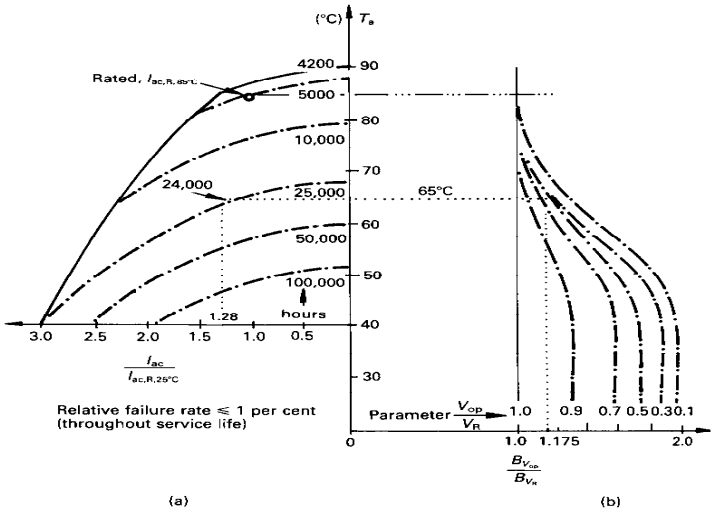


Figure 24.11. *Service life for an aluminium oxide liquid capacitor. Temperature dependence of lifetime variation with: (a) ripple current and (b) operating voltage.*

Example 24.4: **$A1_20_3$ *capacitor service life***

A 1000 μF, 385 V dc aluminium oxide liquid capacitor with a ripple current rating I_{ro} of 2.9 A at 100 Hz and 85°C ambient is used at 5 A, 4 kHz, in a 65°C ambient and on a 240 V dc rail. What is the expected service lifetime of the capacitor?

Solution

From figure 24.9a at 4 kHz, √r =1.35, whence

$$\frac{I_o}{I_{ro}} \times \frac{1}{\sqrt{r}} = \frac{5\text{A}}{2.9\text{A}} \times \frac{1}{1.35} = 1.28$$

From figure 24.11a, the coordinates 1.28 and 65°C correspond to a 24,000 hour lifetime with less than 1 per cent failures. Since a 385 V dc rated capacitor is used on a 240 V dc rail, that is, a ratio 0.64, an increase in service lifetime of 17½ per cent can be expected, according to figure 24.11b. That is, a service lifetime of 28,000 hours or greater than 3½ years is expected with a relative failure rate of less than 1 per cent.

Generally, between 40 and 85°C aluminium electrolytic capacitor lifetime doubles for every 10°C decrease in ambient temperature. A service lifetime of 7 years could be obtained by decreasing the ambient temperature from 65°C to 55°C.

♣

With aluminium electrolytic capacitors, degradation failures are mostly due to factors such as

- aggressiveness of the acidic electrolyte
- diffusion of the electrolyte
- material impurities.

24.2.5ii - Solid, oxide capacitors

The failure rate of solid aluminium and tantalum capacitors is determined by the occurrence of open and short circuits (the dominant failure mode for solid tantalum capacitors) as a result of dielectric oxide layer breakdown or field crystallisation. In general, for a given oxide operating at rated conditions, liquid capacitors have a shorter lifetime than the corresponding solid type. Solid aluminium capacitors are more reliable than solid tantalum types and failure is usually the degradation of leakage current rather than a short circuit.

In comparison with liquid electrolytic capacitors, solid types, and, in particular, tantalum type capacitors, have a number of desirable characteristics:

- higher capacitance per unit / volume due to the higher permittivity of Ta_20_3 and the intrinsically high effective area per unit volume due to the sintered construction
- changes in parameters (C_R, tan δ) are less because the specific resistance of $Mn0_2$ and hence temperature coefficient, is lower than that of liquid electrolytes
- electrolyte is stable, does not evaporate or corrode.

The failure rate of all capacitors can be improved by decreasing the stress factors such as temperature and operating voltage. But reliability of solid tantalum capacitors can be increased by placing a series resistor (of low inductance) in the circuit. The improvement is illustrated by the following design example, which compares the lifetime of both liquid and solid tantalum capacitors based on the conversion curves in figure 24.12.

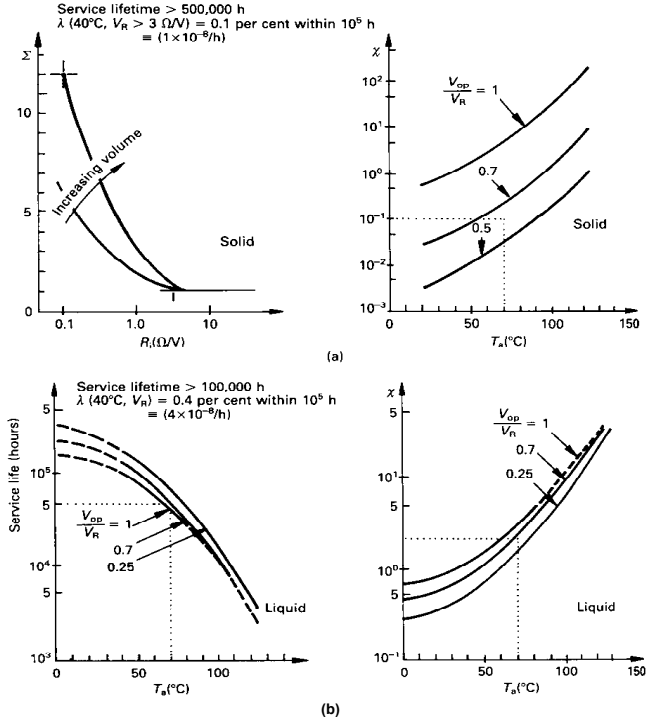


Figure 24.12. *Stress conversion factors for: (a) solid tantalum capacitors and (b) liquid tantalum capacitors.*

Example 24.5: ***Lifetime of tantalum capacitors***

A 22 μF tantalum capacitor is required to operate under the following conditions:
ambient temperature T_a, 70°C
operating voltage V_{op}, 15 V
circuit resistance *i.* 1 Ω
ii. 100 Ω

Calculate the expected lifetime for solid and liquid tantalum capacitors.

		Liquid tantalum	Solid tantalum	
R	Ω	1 and 100	1	100
R_i	Ω	n/a	0.1	3
ΣR_i		(1)	12	1
χ at V_{op}/V_R=0.6 and 70°C		2.2	0.10	0.10
λ_o	/h	4×10^{-8}	1×10^{-8}	1×10^{-8}
$\lambda = \lambda_o \times \Sigma$	/h	$2.2\times4\times10^{-8}$ 8.8×10^{-8}	$12\times0.1\times10^{-8}$ 1.2×10^{-8}	$1\times0.1\times10^{-8}$ 0.1×10^{-8}
fit		88	12	1
τ (% failures) within $\lambda\Delta t$	h	45,000 (0.4%)	83,000 (0.1%)	100,000 (0.1%)

Solution

Capacitor used C_R = 22 µF
V_R = 25 V
For each capacitor type (solid or liquid) the voltage stress factor is
V_{op}/V_R = 0.60
For the solid tantalum, the circuit resistance factor is given by

i. R_i' = 1 Ω /15 V = 0.07 Ω/V which is < 0.1 Ω/V

ii. R_i' = 100 Ω /15 V = 6.6 Ω/V which is **>** 3 Ω/V

Based on figure 24.12, the capacitor lifetime calculation is summarised the previous table.

♣

24.3 Plastic film dielectric capacitors

Plastic (polymer) dielectric type capacitors are non-polarised capacitors and in general offer high *dv/dt* and pulse rating capability compared with oxide type capacitors.
The most common dielectric plastics used (organic, hydrocarbons, as shown in figure 24.13) are:

polye ***t*** hylene-terephthalate (polyester or PEPT or PET)	T
poly ***c*** arbonate (now obsolete)	C
poly ***p*** ropylene (PP)	P
poly ***s*** tyrene	S
polyphenylene sulph ***i*** de	I
polyethylene ***n*** aphthalate (PEN)	N

The letter shown after each type is the symbol generally used to designate the film type. The symbol K is used to designate plastic, which is *Kunststoff* in German.

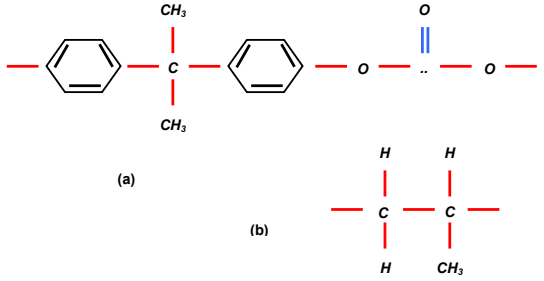


Figure 24.13. *Basic hydrocarbon structure of (a) polyethylene-terephthalate and (b) polypropylene.*

Two basic types of plastic film dielectric capacitors are common. The first type involves *metallisation* deposited on to the plastic and the metal forms the electrodes. Typically such a capacitor would be termed MKP, that is metallised - M, plastic - K, polypropylene - P. A foil capacitor, the second type, results when interleaved and displaced metal foil is used for the electrode. Typically such a foil capacitor would be termed KS, that is plastic - K, polystyrene – S or MFT and MFP (F - foil). The plastic type is designated by the fifth letter of the plastic name, that is the letter after poly, with two exceptions.

Table 24.3: Characteristics of plastic film dielectrics – typical values

dielectric			PEPT	PC	PP	PS	PEN
Dielectric constant, 25°C, 1kHz	ε_r		3.2	2.8	2.2	2.5	3.0
Dielectric strength		V/µm	250	180	350	150	
C drift with time	$\Delta C/C$	%	3		3		2
C temperature coefficient	α_c	10^{-6}/K	+600	+150	-250	-150	+200
Max working temperature		°C	125	125	105	70	
C humidity coefficient (50% …95%)	β_c	10^{-6}/% r.h.	500 … 700		40 … 100		700 … 900
Dissipation factor @ 1kHz	$tan\delta$	10^{-3}	5	1	0.6	0.2	4
Time constant, 25°C	$\tau=R_iC_R$	10^3 s	25		100		25
Water dielectric absorption		% weight	0.2	0.3	0.05	0.1	1.2
density		g/cm^3	1.39	1.21	0.91	1.05	

26.3.1 Construction

24.3.1i - Metallised plastic film dielectric capacitors

The dielectric of these capacitors consists of plastic film on to which metal layers of approximately 0.02-0.1 µm are vacuum deposited. A margin of non-coated film is left as shown in figure 24.14a. The metallised films are either wound in a rolled cylinder or flattened to form a stacked block construction. In this construction, the metallised films are displaced so that one layer extends out at one end of the roll and the next layer extends out the other end as shown in figure 24.14a. This displaced layer construction is termed *extended metallisation* and facilitates electrical contact with the electrodes. A hot lead-free metal spray technique, called s*chooping,* is used for making electrical contact to the extended edges of the metallised plastic winding. This large disk area contact method ensures good ohmic contact, hence low loss and low impedance capacitor characteristics result.
The most common metallised plastic film capacitors are those employing polyester, MKT and polypropylene, MKP. All have self healing properties, hence use thinner dielectrics.
Polyester has a higher dielectric constant than polypropylene, and because of its stronger physical characteristics, is available in thinner gauges than polypropylene. High capacitance values result in the smallest possible volume. But polypropylene has a higher dielectric strength and lower dielectric losses, hence is favoured at higher ac voltages, currents, and frequencies.

24.3.1ii - Foil and plastic film capacitors

Foil capacitors normally use a plastic film dielectric which is a flexible bi-axially aligned electro-insulator, such as polyester. Aluminium foils and/or tin foils (5 to 9µm) are used as the electrodes. The thin strips are wound to form the capacitor as shown in figure 24.14b. An *extended foil* technique similar to the extended metallisation method is used to enable contact to be made to the extended metal foil electrodes. Film foil capacitors do not have self-healing properties, hence use thicker dielectrics.

24.3.1iii - Mixed dielectric capacitors

To further improve the electrical stress capabilities of a capacitor, combinations of different dielectrics are commonly used. Such capacitors use combinations of metallised plastics, metallised paper, discrete foils and dielectrics, and oil impregnation.
Figure 24.14c shows the layers of a mixed dielectric paper and polypropylene capacitor. A thin gauge of polypropylene dielectric is combined with textured metallised paper electrodes. The coarse porous nature of the paper allows for improved fluid impregnation of the dielectric material, which counters the occurrence of gas air bubbles in the dielectric. This construction has the electrical advantages of high dielectric strength, low losses, and a self-healing mechanism, all at high voltages.
Two plastic dielectrics can be combined, as shown in figure 24.14d, to form a *mixed layer* capacitor, with two series connected elements. It involves a double metallised polyethyleneterephthalate film and polypropylene films. These dielectric combinations give low inductance, high dielectric strength, and low losses with high ac voltage capability.
Other extended layer winding designs, involving two internally series connected elements as in figure 24.14d, but single sided, are extended metallised film and extended foil with a metallised film designs.

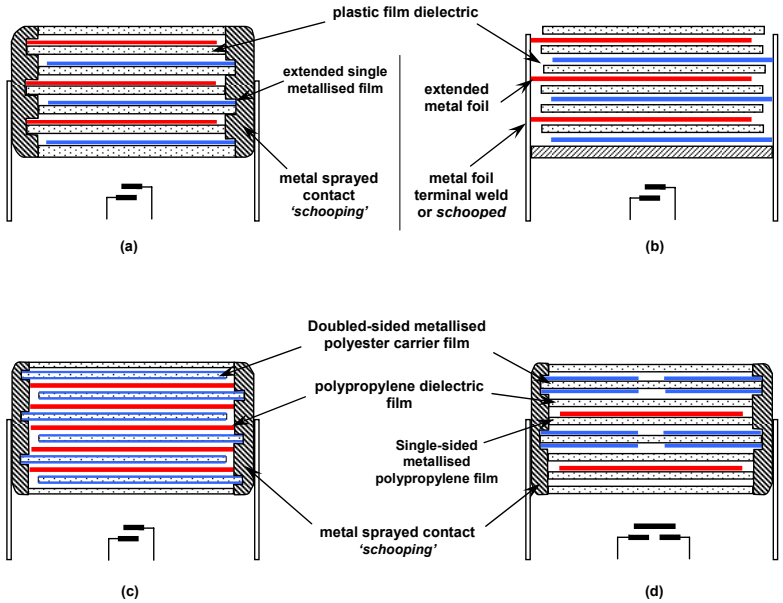


Figure 24.14. *Plastic capacitor constructions: (a) extended single metallization film; (b) extended foil; (c) mixed dielectric, extended double-sided metallised carrier film; and (d) mixed dielectric, extended double-sided metallised carrier film with internal series connection.*

24.3.2 Insulation

The dc resistivity insulation characteristics of a capacitor are indicated either as a resistance value R_i, as shown in Figure 24.2 or for capacitance greater than 0.33μF, as a time constant, $\tau = R_i C_R$. The resistance comprises the insulation resistance of the dielectric (layer to layer) and the insulation resistance between layer and case. This later resistance is determined by the quality of the case insulating material and by the length of the surface leakage paths.
Both the time constant and resistance are dependent on voltage, temperature, and significantly himidity, as is shown in figure 24.15. These characteristics illustrate that extremely high insulation resistance values can be obtained.

24.3.3 Electrical characteristics

24.3.3i - Temperature dependence

The capacitance of plastic film capacitors changes with temperature, humidity, and frequency, as shown in figure 24.16. The dependence is strongly dependent on the dielectric film although some foil types are virtually independent of frequency. Table 24.4 summarises capacitance temperature dependence for a range of dielectrics. The temperature coefficient α_c is measured in parts per million per degree Kelvin, ppm/K.

$$\alpha_c = \frac{C_{T2} - C_{T1}}{C_{20°C} \times (T_2 - T_1)} \qquad (24.33)$$

where C_{T1} and C_{T2} are the measured capacitances at temperatures T_1 and T_2. (See equation (24.23))
Any small irreversible charge in capacitance at rated temperature, after a temperature variation between the allowable temperature extremes, is termed *temperature cyclic capacitance drift*.
The temperature dependence of dissipation factor is shown in figure 24.23a.

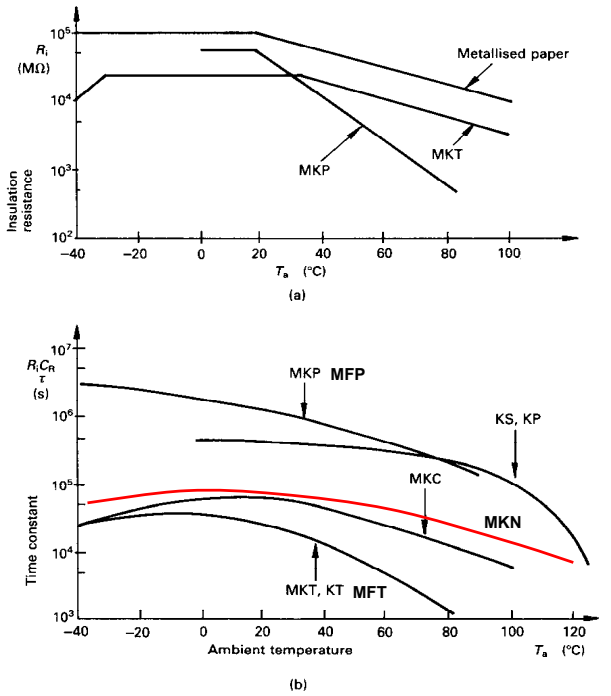


Figure 24.15. *Plastic dielectric insulation resistance temperature dependence characteristics: (a) resistance R_i and (b) time constant τ.*

24.3.3ii - Humidity dependence

The capacitance will undergo a reversible change in value due to ambient humidity variation. The *humidity coefficient* β_c, is define for a 1% change in humidity, at a constant temperature, by

$$\beta_c = \frac{2 \times (C_{F2} - C_{F1})}{(C_{F2} + C_{F1}) \times (F_2 - F_1)}$$

where C_{T1} is the capacitance at relative humidity F_1
C_{T2} is the capacitance at relative humidity F_2
Wide capacitance variations occur at relative humidity levels above 85%. Prolong contact of a film capacitor with high humidity or direct liquid water will produce irreversible effects due to reaction of the film metallisation. Typical plastic dielectric humidity coefficient variations are shown in figure 24.16c.

24.3.3iii - Time dependence

Capacitance charges irreversibly with time, where the drift coefficient i_z=|ΔC/C| is measure over a period of at least a year and at a temperature above ambient, typically a maximum of 40°C. As shown in Table 24.3, typical drift values are about 2% to 3%.

24.3.3iv - Dissipation factor and impedance

Figure 24.17a shows the typical frequency dependent characteristics of the dissipation factor for a range of plastic dielectric capacitor types. It is important to note that polyester types have 50-100 times the losses of polypropylene capacitors. A low loss characteristic is important in power pulse applications where capacitor package heat dissipation may be a limiting factor, as indicated in figure 24.18.
The dissipation factor, hence losses, are dependant on the *ESR*, as shown in equation (24.9). The *ESR* represents a complex set of loss mechanisms, many of which are strongly dependent on the measurement conditions. The *ESR* measured at the resonant frequency is not the worst-case value and is higher at lower frequencies.

Capacitor losses vary as a function of voltage, temperature, and other aspects of the applied waveform. This is because there are a variety of energy loss mechanisms which act within a capacitor. Some of these reside within the dielectric while others involve the conductors carrying the current. Some of the mechanisms and operating parameters which effect the magnitude of the losses, follow.

Dielectric losses are usually the most important losses in a film capacitor. These losses are associated with the polarization and relaxation of the dielectric material in response to the applied capacitor voltage. The magnitude of the dielectric losses in a capacitor are therefore generally both frequency and temperature dependent, when the largest losses occur at low temperatures or high frequencies, which hinder dipole orientation. Dielectric losses essentially are not voltage-dependent.
The dielectric losses of a given material can be described by its Dissipation Factor, *DF*. If the dielectric loss were the only loss mechanism operating in a capacitor, then the *DF* of the capacitor would be independent of its size, geometry and internal configuration. Capacitors of any size made with the same material would have the same *DF* under identical conditions. The *ESR* could then be computed using equation (24.9). However, the capacitor *DF* and *ESR* depend on the electrodes and their configuration.

Ferroelectric hysteresis losses are observed in certain high dielectric constant materials, most notably ceramics. These losses are a strong function of applied voltage. This loss mechanism arises when the internal polarization field has the same order of magnitude as the applied field. Under these conditions the dielectric response saturates. Capacitors made with such materials exhibit permanent polarization, variable capacitance as a function of voltage, and sensitivity to reversals of voltage.

Dielectric conduction losses are caused by the actual transport of charge across the volume of the dielectric or across internal dielectric interfaces. These losses are largest at low frequencies and higher temperatures. Because conduction in a dielectric material can be strongly nonlinear, non-Ohmic, conduction losses are often strongly dependent on applied capacitor voltage.

Interfacial polarization losses are related to dielectric conduction. Many high voltage capacitors contain two or three different materials within their dielectric arrangements, film and oil or paper, film and oil. Each material has different conduction properties and permittivity. As a result, the application of a DC voltage over a period results in a build-up of conducted charge at the internal interfaces between materials. This polarization of the dielectric is a low frequency phenomenon, hence the energy stored in this way is not available for discharge at high frequency. Again, since the conduction is nonlinear, interfacial polarization is nonlinearly voltage dependent.
This loss mechanism is important in pulse discharge applications, where the capacitor is charged over a relatively long period of time and then discharged rapidly.

Partial discharge losses can occur within gas-filled or defective solid capacitors or even in liquid-filled capacitors at high voltages. It is also common to have external corona on capacitor terminals. Partial discharges are most energetic at high rates of change of voltage (high *dv/dt*), such as during a capacitor pulse discharge. Also, reversal of the voltage such as in a highly oscillatory ringing capacitor discharge will cause more numerous, energetic, partial discharges.

Electromechanical losses result from the electrostriction, and sometimes piezoelectricity, acting within the capacitor dielectric itself and the flexing of internal wiring due to the Lorentz forces.

Ohmic resistance losses occur within the metallic electrodes, the internal wiring, and the capacitor terminals. In electrolytic capacitors, ohmic resistance in the electrolyte itself represents the largest loss mechanism. The resistance losses in the metal are constant as a function of temperature and frequency, until the skin depth in the electrodes becomes important, usually above several megahertz. Losses in the internal wiring and the terminal is important in high current applications. When high voltage capacitors are internally configured as a series string of lower voltage capacitor windings or units, the ohmic resistance within a given container size increases at the square of the voltage (or as the number of series elements).

Sparking can occur between conductors or different points on the same conductor during the discharge of pulse capacitors. For example, capacitors manufactured with an inserted tab connection to the electrode foil which is only a pressure contact exhibit points of localized melting after pulse discharge operation resulting from sparks between the adjacent metallic surfaces. This phenomenon is related to a high current rate of change, *di/dt,* during discharge, and is therefore frequency and voltage related.

Eddy current losses are important in pulse forming networks where a high magnetic field can couple into any ferromagnetic materials within the capacitor. These losses depend strongly on frequency. Usually the internal inductance in a capacitor is small and does not generate significant eddy currents.

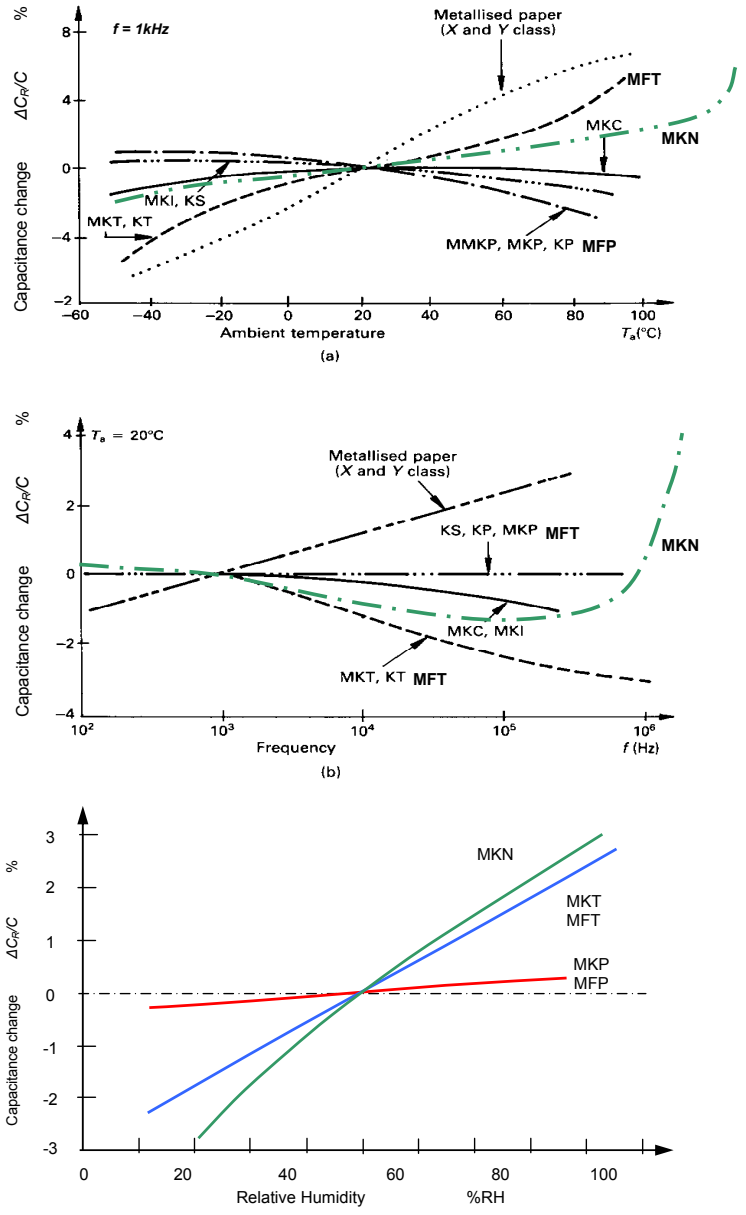


Figure 24.16. *Plastic film dielectric capacitance variation with: (a) ambient temperature; (b) frequency; and (c) relative humidity.*

Generally, tanδ rises with increased frequency and increased capacitance. Tanδ is dominated by dielectric losses and the contact resistance of the leads. The extended foil/metallisation and schooping contact methods provide not only a low and constant ohmic contact, but because of the large contact area, result in a low self-inductance. The resonant frequency of such capacitors, because of their self-

inductance and their capacitance, is high as shown by the minimum impedance in figure 24.17b. Minimum impedance decreases with increased capacitance and each capacitor in the range, here 1.5 nF to 4.7 μF, has its own Y-shaped impedance curve. The self-resonant frequency decreases with increased capacitance and can be used to determine the ESL, by $\omega_o = \sqrt{LC_R}$. In figure 24.17b, the full impedance curves for maximum and minimum capacitance only have been shown.

Table 24.4: Capacitor temperature coefficient for various dielectric materials

Dielectric type	*Temperature coefficient α_c (ppm/K or 10^{-6}/K)*		
	metallised	*other*	*film/foil*
Polypropylene	-170		-120
Polyester	400		400 (non-linear)
Polyethylene naphthalate	180		160
Polycarbonate	150		-50 to -150
Polystyrene			-125
Paper	300		300
Mica		100	
Ceramic		+ 1000 to -1000	(non-linear)
Aluminium		1500	
Tantalum (solid and liquid)		+200 to +1000	

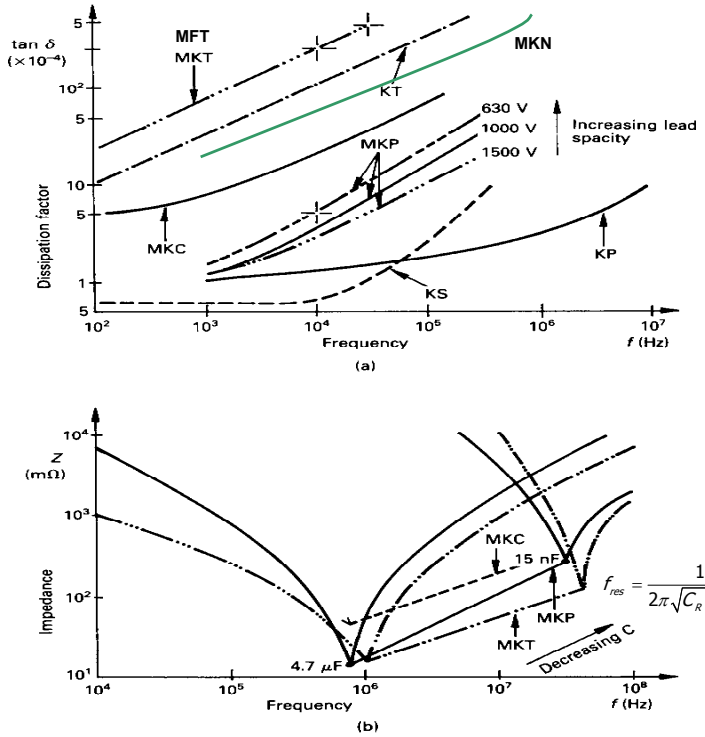


Figure 24.17. *Frequency characteristics for plastic dielectric capacitors: (a) maximum dissipation factor, tan δ and (b) typical impedance characteristics, Z, for metallised plastic dielectric capacitors.*

***24.3.3v -** Voltage derating with temperature*

The ac and dc voltages which may be applied continuously to a capacitor vary with ambient temperature and also frequency in the case of ac voltage rating. Typical characteristics showing temperature and frequency dependence are shown in figures 24.18 and 24.20 for plastic dielectric capacitor types. It will be seen that the ac voltage rating is significantly less than the dc voltage rating, while both voltage ratings are derated above 85°C and at higher frequencies. In all situations, the sum of the dc voltage and peak value of superimposed ac voltage must not exceed the rated dc voltage.
An alternative approach for calculating the maximum ac voltage, allowable V_{ac}, for a capacitor is based on the power dissipation limits, P, of the package.
If R_l and ESL are neglected in the capacitor equivalent circuit shown in figure 24.2, then

$$P = \frac{V_{R_s}^2}{R_s} = I^2 R_s \qquad \text{(W)} \tag{24.34}$$

and

$$V_{R_s}^2 = \frac{R_s^2}{R_s^2 + \dfrac{1}{\omega^2 C_R^2}} V_{ac}^2 \tag{24.35}$$

Since from equation (24.12) for plastic dielectric capacitors

$$\tan\delta = \omega C_R R_s$$

then equation (24.34) can be written as

$$P = \omega C_R \times \tan\delta \times V_{ac}^2 = \left(R_s C_R\right) \omega^2 C_R V_{ac}^2 \qquad \text{(W)} \tag{24.36}$$

or alternatively

$$P = \tan\delta \;\; \omega C_R V_{ac}^2 \qquad (= I_{rms}^2 ESR) \;\; \text{(W)} \tag{24.37}$$

The value of tan δ for equation (24.37) is available from figure 24.17a or, alternatively, the value of $R_s C_R$ for equation (24.36) is available from figure 24.19.
The equivalent series resistance is dominated by leakage and dielectric losses (f^{-1}) at low frequency. At medium frequencies, conductor losses dominate, while at high frequencies losses are dominated by the skin effect ($\sqrt{f}$), as shown in figure 24.19b.
The maximum permissible power dissipation, $\hat{P}$ which depends on the package dimensions and ambient temperature, is given in figure 24.18d. Thus when the power dissipation, for a given ac voltage, has been calculated, figure 24.18d can be used to specify the minimum size (dimensions) capacitor capable of dissipating that power.
The example 24.6 illustrates the design approach outlined.

***24.3.3vi –** Voltage and current derating with frequency*

The ac voltage/current dependence on frequency for film capacitors, has three distinct regions, as shown in figure 24.20.

Region A Below a certain frequency, f_1, the voltage threshold for corona discharge in capacitor air pockets is a limiting factor.

Region B In the mid-frequency region, the package power dissipation limit restricts the internal loss limit. The internal losses are tanδ dependant (equation (24.37)) while the package limit is surface area, A and heat transfer coefficient dependant, h, equation 5.4.
That is, the internal generated losses are

$$P_{\text{int}} = \tan\delta \;\; \omega C_R V_{ac}^2$$

while the heat dissipation from the capacitor surface A with a heat coefficient h, is defined by

$$P_d = hA\Delta T$$

Since the internal losses must be less than that that can be dissipated, for a given internal temperature self-heating, ΔT:

$$V_{ac} \le \sqrt{\frac{hA\Delta T}{\omega C_R \tan\delta}} \quad \text{or} \quad I_{ac} \le \sqrt{\frac{\omega C_R \times hA\Delta T}{\tan\delta}}$$

The voltage and current are approximately related to frequency by

$$\hat{V}_{ac} \propto f^{-3/4} \;\; \text{or} \;\; \hat{I}_{ac} \propto f^{1/4}$$

Region C At higher frequencies, above f_2, with smaller capacitances and short contact lengths, the ac voltage is limited by the maximum current capabilities I_2 of schooped plating connections.

$$V_{rms} \le \frac{I_c}{2\pi f \times C_R} \quad \text{or} \quad I_{rms} \le I_c$$

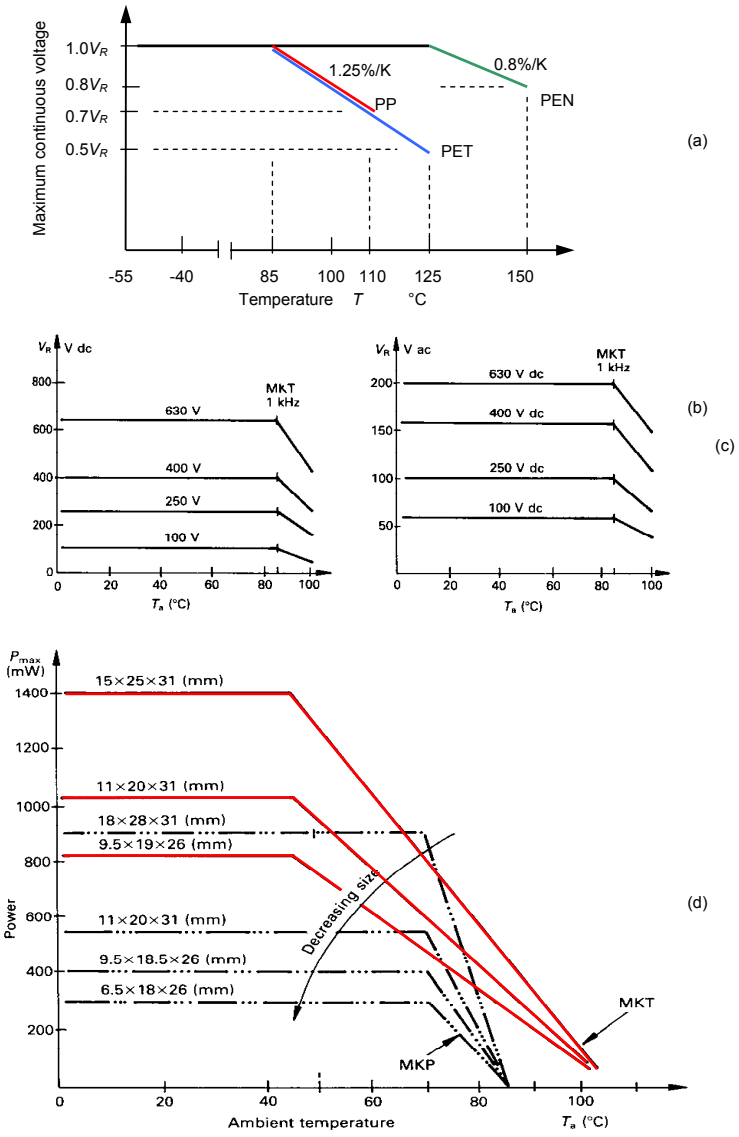


Figure 24.18. *Plastic dielectric capacitor, temperature derating characteristics: (a) general dc voltage derating; (b) dc voltage derating with ambient temperature; (c) ac voltage derating with temperature; and (d) power derating with temperature as a function of capacitor dimensions.*

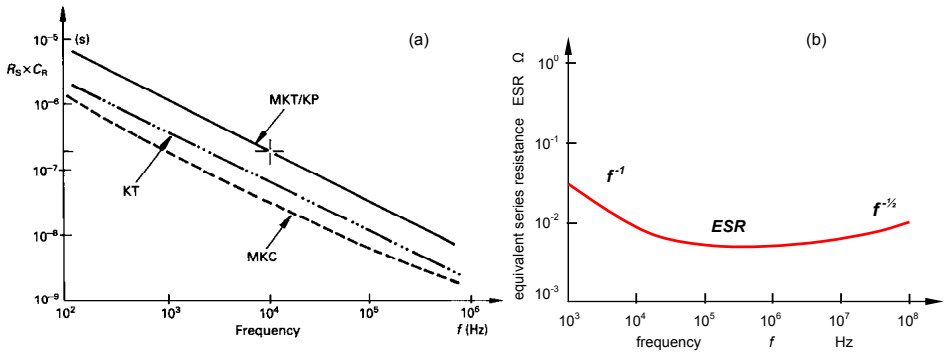


Figure 24.19. *ESR characteristics: (a) maximum product of series resistance, R_s, and rated capacitance, C_R, and (b) ESR as a function of frequency.*

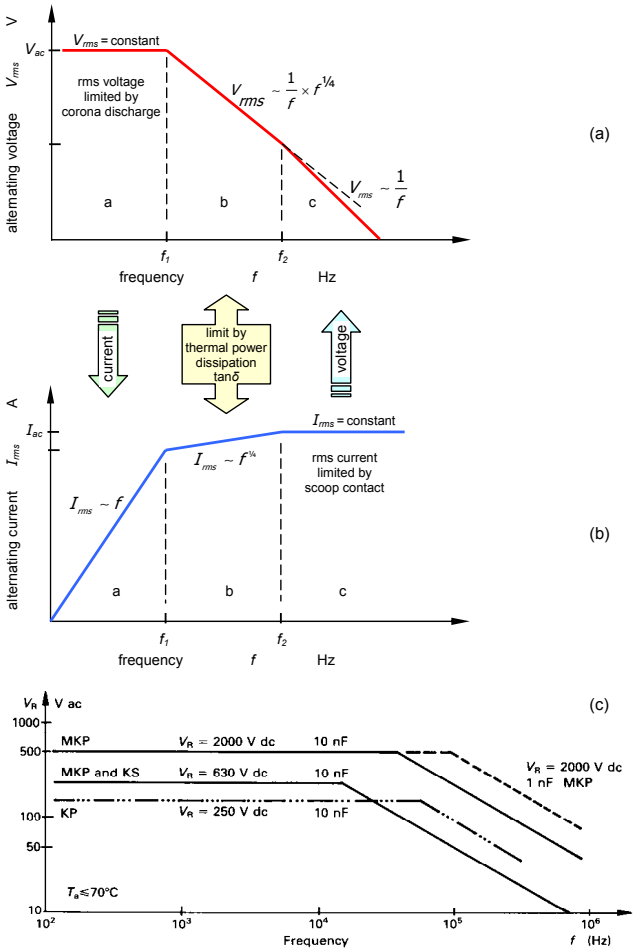


Figure 24.20. *Capacitor rating limits: (a) maximum ac voltage; (b) ac current; and (c) ac voltage derating with frequency of different metallised plastic capacitors.*

Example 24.6: ***Power dissipation limits - ac voltage***

A 0.1 μF plastic capacitor is used in a 100 V ac, 10 kHz and 50°C ambient application. Select suitable metallised polypropylene and polyester capacitors for this application.

Solution

i. Metallised polyester capacitor (MKT)
From equation (24.36)

$$P = (R_s C_R)\,\omega^2 C_R V_{ac}^2 \qquad \text{(W)}$$

From figure 24.19a, $R_s C_R = 2 \times 10^{-7}$ at 10 kHz. Thus

$$P = (2\times10^{-7}) \times (2\pi\times10^{4})^2 \times (0.1\times10^{-6}) \times (100)^2$$
$$= 780\ \text{mW}$$

From figure 24.18d, at 50°C a MKT capacitor of dimensions 11×20×31 (mm) can dissipate 930 mW. The applicable capacitor must have an ac voltage rating in excess of 100 V ac. From figure 24.20c, it can be seen that a 0.1 μF, 400 V dc MKT capacitor is necessary, given that the dimension constraints are met.

ii. Metallised polypropylene capacitor (MKP)
From equation (24.37)

$$P = \tan\delta\ \ \omega C_R V_{ac}^2 \qquad \text{(W)}$$

From figure 24.17a, tan $\delta = 4.0 \times 10^{-4}$ at 10 kHz, for a 600 V dc type. Thus

$$P = (4.0\times10^{-4}) \times (2\pi\times10^{4}) \times (0.1\times10^{-6}) \times 100^2$$
$$= 25.6\ \text{mW}$$

From figure 24.18d, at 50°C, the smallest volume MKP capacitor, of dimensions 6.5×15×26 mm, can dissipate 300 mW. From figure 24.20c it can be seen that a 0.1 μF, 630 V dc (250 V ac) MKP capacitor is necessary.
From figure 24.20c it can be seen that a 250 V dc 0.1 μF polypropylene foil capacitor (KP) is capable of 160 V ac at 10 kHz. Figure 24.17a shows the dissipation factor of KP type capacitors to be under half that of the metallised equivalent. That is, the expected losses are only

$$P = (1.4\times10^{-4}) \times (2\pi\times10^{4}) \times (0.1\times10^{-6}) \times 100^2$$
$$= 9\ \text{mW}$$

♣

24.3.3vii - *Pulse dV_R/dt rating*

Related to the ac voltage rating and power handling capabilities of a capacitor is the rated pulse slope dV_R/dt, which from $i = C_R\, dv/dt$ is specified by

$$R = \frac{V_R}{C_R\, dV/dt_{\max}} = \frac{V_R}{\hat{I}} \qquad (24.38)$$

where *R* is the minimum series resistance including the ESR. The rating test is an accelerated test, carried out for 10,000 pulses at a 1Hz repetition rate. The capacitor is then *dv/dt* rated at 10% of that at which the pulse test was performed.

Generally for a given C_R, *dv/dt* capability increases with rated voltage V_R, and decreases as the distance between the metallised electrode contacts increases. If the capacitor operating voltage V_{op} is decreased below V_R, at which voltage, *dv/dt* capability is specified, *dv/dt* capability increases according to

$$\frac{dV_{op}}{dt} = \frac{dV_R}{dt} \times \frac{V_R}{V_{op}} \qquad \text{(V/s)} \qquad (24.39)$$

The *dv/dt* capability depends on both the dielectric type and layer construction. Generally polystyrene (KS) and polyester (KT) foil type capacitors are not applicable to high *dv/dt* applications. Metallised polycarbonate capacitors (diminishing availability) offer slightly better *dv/dt* properties than those of metallised polyester. Metallised paper capacitors can withstand high levels of *dv/dt,* 30-50 times higher than those for metallised polyester. Capacitors using polypropylene, or even better a mixed dielectric involving polypropylene, offer extremely high *dv/dt* capability. With the construction shown in figure 24.14d, a 1 μF metallised polypropylene capacitor with V_R of 2000 V dc and 1000 V ac, a 2500 V/μs capability is attainable. Practically the *dv/dt* limit may be restricted by the external connections. Such ratings are obtainable with polypropylene because of its extremely low losses, tan *δ*, as indicated in figure 24.17a. Under such high *dv/dt* stresses, it is important to ensure that the power dissipated (whence rms current) does not exceed the package limit.

24.3.4 Non-sinusoidal repetitive voltages

Capacitors used for repetitive transient suppression, and for turn-off snubbers on GTO thyristors and diodes, experience high-magnitude short-duration voltage and current pulses which are not sinusoidal. High *dv/dt* capacitors based on metallised polypropylene are used, which are limited by their internal power losses, hence temperature rise and package power dissipation limit.
A restrictive graphical design approach for capacitor selection with sinusoidal, sawtooth, and trapezoidal pulse trains is shown in figure 24.21. The design approach is illustrated by example 24.7.

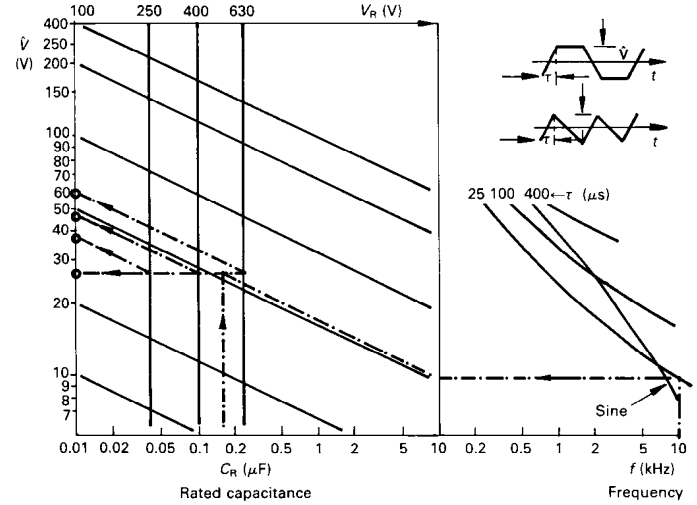


Figure 24.21. *Metallised polyester capacitor selection graph for sinusoidal and non-sinusoidal voltages.*

When capacitor voltages are more abstract, the concept of pulse characteristic can be applied for frequencies lower than 10kHz and low duty cycles, which is based on the internally generated heat.

$$W = \int_o^\tau I^2 R_i dt = \int_o^\tau C^2 \left(\frac{dV}{dt}\right)^2 \times R_i dt = ½ C^2 \times R_i \times k_o$$

Thus k_o, related the heat energy and pulse slope, is defined as

$$k_o = 2\int_o^\tau \left(\frac{dV}{dt}\right)^2 dt \qquad \text{V}^2/\text{s} \qquad (24.40)$$

where *τ* is the pulse width and R_i is the effective internal resistance, dominated by the contact resistance.

For a ramp voltage change

$$\frac{dV}{dt} \approx \frac{V_{pp}}{\tau}$$

such that equation (24.40) gives

$$k_o \cong 2\frac{V_{pp}^2}{\tau} \quad \text{and} \quad \frac{dV}{dt} \cong \frac{V_{pp}}{\tau} = \frac{k_o}{2V_{pp}} \qquad (24.41)$$

This equation (24.41) shows that for a given capacitor, that is, a given k_o, the lower the source peak to peak voltage, the higher the allowable *dv/dt*, hence higher peak current.

For a passive RLC type discharges and short-circuit configurations, k_o is

$$k_o = \frac{V_{ch}^2}{RC}$$

where *R* is the discharge circuit resistance and V_{ch} is the source charging voltage.

Example 24.7: ***Capacitor non-sinusoidal voltage rating***

A 0.15μF MKT capacitor is used to generate a 10 kHz maximum and 25μs rise-time minimum, saw-tooth ac voltage waveform. What voltage rated capacitor is applicable if the output voltage maximum is 100V p-p?

Solution

Worst-case conditions are at maximum frequency, 10 kHz, and minimum risetime, 25 μs.
With reference to figure 24.21, use
f = 10 kHz (repetition frequency)
τ = 25 μs (rise-time)
C = 0.15 μF (capacitance)
According to the dashed line in figure 24.21, starting from f = 10 kHz, yields
V_R = 100 V dc gives maximum peak voltage of 27 V
V_R = 250 V dc gives maximum peak voltage of 38 V
V_R = 400 V dc gives maximum peak voltage of 47 V
V_R = 630 V dc gives maximum peak voltage of 59 V
The peak to peak requirement is 100 V, hence only a 630 V dc 0.1 μF MKT capacitor can fulfil the specification.

♣

An alternative approach to specify the voltage limits for non-sinusoidal repetitive voltages is to sum the power contribution due to each voltage harmonic. The total power due to all harmonics must not exceed the capacitor package power limits.
The non-sinusoidal voltage v can be expressed in the form

$$v = \sum_{V_i} V_i \sin\left(i\omega t + \phi_i\right) \qquad (24.42)$$

where V_i is the magnitude of the *ith* voltage harmonic, which has an rms value of

$$v_i = \frac{V_i}{\sqrt{2}}$$

From equations (24.12) and (24.36), assuming capacitance is frequency independent

$$P_i = \left(R_s C_R\right)_i \omega_i^2 C_{R_i} v_i^2 \qquad (24.43)$$

or

$$P_i = \tan\delta_i \; \omega_i C_{R_i} v_i^2 \qquad (24.44)$$

The total power dissipated is the sum of the powers associated with each frequency. The near-linear frequency dependence of tan δ and $R_S C_R$, as shown in figures 24.17a and 24.19, may be utilised to simplify the calculation procedure. Assuming the rated capacitance is independent of frequency may be a valid and helpful simplification, while the temperature dependence of C_R initially could be accounted for by using a value at 10 K above ambient.

Example 24.8: ***Capacitor power rating for non-sinusoidal voltages***

The applied voltage across a 1 μF MKP capacitor, at 40°C ambient is $\sqrt{2}$ 100 sin($2\pi \times 10^4 t$) + $\sqrt{2}$ Y sin($2\pi \times 3 \times 10^4 t$)
What is the maximum allowable third harmonic voltage Y?

Solution

From equation (24.44), the total power is given by

$$P_t = \tan\delta_1 \; \omega_1 C_{R_1} v_1^2 + \tan\delta_3 \; \omega_3 C_{R_3} v_3^2$$

From figure 24.16b we may assume that capacitance is independent of frequency for polypropylene types. From figure 24.16a, at 50°C, rated capacitance has reduced by only 1 per cent - thus temperature effects on C_R may be neglected.
From figure 24.17a, for a 600 V MKT capacitor
tan δ_1 at 10 kHz (ω_1) = 2.5×10^{-4}
tan δ_3 at 30 kHz (ω_3) = 4.2×10^{-4}
From figure 24.18d it can be seen that 880 mW can be dissipated in the largest package at 50°C.
Total power is given by

$$0.88\text{W} = 2.5\times10^{-4} \times 2\pi\times10^{4} \times 1\times10^{-6} \times 100^2 + 4.2\times10^{-4} \times 6\pi\times10^{4} \times 1\times10^{-6} \times Y^2 \quad \text{(W)}$$

Solving for Y, Y = 30.2 V rms.

♣

24.3.5 DC plastic capacitors

Significant increase in capacitance per unit volume per unit volt can be gained if plastic based capacitors are designed specifically for dc applications. Such applications include dc-link decoupling where electrolytic capacitors are avoided because of lifetime constraints and high-energy discharge capacitors; where voltage reversal is a critical parameter. Voltage reversal is the changing of the relative polarity of the capacitor terminals, such as experienced during ringing or oscillating pulse discharge, during ac operation, or as the result of dc charging the capacitor in the opposite polarity from which it had been previously dc charged.
Voltage reversal is the percentage of the peak voltage that is experienced in the reverse polarity. In an ac application, the reversal is 100 %. Oscillating pulse discharges usually have between 0% and 100% voltage reversal. DC capacitors are designed for the highest level of voltage reversal (normal or fault) that may be experienced in service. Adversely, high reversal ratings result in significant reductions in energy density and increases in size and cost.
The damage inflicted on a capacitor by a transient voltage reversal is a nonlinear function of the degree of reversal. As shown in Figure 24.22a, the change in life with between 80% and 85% reversal is much greater than the change between 20% and 30% reversal. The magnitude of the damage also depends on the rate of change of voltage during the reversal. The least deterioration is when the rate of change of voltage is slow, as in the case of dc charging the capacitor with the terminal connections reversed. The greatest damage occurs when the capacitor voltage 'rings' or oscillates at a high frequency, the effect of frequency on life is shown in Figure 24.22b.

Capacitor life is extended by minimizing the degree of voltage reversal in the normal operating mode. A diode and series resistance in parallel with the capacitor can reduce voltage reversal. The smaller the series resistance, the lower the reversal on the capacitor.

Operating mechanism during voltage reversal

i. Electric Fields
Voltage reversals impact on the electric field magnitude in the capacitor dielectric. Dielectric over-stresses result from the superposition of the applied reverse electric field and the remnant polarization field from the original dc polarity. At typical pulse capacitor discharge rates, the electronic, atomic, and permanent dipole polarizations reverse virtually in phase with the applied field. However, inherently 'slow' polarization mechanisms acting in the dielectric, such as interfacial polarization associated with charge injection and ionic conduction, do not. The longer the capacitor remains dc charged, the greater the remnant polarization field magnitude. This field (which is anti-parallel to the applied dc field) is added to the applied field during a voltage reversal, increasing the total field within the dielectric. Excessive fields can result in immediate breakdown or may produce partial discharging, treeing, or other degradation.
Even in ac applications, where interfacial polarization may not have time to build up, charge can be injected from the electrodes into the adjacent dielectric, especially at sharp edges, one half-cycle, and then return to the electrode in the next half-cycle in a partial discharge process. Such discharges degrade the dielectric locally and eventually result in breakdown. Therefore, long-life ac capacitor elements are designed to operate at voltage levels where such charge injection is negligible. DC capacitors, on the other hand, can usually be operated at much higher stresses and can therefore be made smaller.

ii. Heating
The current waveform is used to determine the internal heating of a capacitor due to various energy loss mechanisms.
The energy dissipated in the capacitor during a single charge/discharge cycle (J_{cap}) depends on the I^2t integral of the current waveform and the equivalent series resistance (ESR) of the capacitor. The action involves the Joule's integral:

$$J_{cap} = ESR \times \int_0^T I^2(t)\,dt \qquad \left[\text{Joules}\right]$$

The ESR is not a true ohmic resistance and is a function of frequency, voltage, voltage reversal, temperature, and other parameters. The ESR includes a number of energy loss mechanisms, the two most important of which, in terms of voltage reversal mechanisms, are dielectric loss and electrode resistance.

- The dielectric loss results from motion of bound charge within the dielectric (displacement current) such as molecular dipole rotation, in response to an applied electric field. It is the dissipation factor (DF), since the losses vary linearly with capacitance.
- The electrode resistance is purely ohmic, with the skin effect becoming important at high frequencies. There are two basic types of electrodes used in film capacitors, discrete foils and

metallisation. Foil electrode capacitors provide minimum ESR at high frequencies. Metallised electrodes can be used for relatively low frequency discharges (less than 10 kHz) where the ESR is dominated by the dielectric loss.

Capacitor design for voltage reversal

The critical aspects of the capacitor design in relation to voltage reversal effects are the dielectric materials, the rated voltage and rated electric field, and the type of electrode and internal connections. The resistance of different dielectrics to voltage reversal effects vary.

- A paper dielectric is suitable for high reversal discharge applications. Mixed dielectrics such as paper and polypropylene laminates are more susceptible to damage in voltage reversal than all-paper dielectrics because of interfacial polarization. All-polypropylene dc capacitors are highly susceptible to foil edge failure at moderate voltage reversal.
- Dry capacitors should be used in high reversal applications only at low voltage (less than 1kV) and low stress. A liquid impregnant should be incorporated in the dielectric to suppress partial discharges. Some impregnant materials are able to absorb gases and other decomposition products better than others.
- Metallised capacitors are more robust than foil capacitors in terms of ability to survive high reversal discharges without immediate failure. As long as ratings are not exceeded, high reversals simply accelerate the rate of capacitance loss. If peak current or ratings are exceeded, however, failure of the internal connections may occur, resulting in large capacitance loss, increased DF and ESR, and reduced voltage capability.

Capacitors designed to operate for long lifetimes (>10^7 pulses) at relatively low electric field stresses are more robust in the event of fault condition reversals, and capacitors being operated at well below their rated voltage are more likely to survive.

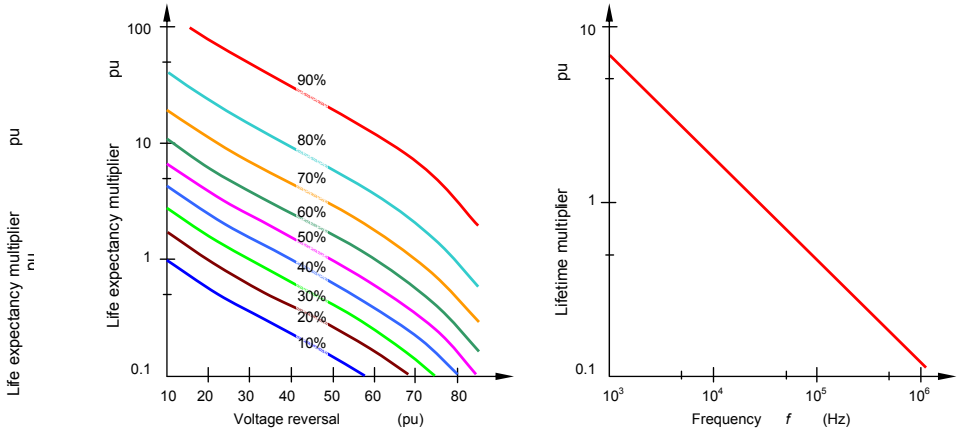


Figure 24.22. *Effects of: (a) dc capacitor voltage reversal on lifetime and (b) lifetime versus frequency of voltage reversal oscillation.*

The key properties of plastic type non-polarised capacitors are summarised in Table 24.5. The excellent dielectric properties of polypropylene lead to metallised polypropylene capacitors being extensively used in power applications.

Polycarbonate film based capacitors (KC and MKC) are obsolete. Mixed dielectric alternatives, based on polyethylene-terephthalate and polypropylene are recommended, but no alternative matches the excellent temperature and high frequency properties of polycarbonate.

Table 24.5: Properties of non-polarised plastic type capacitors

dielectric type	ε_r	tanδ	λ_o	*dv/dt*	self-healing
polypropylene	low	low	good	high	good
polyester	medium	high	poor	medium	good
polystyrene	low	low	good	high	poor
polycarbonate	low	medium	good	medium	good
mixed dielectric	medium	medium	good	medium	good
paper	high	high	very good	high	very good

24.4 Emi suppression capacitors

Non-polarised capacitors are used in rfi filters for electrical appliances and equipment, as was introduced in 10.2.4. The capacitors (7 classifications in all) used between line and neutral are termed class X while those used to earth are termed class Y.

24.4.1 Class X capacitors

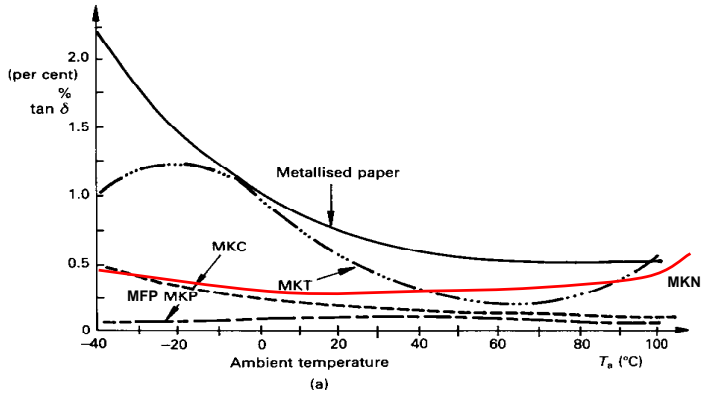
X capacitors are suitable for use in situations where failure of the capacitor would not lead to danger of electric shock. X capacitors, for 250V ac application, are divided into two subclasses according to the ac power line voltage applied.

- The X1 subclass must support a peak voltage in excess of 1.2 kV in service, while
- X2 capacitors have peak service voltage capabilities of less than 1.2 kV.

In order to obtain the peak voltage requirement of X1 capacitors, a construction comprising impregnated paper dielectric and metal foil electrodes is essential. The common X1 capacitance range is 10 nF to 0.2 µF. Class X1 is impulse tested to 4kV and 2.5kV for X2. Both are tested to higher voltages if $C \geq 1\mu F$.
The lower peak voltage requirement of X2 capacitors allows the use of a metallised plastic dielectric, of which polyester and polypropylene are common. Impregnated paper dielectrics may also be employed. Advantageously, metallised plastic film suppression capacitors yield high *dv/dt* capability with low associated losses, tan δ, as shown in figure 24.17a. These films also offer good insulation properties as shown in figure 24.15. Variation of capacitance with frequency and temperature is shown in figure 24.16, while percentage variation of losses, tan δ, with frequency and temperature is shown in figure 24.23. The typical capacitance range of X2 capacitors is from 10 nF to 1 µF, rated for 250 V ac application.

24.4.2 Class Y capacitors

Class Y capacitors are suitable for use in situations where failure of the capacitor could lead to danger of electric shock. These capacitors have high electrical and mechanical safety margins so as to increase reliability and prevent short circuit. They are limited in capacitance so as to restrict any ac current flowing through the capacitor, hence decreasing the stored energy to a non-dangerous level.
An impregnated paper dielectric with metal foil electrodes is a common construction and values between 2.5nF and 35nF are extensively used. Capacitance as low as 0.5nF is not uncommon.
A Y-class capacitor for 250V ac application can typically withstand over 2500V dc for 2s, layer to layer. On an ac supply, 425V ac ($\sqrt{3}\ V_R$) for 1000 hours is a common continuous ac voltage test. Class Y1 is impulse voltage tested to 8kV, and 5kV for Y2.
If *dv/dt* capability is required, polypropylene film dielectric Y-class capacitors are available, but offer lower withstand voltage capability than paper types. Generally paper dielectric capacitors offer superior insulation resistance properties, as shown in figure 24.15a.
Metallised paper capacitors are also preferred to metallised plastic types because they have better self-healing characteristics. Breakdown in metallised plastic film dielectrics causes a reduction of the insulation resistance because of a higher carbon deposit in the breakdown channel than results with paper dielectrics.



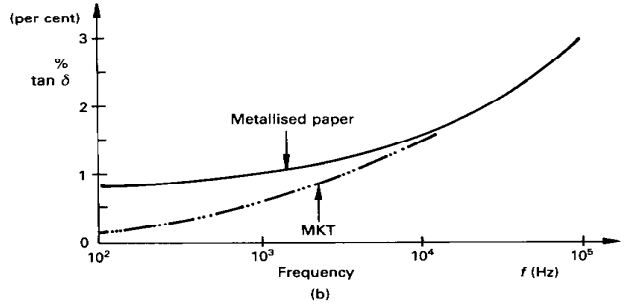
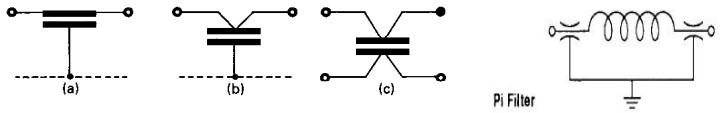


Figure 24.23. *RFI capacitance variation with:*
(a) ambient free-air temperature and (b) frequency.



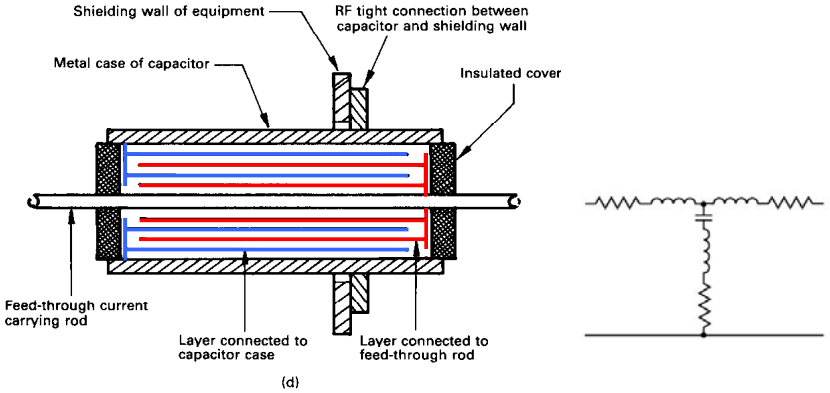


Figure 24.23. *Feed-through capacitors for RFI attenuation: (a), (b) three user terminals; (c) four terminals; and (d) coaxial feed-through capacitor construction.*

24.4.3 Feed-through capacitors

Feed-through or four-terminal capacitors are capacitors in which the operating current flows through or across the electrodes. High frequency rfi is attenuated by the capacitor and the main power is transmitted unaffected. That is they suppress emi penetration into or from shielded equipment via the signal or power path.
Figures 24.23a and b show three terminal feed-through capacitors while figure 24.23c is a four-terminal capacitor. A three-terminal coaxial feed-through, wound capacitor cross-section is shown in figure 24.23d. The feed-through rod is the central current-carrying conductor: the outer case performs the function of an electrode plate and connector to produce an RF seal between the capacitor case and shielding wall.
These capacitors are effective from audio frequencies up to and above the SW and VHF band (>300 MHz). Current ratings from signal levels to 1600 A dc, 1200 A ac are available, in classes X1 and X2, rated at 240 V ac, 440 V ac, and 1000 V dc. Class Y feed-through capacitors rated at 25 A and 440 V ac, 600 V dc are available.

Important note: *This section on emi-suppression capacitors does not imply those requirements necessary to conform to governmental safety and design standards.*

24.5 Ceramic dielectric capacitors

Ceramic capacitors as a group have in common an oxide ceramic dielectric. The dielectric is an inorganic, non-metal polycrystalline structure formed into a solid body by high temperature sintering at 1000 to 1300°C. The resultant crystals are usually between 1 and 100μm in diameter.
The basic oxide material for ceramic capacitors is titanium dioxide ($Ti0_2$) which has a relative permittivity of about 100. This oxide together with barium oxide ($Ba0_2$) forms barium titanate ($BaTi0_3$) which is a ferro-electric material with a high permittivity, typically 10^4. Alternatively, strontium titanate may be utilised. These same materials are used to make positive temperature coefficient resistors - thermistors, where dopants are added to allow conduction.
Metal plates of silver or nickel (with minimal palladium and platinum) are used to form the capacitor. Single plate, or a disc construction, is common as is a multi-layer monolithic type construction.
The ceramic dielectric is split into two classes, as shown in Table 24.6.

Table 24.6: Ceramic dielectric capacitor characteristics

Dielectric class			*I* (ε_r < 500) *Low K*	*II* (ε_r > 500) *Moderately high K*		*High K*
EIA-designation* IEC/CECC designation	see table 24.8		COG CG	X7R 2C1		Z5U 2F4
Temperature range		°C	-55 to 125	-55 to 125		+ 10 to 85
Dielectric constant	ε_r		13 - 470	700	to	50,000
Temperature coefficient of C_R (typical)			(N150) -150 ± 60 ppm	(X7R) ±15%		(Z5U) +22% / -56%
Dissipation factor	tan δ		0.15% @ 1 MHz	2.5%		3%
Capacitance	C	nF	< 0.2	< 4.7		< 40
Rated voltage	V_R	V	500-1k	100 to >2k		

* In EIA designation, first letter and number indicate temperature range while last letter indicates capacitance change.

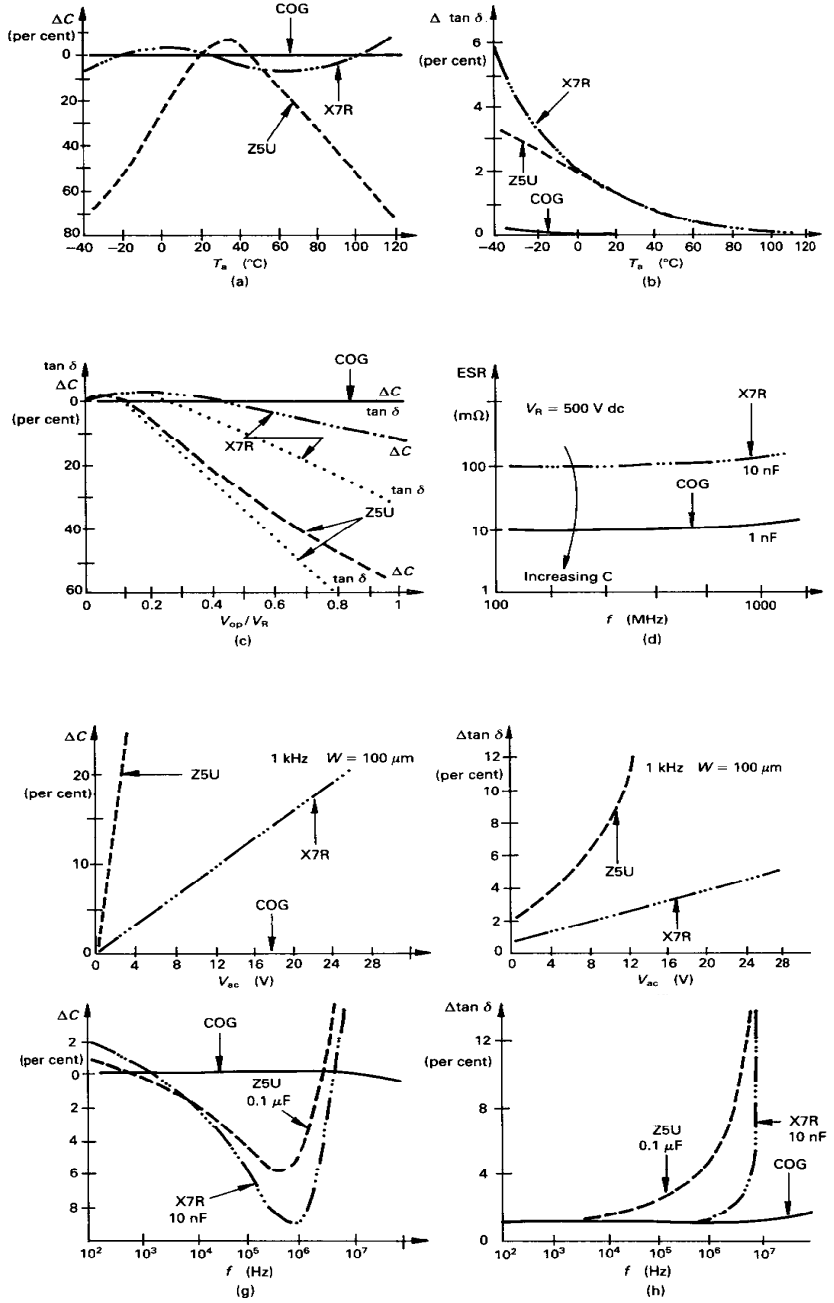


Figure 24.25. *Typical properties of commercial ceramic capacitors: (a) capacitance change with temperature; (b) dissipation variation with temperature; (c) capacitance change with dc voltage; (d) ESR change with frequency; (e) capacitance change with ac voltage; (f) dissipation factor variation with ac voltage; (g) capacitance change with frequency; and (h) dissipation factor variation with frequency.*

Table 24.7: Characteristics of class I and II type dielectrics

Class I	*Class II*
Almost linear capacitance/temperature function	Non-linear capacitance/temperature function
No voltage dependency of capacitance and loss angle	
No ageing	Slight ageing of capacitance
High insulation resistance	High insulation resistance
	Extremely high capacitance value per unit volume
Very small dielectric loss	
High dielectric strength	
Normal capacitance tolerance ±1% to ±10%	Normal capacitance tolerance ±5% to -20+80%

Table 24.8: Class II ceramic capacitor parameter coding – EIA designation

Dielectric class II (and higher) capacitor ceramic code					
Letter code	lower temperature	Number code	upper temperature	Letter code	ΔC over temperature range, ΔT
Z	+10°C	2	+45°C	A	±1.0%
				B	±1.5%
				C	±2.2%
		4	+65°C	D	±3.3%
Y	-30°C			E	±4.7%
				F	±7.5%
		5	+85°C	P	±10.0%
				R	+15.0%
		6	+105°C	S	±22.0%
X	-55°C			T	+22% to -33%
		7	+125°C	U	+22% to -56%
		8	+150°C		
		9	+200°C	V	+22% to -82%

24.5.1 *Class I dielectrics*

This class of dielectric consists mainly of $Ti0_2$ and additions of Ba0, La_20_3 or Nd_20_5, which provides a virtually linear, approximately constant and low temperature coefficient as shown in figure 24.25a.
COG [EIA or industry code alternative NP0] capacitors belong to the class I dielectrics and have a low temperature coefficient over a wide temperature range, as seen in Table 24.6. They provide stability and have minimum dissipation properties. In attaining these properties, a low dielectric constant results and these capacitors are termed *low K*. Because of the low dielectric constant, capacitance is limited.

24.5.2 *Class II dielectrics*

Ceramic capacitors in this class are usually based on a high permittivity ferroelectric dielectric, $BaTi0_3$, hence termed *hi K*. Large capacitance in a small volume can be attained, but only by sacrificing the temperature, frequency, and voltage properties, all of which are non-linear. Typical characteristics are shown in figure 24.25. Their characteristics are less stable, non-linear, and have higher losses than class I ceramic, as seen in table 24.6. See Table 24.7 for a comparison between types and Table 24.7 for class II dielectric coding for capacitance variation for different temperature ranges, e.g. X7R, Z5U.

The mean time between failure λ_o is transformed from the rated voltage and temperature conditions to the operating conditions by equation (24.16), that is

$$\lambda = \lambda_o \times \text{acceleration factors} = \lambda_o \left(\frac{V_{op}}{V_R}\right)^n e^{-\frac{E_a}{K}\left(\frac{1}{T}-\frac{1}{T_O}\right)}$$

where the activation energy E_a and voltage index n are shown in Table 24.9. K is Boltzmann's constant, 8.625×10^{5} eV/K.

Table 24.9: Accelerated MTBF factors

	E_a	n
Dielectric type	activation nergy eV	voltage index
NP0	1.15	2.9
X7R	1.15	2.9
Y5V	1.07	2.4

24.5.3 *Applications*

Flat circular disc ceramic (Z5U dielectric, high K) capacitors have a 2000 V dc, 550 V ac rating with capacitances of up to 47 nF. An exploitable drawback of such a ceramic capacitor is that its permittivity decreases with increased voltage. That is, the capacitance decreases with increased voltage as shown in figure 24.25c. Such a capacitor can be used in the turn-off snubber for the GTO thyristor and diodes which are considered in 8.1.3 and 8.1. High snubbering action is required at the commencement of turn-off, and can subsequently diminish without adversely affecting losses or the switching area trajectory tailoring. The capacitor action is a dual to that performed by a saturable reactor, as considered in 8.3.4. Exploitation of voltage dependence capacitance is generally outside the capacitor specification. Advantageously, the disc ceramic capacitor has low inductance, but the high dissipation factor may limit the frequency of operation. Multi-layer ceramic capacitors can be used in switched mode power supply input and output filters. *Piezoelectric effects* (change of physical size when an electric field is applied) can cause failure due capacitor cracking in traditional X7R class *II* ceramic capacitors.

24.6 Mica dielectric capacitors

The dielectric mica can be one of 28 mica types. It is a naturally occurring inorganic, chemical resistant, clear mineral aluminosilicate (usually India Ruby muscovite, is a hydrated silicate of potassium and aluminum, $H_2KAl_3(SiO_4)_3$) which has a plane of easy cleavage enabling large sheets of single crystal to be split into thin 20-100 µm plates, typically 50 µm.
The general formula for mica is $AB_{2-3}(A\ell, Si)Si_{-3}O_{-10}(F, OH)_{-2}$. In most micas the A is usually potassium, K, but can be calcium, Ca, sodium, Na, barium, Ba, or some other elements in the rarer micas. The B in most micas can be aluminum, Aℓ, and/or lithium, Li, and/or iron, Fe, and/or magnesium, Mg. Phlogopite mica is a hydrated silicate of potassium and magnesium.
Ultra thin silver electrodes are screen printed on to both sides of the mica (and over the edge), as shown in figure 24.26b, which is then fired in an oxidisation atmosphere to obtain a permanent plated adhesive bond between the mica and silver. Variation of silver electrode thickness affects the dissipation factor, while the overlapped printed silver area and mica thickness control the capacitance.
A number of different techniques (and combination of different techniques) are used to parallel connect (parallel stack) the silver coated mica plates in order to give high capacitances.

- Multiple layers of over-the-edge printed mica plates are stacked together (without any interposing foils), as shown in figure 24.26a. The printed silver at each opposite edge is bonded with silver paste, on to which the terminals are directly soldered.
- Multiple silver printed plates are stacked interleaved with metal outer foils for contacts, The foils are made of silver, copper, brass, tin or lead. The foil alternately extends from each end and covers a portion of the plated area. Joining the extended foils at each end, parallel connects each individual mica plate. The stack is held and compressed together either by the encapsulation or by bending the extended foils over the top of the stack which is held by a brass metal (tin coated) crimp, which also acts as a heat sink. Copper clad (for at least 30% conductivity) steel leads are spot-welded to each clip, which is then solder coated.

The assembled unit is encapsulated by dipping it into high melting temperature microcrystalline wax or by dipping in phenolic resin, then vacuum impregnating with liquid epoxy resin.
Mica capacitors are non-magnetic, non-polar, low loss, and stable up to about 30 MHz, where the lead length and electrodes dominate as inductance, typically 5 to 10nH. They are characterised by extremely low capacitor coefficients of temperature and voltage over a wide parameter operating range. Mica has a typical impregnated relative permittivity of 4.5 to 6.5 and a density of 1.65 g/ cm^3.

Alternatively, synthetic fluorine mica, fluorophlogopite $KMg_3(AlSi_3O_{10})F_2$, contains no $(OH)^-$ of the natural phlogopite $KMg_3(AlSi_3O_{10})(OH)_2$. The $(OH)^-$ is fully substituted with F^-. Large fluorine mica crystals of high quality are grown using platinum crucibles with seeds. It has properties similar to natural mica:

Melting temperature	°C		1378
Density	g/cm^3		2.8
Dielectric strength	kV/mm		~180
Volume resistivity	Ω-cm		4×10^{16}
Surface resistivity	Ω-cm		3×10^{12}
Dielectric constant	ε		~6
Dielectric loss	tgδ	1MHz	3×10^{-4}
Tensile strength	kg/cm^2		~1500

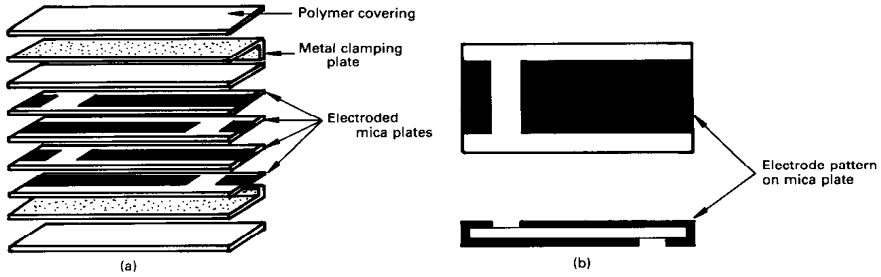


Figure 24.26. *Silver mica capacitor:*
(a) exploded construction view and (b) electrode pattern of a silvered mica plate.

24.6.1 *Properties and applications*

Maximum ratings are a few nanofarads at 50 kV, to 5µF at 1500V, with dissipation factors of 0.1 per cent at 1 kHz. Low dissipation factor is countered by poor dielectric absorption. A 10nF, 50kV mica capacitor in a cylindrical volume of Φ = 150mm×H = 120mm, has <100nH of internal inductance and is capable of 1kA-200ns pulses. For capacitance less than 1 nF, a 0.1 per cent dissipation factor is obtainable at 1 MHz. An insulation resistance of 10^5 MΩ at 20°C down to 10^4 MΩ at 125°C is common for capacitance to 10 nF, after which resistance falls off. Typical operating temperature range is from -55°C to 125°C, with a low capacitance temperature coefficient of 0 to +70 ppm/K. The Indian Ruby mica itself, is thermally stable to 500°C. A capacitance voltage coefficient of ±0.1% over the full voltage range is typical. The maximum current depends on the edge connections and electrodes, so for each physical design the factor is different and is expressed in mA/pF. Because losses are low, typical 0.05% of the throughput VA, energy densities of between 15 and 125 mJ/cm^3 are readily attained.
In power applications, mica capacitors are used to produce stable, high current pulses at high dv/dt (20kV/µs), as for NMR MRI coils, because of its excellent high voltage breakdown properties and corona resistance. Mica dielectric capacitors are sensitive to pressure.
Because of their relatively high cost of manufacture, as a result of the high labour content and diminishing number of mines, the ceramic capacitor, particularly the monolithic multi-layer type, is favoured.

22.7 Capacitor type comparison based on key properties

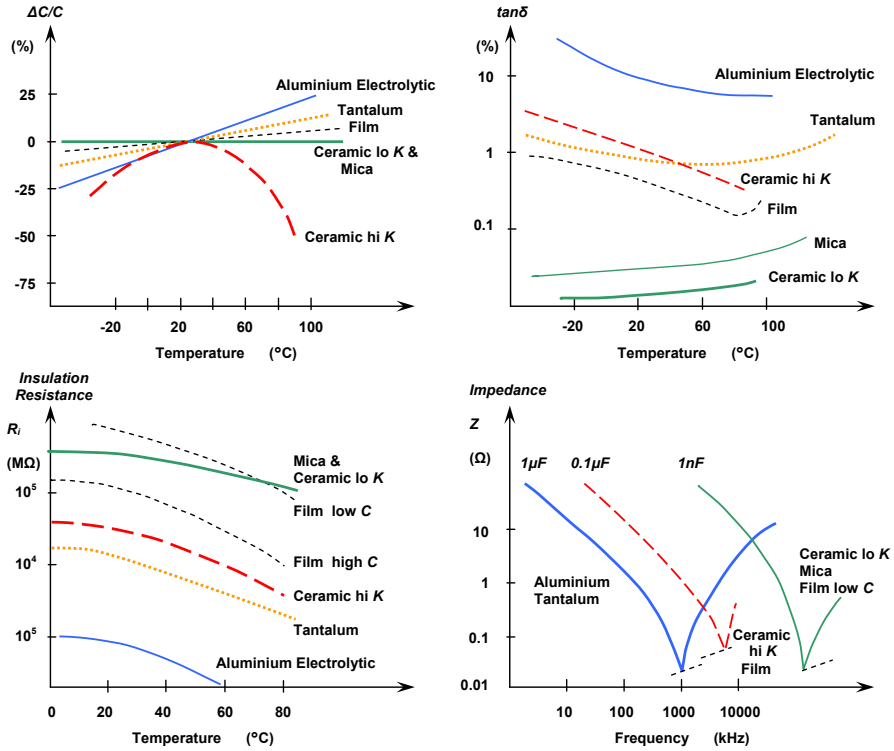


Figure 24.27. *Capacitor type characteristic comparison.*

24.8 Appendix: Minimisation of stray capacitance

Unexpected component stray capacitance, and inductance, can have disastrous power circuit consequences. Figure 24.28 shows four examples of electronic components which have stray capacitance between two parts of the component used at different potentials. When the isolated part rapidly changes its relative potential, a charging current flows according to $i = C\,dv/dt$. With just 1 pF of capacitance, and at 10,000 V/μs, which is possible with MOSFETs and IGBTs, 10 mA of current flows. This current coupled from the power level to the signal level would affect cmos or ttl circuitry, leading to malfunction and possible failure, if precautions are not taken.

Figure 24.28a shows a power package electrically isolated from its heatsink, which is grounded (to 0V or V_s) in order to minimise rfi radiation. Large power blocks have over 100 pF of isolation capacitance. Other than injecting noise, the level may be sufficient to activate earthing leakage circuitry, if connected to ground. Increasing the ceramic substrate or mica thickness decreases capacitance according to equation (24.3), but at the expense of increasing thermal resistance. Aluminium nitride reduces the thermal impedance compared to Al_20_3, but at the expense of increased cost.

Transformer interwinding capacitance, shown in figure 24.28b, is important in switch mode power supplies and other applications using transformers. By winding the primary and secondary in different bobbin sections, the interwinding capacitance is decreased since their physical separation is increased. Alternatively, an overlapped copper foil ground shield layer is wrapped between the two windings. The copper strip is connected to a transformer terminal, a supply rail or earthed so that charging currents bypass sensitive circuitry. Experimentation will reveal the best connection potential and location position. The copper foil overlapped turn ends must not make electrical contact, otherwise a short circuit turn results. Minimise winding start to finish turns capacitance by using the winding method shown in figure 20.23b.

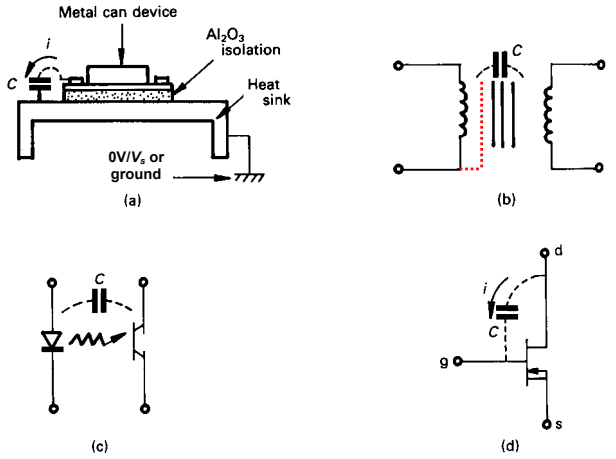


Figure 24.28 *Component stray capacitance C:*
(a) when isolating power devices; (b) between transformer windings; (c) in opto-couplers; and (d) between terminals of metal oxide semiconductor devices.

A similar solution is used in opto-coupler packages. A grounded (with respect to the diode) Faraday's grid is placed between the emitter and receiver in order to divert charging current. High *dv/dt* opto-couplers, with less than 1 pF capacitance input to output, are guaranteed to 15000 V/μs at 200 V dc levels. This *dv/dt* limit decreases to 1000 V/μs on a 600 V dc rail. The effects of capacitive charging current can be minimised by driving the emitting diode from a low impedance source, both when on and off. Speed and current transfer ratio can be traded for higher *dv/dt* capability by increasing isolation separation. For high voltages and high *dv/dt*, a fibre optic is an expensive alternative, but unlike the pulse transformer, has no lower cut-off frequency.

Figure 24.28d shows the Miller capacitance associated with the MOSFET and IGBT. During switching, the Miller capacitance charging and discharging currents slow the switching transition as power level current is injected into and from the gate level circuitry. A low impedance gate drives reduces the Miller capacitance effects.

A commonly overlooked capacitively injected current is that associated with the use of oscilloscope probes, when measuring power level signals. The scope probe ground should be physically connected to an appropriate power ground point, rather than signal ground. Always use the highest possible voltage step-down ratio probes, since capacitance tends to decrease with increased step down ratio.

24.9 Appendix: Capacitor lifetime derating

General formulae for estimating lifetime τ dependence on operating temperature T_{op} and voltage V_{op} are

Film capacitor

$$\tau\left(V_{op},T_{op}\right)=\tau_R\times\left(\frac{V_R}{V_{op}}\right)^7\times 2^{\frac{T_{core}-T_R}{10}} \tag{24.45}$$

where τ_R, V_R, and T_R are rate lifetime, voltage, and temperature. T_{core} is the capacitor core temperature.

Ceramic capacitor

$$\tau\left(V_{op},T_{op}\right)=\tau_R\times\left(\frac{V_R}{V_{op}}\right)^3\times\left(\frac{T_R}{T_{op}}\right)^8 \tag{24.46}$$

Aluminium electrolytic capacitor

$$\tau\left(V_{op},T_{op}\right)=\tau_R\times\left(\frac{V_R}{V_{op}}\right)^1\times 2^{\frac{T_{core}-T_R}{10}} \tag{24.47}$$

The voltage multiplier term may be expressed in a more specific capacitor type form, for example

$$4.3 - 3.3 \times V_{op} / V_R \tag{24.48}$$

Solid tantalum capacitor

$$\tau\left(V_{op}, T_{op}\right) = \tau_R \times \left(\frac{V_R}{V_{op}}\right)^3 \times 2^{\frac{T_{core} - T_R}{10}} \tag{24.49}$$

Reading list

Siemens, *Components,* 1986.

Mullard, *Book 3 Parts 1b and 1e,* 1984/85.

http://my.execpc.com

Rohm.com

Kemet.com

AVXcorp.com

Vishay.com

Epcos.com

Evox-Rifa.com

Cornell-Dubilier.com

blank

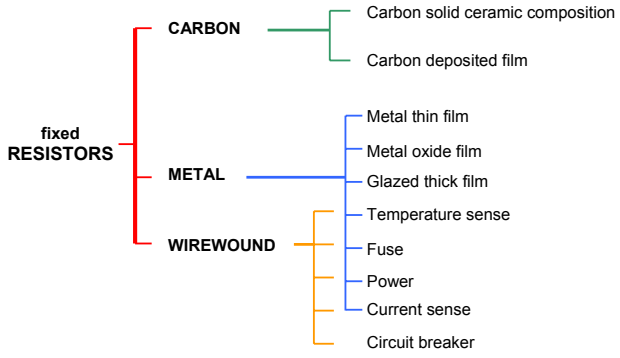
25

Resistors

Power resistors (≥1W) are used extensively in power electronic circuits, either as a pure dissipative element, or to provide a current limiting path for charging/discharging currents. Depending on the application, these energy transfer paths may need to be either inductive or non-inductive. Power resistors are used for the following non-inductive resistance applications.

- Series *R-C* circuit for diode, mosfet and thyristor snubbers (non-inductive) (6.1.2,8.1, 8.3)
- Series turn-on *L-R-D* snubbers (6.1.2, 8.3.3)
- *R-C-D* turn-off snubbers for GTO thyristors - inductance in *R* is allowable (8.2)
- Static voltage sharing for series connected capacitors and semiconductors (10.1.1)
- Static current sharing for parallel connected semiconductors (10.1.2)
- Resistor divider for proportional voltage sensing (10.2.3)
- Current sensing (10.2.3)
- Damping, clamping, and voltage dropping circuits

The resistive element specification can be more than just fulfilling resistance and power dissipation requirements. For example, a current shunt resistor should be non-inductive in order to give good high frequency performance. Conversely, the resistor of the *R-C-D* turn-off snubber considered in 8.3 can be inductive thereby reducing the high initial peak current associated with an *R-C* discharge. An important resistor requirement is working voltage and the dielectric withstand voltage. High voltages (>200 V) are common in power circuits and the physical construction of a resistor places a limit to allowable voltage stress levels. Certain applications within the realm of power electronics may necessitate a power resistor with a low temperature coefficient of resistance (or even a negative temperature coefficient), a high operating temperature, a high pulse power ability or even a low thermoelectric voltage. Any one of these constraints would restrict the type and construction of resistor applicable.



25.1 Resistor types

The resistor tree illustrates the main types of resistors used in electrical power applications. The three main resistor types are carbon/metal film, solid, and wire wound. The main electrical and thermal properties of each resistor type are summarised in table 25.1. Typical property values for power resistors are shown, which may vary significantly with physical size and resistance value.

25.2 Resistor construction

Almost all types of power resistors (≥1W) have a cylindrical high purity ceramic core, either rod or tube as shown in figure 25.1. The core has a high thermal conductivity, is impervious to moisture penetration, is chemically inert, and is capable of withstanding thermal shock. The resistive element is either a carbon film, a homogeneous metal-based film or a wound wire element around the ceramic body. For high accuracy and reliability, a computer-controlled helical groove is cut into the film types in order to trim the required ohmic resistance. The resistance tolerance can be typically ± 5% for wire wound resistors and better than ± 0.1% for trimmed film types.
The terminations are usually nickel-plated steel, or occasionally brass, force fitted to each end of the cylindrical former in order to provide excellent electrical and thermal contact between the resistive layer and the end-cap. Tinned connecting wires of electrolytic copper or copper-clad iron are welded to the end-caps, thereby completing the terminations. Axial cylindrical resistors without leads, used as surface mount resistive devices (SMD), are termed *metal electrode leadless face*, MELF.
All fixed resistance resistor bodies are coated with a protective moisture-resistant, high dielectric field strength, and some times conformal coating, such that the wire terminations remain clear and clean.
The resistors are either colour coded by colour bands or provided with an identification stamp of alphanumeric data.

Table 25.1: The main characteristics of electrical power resistors

			carbon composition	carbon deposited film	metal thin film	glazed thick film		metal oxide film	power wire wound	fusible	circuit breaker	temperature sense	current sense
						LV	*HV*						
Resistance range	R	Ω	10-22M	1-10M	1-5M	1-2M	300k-1G	15-100k	0.1-1.5M	0.1-3.9k	0.27-82k	0.1-300	0.01 - 10
Watts @ 70°C	P_R	W	1	2	2.5	2	90	7	>300	2	6	2	9
Maximum temperature	T_h	°C	150	125	300	175	100	235	275	160	150	200	250
Working voltage	V_m	V	500	500	500	1k	100k	650	2.5k	160	500	700	$\sqrt{PR}$
Voltage coefficient	φ	10^{-6}/V	200	50	5	10	-	0.1	<1	-	-	-	-
Residual capacitance	C_R	pF	¼	½	-	-	$\frac{1}{5}$	½	-	-	-	-	-
Temperature coefficient	α	10^{-6}/K	-500 -1000	+50 -350	±350	±200	±150	±500	50	500	-80 +500	-3000 +5500	100
Thermal resistance	R_θ	K/W	80	27	90	35	13	26	0.3	50	14	0.55	20
Reliability	λ	10^{-9}/hr fit	1	10	1		-	3	300	-	-	-	-
Stability $\frac{\Delta R}{R}$%	@ P_R, T_a = 70°C, @10^3 hours		5	3	5	½	2	3	3	5	2 @ 150°C	0.1	3
Tolerance %			10	5		1	1	1		0.1			

25.2.1 Film resistor construction

Figure 25.1a shows a sectional view of a typical film resistor having a construction as previously described. The resistive film element is produced in one of four ways:

- cracked carbon film
- glaze of glass powder mixed with metals and metal compounds fired at 1000°C, giving a firmly bonded glass-like film on the core

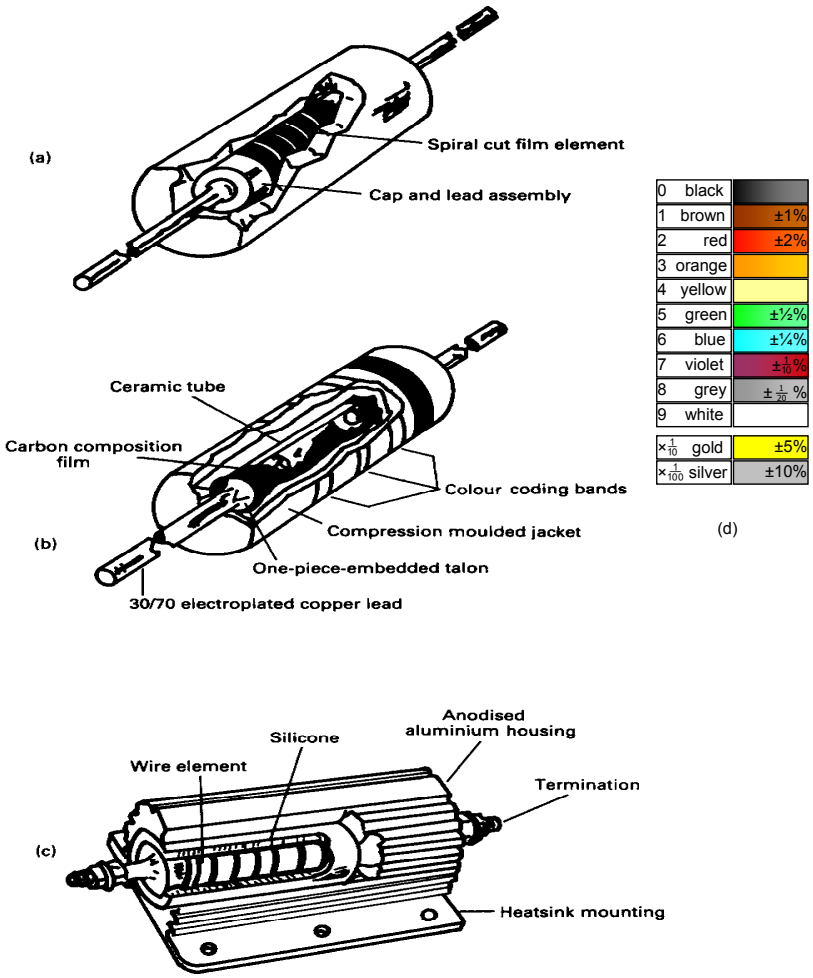


Figure 25.1. *Power resistor construction: (a) metal glaze, thick film; (b) moulded carbon composition film; (c) wire wound aluminium clad; and (d) resistance colour code and tolerances.*

- precisely controlled thin film of metal alloy (Cr/Ni or Au/Pt) evaporated, baked or vacuum sputtered (vacuum deposition) on to the inert core and of thickness between 10^{-8}m and 10^{-7}m
- metal oxide ($Sn0_2$) resistive film deposited or sintered on to the core.

The film materials exhibit a wide range of resistivity, ρ, which extends from 40 × 10^{-6} Ω cm for gold/platinum to in excess of 10^{2} Ω cm for layers of thick film mixture. The thinnest possible film, for maximum resistance, is limited by the need for a cohesive conductive film on the ceramic substrate while the thickest film, for minimum resistance, is associated with the problem of adhesion of the resistance film to the substrate.
The helical groove shown in figure 25.1a, used to trim resistance (by increasing the length of the film width), is shown clearly and is either laser or diamond (abrasive) cut. The residual inductance is significantly increased because of the formed winding which is a spiral around the core. Below 100 Ω, a helical groove may not be used.
The difference between thick and thin film is how the film is applied (not necessarily the film thickness).

Thick film, as used for low cost surface mount (SMD) resistors, employs traditional liquid screen printing technology baked at 850°C, as opposed to spluttering type technology for thin film resistors.
As a result thin film resistors are

- more expensive than thick film resistors
- have better tolerances, typical 0.1% to 1% and a temperature coefficient of ±5 to ±25 ppm, because the spluttering time can be used to control the film thickness

Thick film tolerances are typically: 1 to 2% on resistance and a temperature coefficient of ±200ppm.

Resistive materials such as polymer and ruthenium oxide RuO_2 are used for high voltage and high resistance resistors up to 100 kV, 1 GΩ and 20 kV, 150 GΩ respectively. Ruthenium oxide is called a *cermet* since it is a composite of a *cer*amic and a *met*al. At lower voltage ratings, the same oxide is used to produce planar thick film power resistors, which are mounted on an alumina substrate to give a low thermal resistance for better cooling. The planar strip structure gives a low inductance element with a high surface area for better heat transfer. Dissipation levels of over 200W, up to 900W, are possible when packaged in TO220, ISOTOP, etc. type packages, which can be heatsink mounted.

25.2.2 Carbon composition film resistor construction

A sectional view of a moulded carbon composition film resistor is shown in figure 25.1b. The resistive carbon film is cured at 500°C and is unspiralled, hence non-inductive with excellent high frequency characteristics. The resistance value is obtained by variation of film composition and thickness, which may be between 0.01 µm and 30 µm. A component with a wide nominal value tolerance results since the film is not helically trimmed.
A special formed one-piece talon lead assembly is deeply imbedded into the substrate for good uniform heat dissipation. These terminations are capless.
The following example illustrates typical parameters and dimensions for carbon film resistors.

Example 25.1: ***Carbon film resistor***

A 470 Ω resistor is constructed from a film of carbon with a resistivity 3.5 × 10^{-5} Ω m, deposited on a non-conducting ceramic bar 3 mm in diameter and 6 mm long.
Calculate the thickness of film required, ignoring end connection effects.

Solution

Let thickness of the film be t metre, then

cross-sectional area ≈ $\pi \times 3 \times 10^{-3} \times t$ (m^2)

Now $R = \rho\ell$/area, that is

$470\Omega = 3.5\times10^{-5}\times6\times10^{-3}/\pi\times3\times10^{-3}\times t$

t = 0.0474 µm

♣

25.2.3 Solid carbon ceramic resistor construction

Mixtures of finely ground, powdered clay, alumina, and carbon are blended with resin, pressurised into the desired shape (diameters from less than 0.3mm to over 15mm), and fired in a tunnel kiln, at high temperature and controlled pressure. The higher the carbon concentration, the higher the resistivity (resistance). This sintering process results in a 100% active homogeneous, solid volume resistive element, in a minimum size. Aluminium (and/or copper brass, silver, nickel) is then flame-sprayed on to the appropriate surfaces to provide electrical contact, followed by gold plated spring pin terminations, if to be used for pcb mounting. Then an anti-tracking epoxy resin coating is applied using a fluidised bed technique, to improve dielectric withstand, mechanical robustness, and minimise corrosion. Although the coating reduces the rate of moisture ingress, the element is not impervious to liquid, so after drying for 24 hours at 110-120°C, it is silca gel coated if the resistor is to be immersed in SF_6 gas or oil.
Because of the solid construction, the ceramic carbon element has a high surge energy rating, high voltage withstand, high transient voltage impulse withstand, inherent low inductance, higher thermal capacity, and is mechanical robust; with compact size and a wide range of geometries. It is brittle to direct mechanical impact. Also, because of its homogeneous physical and chemical structure, resistor mechanical, electrical, and thermal properties and characteristics can be defined mathematically or empirically. At high electrical stress levels, the resistivity property change. Mechanical, thermal, and electrical data is presented in Appendix 25.8 for solid ceramic carbon. Rod, disc, and tube shapes are common, with resistance tolerances of not better than ±5%.

25.2.4 Wire-wound resistor construction

The sectional view of an aluminium-housed power wire-wound resistor, shown in figure 25.1c can dissipate up to 300 W with a suitable heatsink in air or up to 900 W when water-cooled.
The central former is a high purity, high thermal conductivity ceramic, of either Steatite or Alumina tube, depending on size. The matching resistive element is iron-free, 80/20 nickel-chromium for high resistance values or copper-nickel alloy for low resistance. These alloys result in a wire or tape which has a high tensile strength and low temperature coefficient. The tape or wire is evenly wound on to the tube former with a uniform tension throughout. This construction is inductive but gives a resistor which can withstand repeated heat cycling without damage.
The assembled and wound rod is encapsulated in a high temperature thermal conducting silicone moulding material and then cladded in an extruded, hard, anodised aluminium housing, ensuring electrical and thermal stability and reliability.
Alternatives to the aluminium-clad resistor are to encapsulate the wound rod in a vitreous enamel or a fire-proof ceramic housing.
Power wire-wound resistors with a low temperature coefficient, of less than $\pm 20\times10^{-6}$ /K, use a resistive element made of Constantan (Nickel and Copper) or Nichrome (Nickel and Chromium). Constantan is used for lower resistance, up to several kilo-ohm, while Nichrome is applicable up to several hundred kilo-ohm. The resistance ranges depend on the ceramic core dimensions, hence power rating. The element is wound under negligible mechanical tension, resulting in a reliable, low temperature coefficient resistor which at rated power can safely attain surface temperatures of over 350°C in a 70°C ambient. Because these resistors can be used at high temperatures, the thermally generated emf developed at the interface between the resistive element and the copper termination can be significant, particularly in the case of Constantan which produces -40 µV/K. Nichrome has a coefficient of only + 1 µV/K, while gold, silver, and aluminium give +0.2, -0.2, and -4 µV/K, respectively, when interfacing with copper. EMF polarities cancel with identical resistor terminations.

Ayrton-Perry wound wire elements can be used for low inductance applications. The resistive element is effectively wound back on itself, such that the current direction in parallel conductors oppose. Either a bifilar winding or an opposing chamber winding is used. The net effect is that a minimal magnetic field is created, hence residual inductance is low. The maximum resistance is one-quarter that for a standard winding, while the limiting element voltage is reduced, by dividing by √2. The low inductance winding method is ineffective below 4.7Ω. This winding style also lowers the maximum permissible winding temperature, called *hot spot temperature*, T_h. The hot spot temperature is the resistor surface temperature at the centre of its length.

25.3 Electrical properties

An electrical equivalent circuit for a wirewound resistor is shown in figure 25.2. The ideal resistor is denoted by the rated resistance, R_R, and the lumped residual capacitance and residual inductance are denoted by C_r and L_r respectively. A film type resistor is better modelled with the capacitance in parallel with the resistive component.
The terminal resistance of a homogeneous element of length ℓ and area A is given by

$$R(v,f,T) = \rho(v,f,T)\frac{\ell}{A} = \frac{\ell}{\sigma A} \qquad (\Omega) \tag{25.1}$$

where ρ is the resistivity of the resistive element and σ is the conductivity (= $1/\rho$).
The total effective resistance of series and parallel connected resistive elements are given by

$$R_s = R_1 + R_2 + R_3 + \ldots$$
$$\frac{1}{R_p} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \ldots \tag{25.2}$$

$$Z = \frac{R_R}{\frac{R_R^2}{X_{Cr}^2} + \frac{(X_{Lr} - X_{Cr})^2}{X_{Cr}^2}} + j\,\frac{X_{Lr} - \left[\left(R_R^2 + X_{Lr}^2\right)/X_{Cr}\right]}{\frac{R_R^2}{X_{Cr}^2} + \frac{(X_{Lr} - X_{Cr})^2}{X_{Cr}^2}}$$

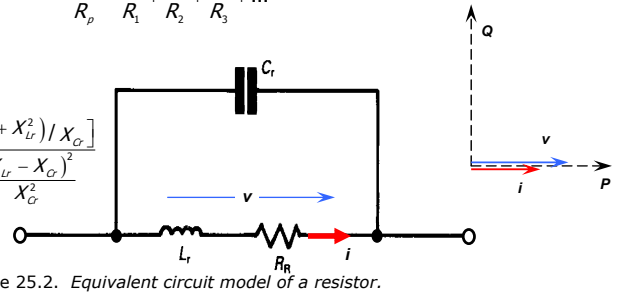


Figure 25.2. *Equivalent circuit model of a resistor.*

The terminal resistance is a function of temperature, voltage, and frequency. Temperature dependence is due to the temperature dependence of resistivity α, typical values of which are shown in table 25.1. The temperature coefficient may vary with either or both temperature, as with carbon and metal film resistors, or resistance, as with thick film and noble metal film resistors. A reference for measurement is usually 25°C. Frequency dependence is due to a number of factors, depending on the type of resistive element and its resistance value. Typical factors are due to skin effects in the case of wire-wound resistors or residual capacitance in film types. Frequency dependent resistance, R_{ac}, for carbon composition film and metal glaze thick film resistors, is shown in figure 25.3.
A voltage dependence factor, which is called ϕ, is given in table 25.1 and is resistive element type dependent. For operation at low frequencies, resistance is given by

$$R(v,T) = R(0,25°\text{C})(1+\phi v)(1+\alpha T) \qquad (\Omega) \tag{25.3}$$

Values for linearity coefficients ϕ and α are given in table 25.1.

Ideally, electrically, the terminal voltage and current are in phase and related by Ohm's law, namely

$$v = i \times R \qquad (\text{V}) \tag{25.4}$$

where it is usually assumed that R is constant. This electrical relationship is shown in the phase diagram in figure 25.2. In practice when a pure sinusoidal current is passed through a resistor its terminal voltage may not be a pure sinusoid, and may contain harmonic components. This voltage distortion is termed nonlinearity and is the harmonic deviation in the behaviour of a fixed resistor from Ohm's law, equation (25.4). Another resistor imperfection is *current noise* which is produced by the thermal agitation of electrons due to resistive element conductivity fluctuation. The noise voltage is proportional to current flow. Johnson noise is given by $E_{rms}^{noise}(t) = \sqrt{4k\,RT\,\Delta f}$, where Δf is the measurement bandwidth and k is Boltzmann's constant. Wire-wound resistors generate negligible current noise. The resistance value itself can change: *long term drift* due to chemical-physical processes such as oxidation, re-crystallisation corrosion, electrolysis, and diffusion, as may be appropriate to the particular resistive element.

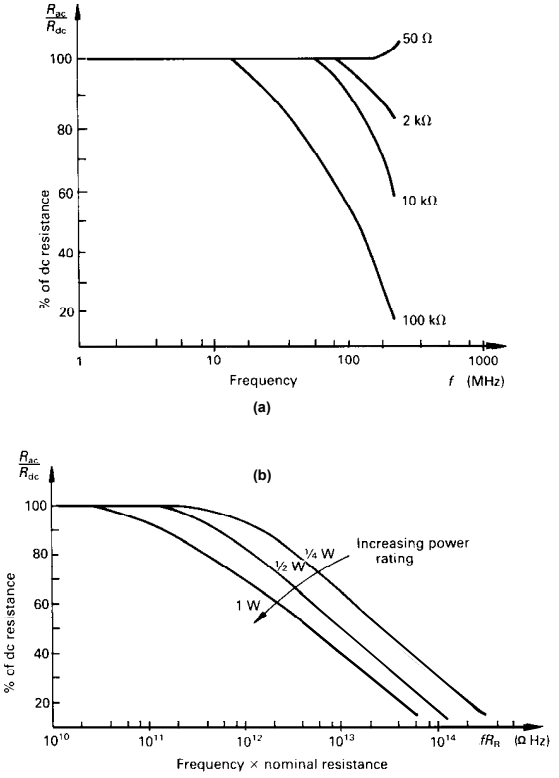


Figure 25.3. *Resistor high frequency characteristics:*
(a) of a metal glaze thick film 3 W resistor and (b) moulded carbon composition film resistors.

The power dissipated, P_d, in an ideal resistor, in general, is specified by

$$P_d = vi = i_{rms}^2 R = v_{rms}^2 / R \qquad \text{(W)} \qquad (25.5)$$

where the power dissipated is limited by the power rating, P_R, of the resistor. In turn, the power limit may also set the maximum voltage that can be withstood safely. Joules law of heat generated is $Q=I^2Rt$.

25.3.1 Resistor/Resistance coefficients

Three mechanical, thermal or electrical coefficients are relevant to a resistor and its resistance

- coefficient of linear expansion
- temperature coefficient of resistance, α
- voltage coefficient of resistance, Φ

The coefficient of linear expansion is only relevant to solid carbon ceramic type resistors. It is dependant on resistivity, ρ, and is in the range $+4\times10^{-6}$ to $+10\times10^{-6}$ /°C. The voltage and temperature coefficients of resistance are also dependant on the resistivity of the resistive element.

25.3.1i - Temperature coefficient of resistance - α

The resistance temperature coefficient α in equation (25.3) and in table 25.1 should not be taken as to imply a linear relationship between temperature and resistance, $1+\Phi\times V$. Quite the contrary, for simplicity, it is only a linear approximation to a quadratic approximation of a non-linear function, as shown in figure 25.4. This figure shows the resistance behaviour of a typical thick film resistor, with varied temperature. The lower the temperature coefficient, the flatter the resistance-temperature curve. The temperature dependant relationship is modelled by a quadratic equation:

$$R(T) = R_o + aT^2 + bT + c \qquad (25.6)$$

where a, b, and c are the quadratic coefficients and R_o is the minimum resistance value, which for this resistor type, usually occurs around 35°C. The temperature at which the minimum resistance R_o occurs can be varied by changing the resistor chemical composition. The temperature co-efficient α is related to the slope $\Delta R/\Delta T$ and is defined by

$$\alpha(T) = \frac{R(T) - R(25°C)}{\Delta T \times R(25°C)} \times 10^6 \quad = \quad \frac{\Delta R}{\Delta T} \times \frac{10^6}{R(25°C)} \qquad (10^{-6}/°C \text{ or ppm/°C}) \qquad (25.7)$$

Given the normal temperature operating range for resistors, it is usual when characterising resistors, to use 25°C as a reference with a +50°C temperature change, that is ΔT = +50°C. The resultant resistance change for this ΔT = +50°C defines the temperature co-efficient. It will be seen that the definition coefficient slope is not necessarily the maximum slope in the range 25°C to 75°C.

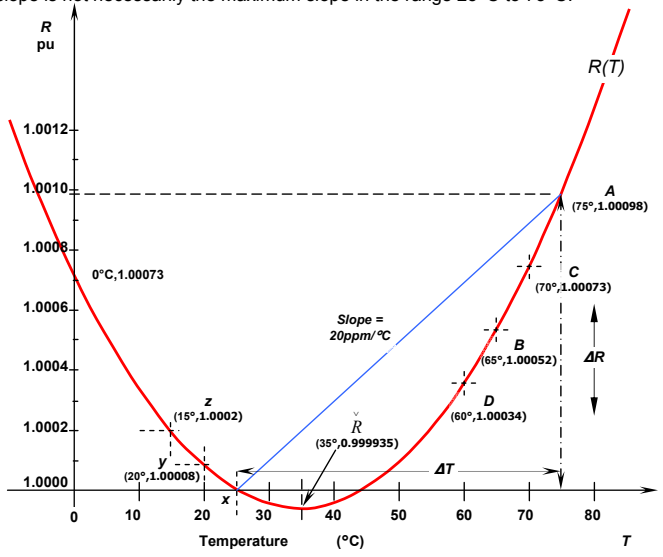


Figure 25.4. *Typical temperature characteristics of resistance for thick film resistors.*

Alternatively, solid carbon ceramic resistive elements may have a temperature coefficient defined in terms of resistivity (in Ω-cm), as an empirical formula such as:

$$\alpha = +1600 \times e^{-\log\rho/1.4} - 1350 \qquad (10^{-6}/°C \text{ or ppm/°C}) \qquad (25.8)$$

Example 25.2: ***Temperature coefficient of resistance for a thick film resistor***

Figure 25.4 shows a quadratic approximation for the variation of per unit resistance with temperature, with a minimum resistance of 0.999935pu occurring at 35°C, and 1pu resistance at 25°C and 45°C.

Determine

i. suitable model quadratic coefficients a, b, c
ii. the 0°C pu resistance
iii. the temperature co-efficient of resistance based on ΔT = +50°C and ΔT = +40°C and the resistance change in each case, with a 25°C reference, point *x*
iv. the incremental temperature coefficient over the range 25°C to 75°C
v. the temperature co-efficient of resistance based on ΔT = +50°C and ΔT = +40°C and the resistance change in each case, with a 20°C reference, point *y*
vi. the minimum resistance occurs at 45°C, with the identical shape. Determine the resistance temperature coefficient (ΔT = 75°C - 25°C) of the resultant resistor

Solution

i. Assuming the temperature dependant resistance data in figure 25.4 is curve fitted by a best fit quadratic, then that quadratic has roots at T = 25°C and T = 45°C, thus after shifting the Y-axis (R_{pu}) by 1pu:

$$R_{pu} - 1 = a(T - 25°C)(T - 45°C)$$

To ensure accuracy in the normal operating range, the minimum resistance point $\hat{R}$, (35°C, 0.999935) is used to find the second order temperature term co-efficient, and

$$R_{pu} - 1 = a \times (T - 25°C)(T - 45°C)$$

$$0.999935 - 1 = a \times (35°C - 25°C)(35°C - 45°C) \quad \Rightarrow \quad a = 6.5\times10^{-7}$$

The quadratic coefficients are thus

$$R_{pu} - 1 = 6.5\times10^{-7} \times (T - 25°C)(T - 45°C)$$

$$R_{pu} - 1 = 6.5\times10^{-7} T^2 - 4.55\times10^{-5} T + 7.3125\times10^{-4} \quad \text{or}$$

$$R_{pu} = 6.5\times10^{-7} T^2 - 4.55\times10^{-5} T + 1.00073125$$

ii. Substituting 0°C into the quadratic gives the constant c, that is 1.000731 - 0.999935 = 0.00080, as can be confirmed from the best fit plot in figure 25.4.

iii. When using 25°C as the parameter specification reference, point x, the quadratic is used to give the pu resistance at points A (75°C) for ΔT = +50°C and B (65°C) for ΔT = +40°C shown in figure 25.4.

For point A:

$$\alpha_{\substack{T_a=25°C \\ \Delta T=50°C}} = \frac{R_{pu-A} - R_{pu-a}}{T_A - T_a}$$

$$= \frac{1.00098\text{pu} - 1\text{pu}}{75°C - 25°C} = 19.6\times10^{-6}\,(/°C) \text{ or } 19.6\text{ppm/°C}$$

For point B:

$$\alpha_{\substack{T_a=25°C \\ \Delta T=40°C}} = \frac{R_{pu-B} - R_{pu-a}}{T_B - T_a}$$

$$= \frac{1.00052\text{pu} - 1\text{pu}}{65°C - 25°C} = 13.0\times10^{-6}\,(/°C) \text{ or } 13\text{ ppm/°C}$$

iv. The tangential slope to the quadratic is the differential of the quadratic:

$$\alpha = 13.0\times10^{-7} \times T - 4.55\times10^{-5}$$

$$\alpha_{25°C} = 13.0\times10^{-7} \times 25°C - 4.55\times10^{-5} = -13\text{ ppm/°C}$$

$$\alpha_{35°C} = 13.0\times10^{-7} \times 35°C - 4.55\times10^{-5} = \ 0\text{ ppm/°C}$$

$$\alpha_{75°C} = 13.0\times10^{-7} \times 75°C - 4.55\times10^{-5} = +52\text{ppm/°C}$$

The temperature coefficient ranges from -13 to 52 ppm/°C.

v. When using 20°C as a parameter specification reference, point *y*, the quadratic is used to give the pu resistance at points C, ΔT = +50°C, and D, ΔT = +40°C, shown in figure 25.4.

For point C:

$$\alpha_{\Delta T=50^\circ C}^{T_a=20^\circ C} = \frac{R_{pu-C} - R_{pu-b}}{T_A - T_a}$$

$$= \frac{1.00073\text{pu} - 100008\text{pu}}{70^\circ\text{C} - 20^\circ\text{C}} = 13.0\times10^{-6}\,(/^\circ\text{C}) \text{ or } 13.0\text{ ppm/°C}$$

For point D:

$$\alpha_{\Delta T=40^\circ C}^{T_a=20^\circ C} = \frac{R_{pu-D} - R_{pu-a}}{T_D - T_a}$$

$$= \frac{1.00034\text{pu} - 100008\text{pu}}{65^\circ\text{C} - 25^\circ\text{C}} = 6.5\times10^{-6}\,(/^\circ\text{C}) \text{ or } 6.5\text{ ppm/°C}$$

vi. The data for a shift of the minimum resistance from 35°C to 45°C can be obtained by considering the appropriate quadratic graph translation. The point z becomes the new 25°C reference and the 50°C temperature increase point becomes data point B.

$$\alpha_{\Delta T=50^\circ C}^{T_c=15^\circ C} = \frac{R_{pu-B} - R_{pu-c}}{T_B - T_c}$$

$$= \frac{1.00052\text{pu} - 1.0002\text{pu}}{65^\circ\text{C} - 15^\circ\text{C}} = 6.4\times10^{-6}\,(/^\circ\text{C}) \text{ or } 6.4\text{ ppm/°C}$$

♣

25.3.1ii - Voltage coefficient of resistance - Φ

The voltage coefficient of resistance is resistivity, ρ, dependant and usually assumed linear for simplicity. For resistors intended for high voltage application, such as carbon ceramics, more accurate voltage dependence is necessary and element length dependant, empirical formula are provided, for example

$$\phi = -0.62\times\rho^{0.22} \qquad (\%/\text{kV/cm}) \qquad (25.9)$$

where resistivity, ρ, is in Ω-cm.

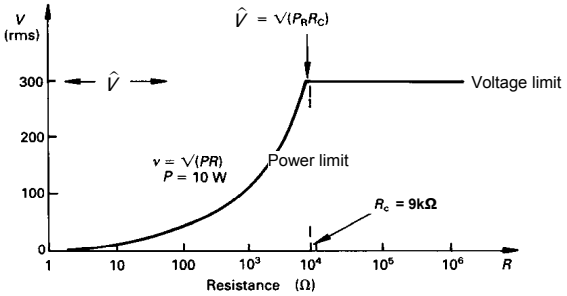


Figure 25.5. *Resistor voltage limits for a given power rating.*

25.3.2 Maximum working voltage

The maximum working voltage $\hat{V}$, either dc or ac rms, is the limiting element voltage that may be continuously applied to a resistor without flashover, subject to the maximum power rating P_R not being exceeded. A typical characteristic is shown in figure 25.5, which illustrates the allowable voltage bounds for a 10 W resistor range, having a limiting flashover voltage of 300 V rms.

At lower resistance, power dissipation capability limits the allowable element voltage, and above a certain resistance level, termed *critical resistance*, R_c, the maximum working voltage, $\hat{V}$, is the constraint. Maximum working voltage decreases with decreased air pressure, typically a 30 per cent reduction for low pressures.

Resistive elements intended for high voltage and surge applications have the maximum working voltage more rigorously defined, and for different possible operating conditions. The following empirical expressions are valid for solid carbon ceramic resistors. It will be seen that the maximum working voltage is fundamentally a property of resistivity, not resistance.

For voltage pulses in the range of t =10ms to 50ms, the maximum working voltage per cm of resistor length is

$$\hat{V} = 1.00\times\left(\frac{\rho}{t}\right)^{0.335} = 1.00\times\left(\frac{R}{t}\frac{A}{\ell}\right)^{0.335} \quad (\text{kV/cm}) \text{ in } SF_6 \text{ gas}$$

$$\hat{V} = 0.87\times\left(\frac{\rho}{t}\right)^{0.3} = 0.87\times\left(\frac{R}{t}\frac{A}{\ell}\right)^{0.3} \quad (\text{kV/cm}) \text{ in air} \qquad (25.10)$$

Low resistance values (<100Ω):

For transient impulse voltages across low resistances, the maximum working voltage per cm of length is

$$\hat{V} = 8.0\times\sqrt[1.2]{\log\left(\frac{\rho}{2.54}\right)} = 8.0\times\sqrt[1.2]{\log\left(\frac{R}{2.54}\frac{A}{\ell}\right)} \quad (\text{kV/cm}) \text{ 1.2μs/50μs waveform, in } SF_6 \text{ gas}$$

$$\hat{V} = 4.3\times\sqrt[1.2]{\log\left(\frac{\rho}{2.54}\right)} = 4.3\times\sqrt[1.2]{\log\left(\frac{R}{2.54}\frac{A}{\ell}\right)} \quad (\text{kV/cm}) \text{ 1.2μs/50μs waveform, in air} \qquad (25.11)$$

$$\hat{V} = 3.0\times\log\left(\frac{\rho}{2.54}\right) = 3.0\times\left(\frac{R}{2.54}\frac{A}{\ell}\right) \quad (\text{kV/cm}) \text{ 50μs/1000μs waveform, in air}$$

$$\hat{V} = 1.5\times\left(\log\left(\frac{\rho}{2.54}\right)\right)^{\frac{5}{4}} = 1.5\times\left(\log\left(\frac{R}{2.54}\frac{A}{\ell}\right)\right)^{\frac{5}{4}} \quad (\text{kV/cm}) \text{ 100μs/10000μs waveform, in air} \qquad (25.12)$$

High resistance values (>100Ω):

For transient impulse voltages across high resistances, the maximum working voltage per cm length is

$$\hat{V} = c\frac{\rho}{A}\times\left(-1+\sqrt{1+\frac{a}{\frac{\rho\ell}{A}}}\right) = bR\times\left(-1+\sqrt{1+a/R}\right) \quad (\text{kV}) \qquad (25.13)$$

where *a* and *b* are specified for a given resistance (area and length).

a increases with length and decreases with area

b is independent of length and increases with area

The constant *a* increases and *b* decreases, as the impulse period (time) increases.

25.3.3 Residual capacitance and residual inductance

Generally, all resistors have residual inductance and distributed shunt capacitance. Inductance increases with both resistance and power rating as shown by figure 25.6, while residual capacitance increases mainly with increased pulse rating. For example, a 2 W metal oxide film resistor (typical of film resistors) has ½ pF residual capacitance and inductance varying from 16 nH to 200 nH. A ¼ W family member has 0.13 pF of capacitance and 3 to 9 nH of inductance.

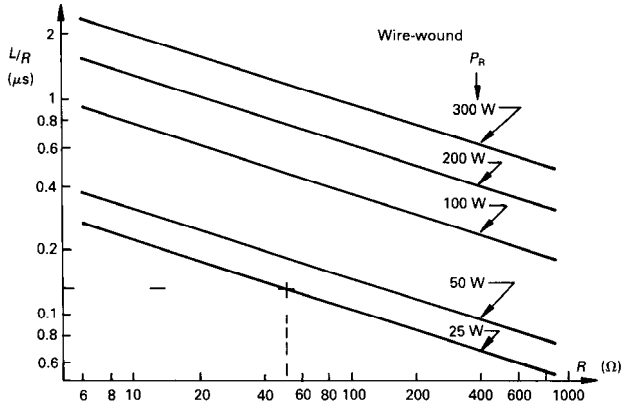


Figure 25.6. *Resistor time constant versus resistance and power rating.*

Film resistors, with resistance above 1 kΩ, having a helical grove, tend to be dominated by interspiral capacitance effects at frequencies above 10 MHz, as seen in figure 25.3a.
Non-inductive elements have low shunt capacitance, such as in the case of carbon composition, while wire-wound resistors can have microhenries of inductance. For example from figure 25.6, a 25 W, 47 Ω, wire-wound resistor may have 6 μH of inductance. Residual inductance increases with resistance and decreases with frequency.
Solid resistive elements, like carbon ceramic, have minimal self-inductance. The method of circuit connection tends to dominate inductance values. Their dielectric constant, ε_r, is about 5, and depends on resistivity.

Example 25.3: ***Coefficients of resistance for a solid carbon ceramic resistor***

Solid carbon ceramic rods of length 1 cm and area 0.25 cm^2, are used as voltage sharing resistors across each series connected switching device. If the ceramic resistivity is 7500Ω-cm, using equations (25.8) and (25.9), determine

i. the resistance at 25°C, with zero supporting voltage
ii. resistance for a 50°C temperature increase
iii. resistance for a 1000V 20ms voltage pulse
iv. resistance at 50°C and 1000V for 20ms, simultaneously
v. maximum working voltage at (a) 50Hz in air and in SF_6
(b) 10μs/1000 μs impulse in air, $a = 2128$ and $b = 0.031$

Solution

i. From equation (25.1), resistance is

$$R = \rho\frac{\ell}{A} = 7500\Omega\text{-cm}\times\frac{1\text{cm}}{0.25\text{cm}^2} = 30\text{k}\Omega$$

ii. Resistance for a 50°C temperature increase is given from equation (25.8)

$$R_{75°C} = R_{25°C}\left(1+\alpha\times\Delta T\right)$$
$$= 30\text{k}\Omega\left(1+\left(1600\times e^{-\log\rho/1.4}-1350\right)\times10^{-6}\times50°\text{C}\right) = 0.9375\times30\text{k}\Omega = 28.12\text{k}\Omega$$

iii. Resistance for a 20ms 1000V pulse is given by equation (25.9)

$$R_{1000V,20ms} = R_{0V}\left(1+\phi\times\ell\times V\right)$$
$$= 30\text{k}\Omega\left(1-\left(0.62\times\rho^{0.22}\right)\times10^{-2}\times1\text{cm}\times1000\text{V}\right) = 0.9558\times30\text{k}\Omega = 28.67\text{k}\Omega$$

iv. For both a 50°C temperature increase and a 1000V, 20ms pulse, assume independence and superposition hold, that is, from parts i, ii, and iii

$$R_{75°C,1000V} = R_{25°C}\left(1+\alpha\times\Delta T\right)\left(1+\phi\times\ell\times V\right)$$
$$= 30\text{k}\Omega\times\left(1+\left(1600\times e^{-\log\rho/1.4}-1350\right)\times10^{-6}\times50°\text{C}\right)\times\left(1-\left(0.62\times\rho^{0.22}\right)\times10^{-2}\times1\text{cm}\times1000\text{V}\right)$$
$$= 30\text{k}\Omega\times0.9375\times0.9558 = 26.88\text{k}\Omega$$

v. The maximum working voltage at 50Hz is given by equation (25.10), that is

$$\hat{V} = 1.00\times\left(\frac{\rho}{t}\right)^{0.335}\times\ell = 1.00\times\left(\frac{7500\Omega\text{-cm}}{10\text{ms}}\right)^{0.335}\times1\text{cm} = 9.18\text{ kV in }SF_6\text{ gas}$$

$$\hat{V} = 0.87\times\left(\frac{\rho}{t}\right)^{0.3}\times\ell = 0.87\times\left(\frac{7500\Omega\text{-cm}}{10\text{ms}}\right)^{0.3}\times1\text{cm} = 6.33\text{ kV in air}$$

The maximum working voltage, in air, for a 10μs/1000μs impulse is given by equation (25.13)

$$\hat{V} = bR\times\left(-1+\sqrt{1+a/R}\right)$$
$$= 0.031\times30\text{k}\Omega\times\left(-1+\sqrt{1+2128/30\text{k}\Omega}\right) = 13.7\text{ kV}$$

♣

25.4 Thermal properties

The continuous power rating of a resistor, P_R, is based on three factors:

- Maximum surface temperature, in free air, over the usable ambient temperature range, typically from -55°C to well over 100°C.

- Stability of resistance when subjected to a dc cyclic load. Typical power dissipation is limited to give 5 per cent resistance change for 2000 hours continuous operation in a 70°C ambient air temperature.
- Proximity of other heat sources and the flow of cooling air.

The temperature rise of a resistor due to power dissipation is determined by the laws of conduction, convection, and radiation (see Chapter 5). The maximum body external temperature, the hot spot temperature, occurs on the surface - at the middle of the resistor length. As previously considered, any temperature rise will cause a change in resistance, depending on a temperature coefficient; examples are given in table 25.1.
Within the nominal operating temperature range of a resistor, the hot spot temperature, T_h, is given by (similar to equation 5.10)

$$T_h = T_a + R_{\theta_{h-a}}P_d \quad (\text{K}) \tag{25.14}$$

The steady state power dissipation is related to temperature rise (°C or K) and the exposed surface area (m^2). A solid carbon ceramic resistor yields an excellent model for the relationship between temperature and power since it has a homogeneous composition and uniform cross section. For example, heat loss (power dissipated) by radiation and convection in still 25°C air, is given by

$$P_d = 2.6\times\left(\Delta T\right)^{1.4}\times A_{surface} \quad (\text{W}) \tag{25.15}$$

where $\Delta T = T_h - T_a$ and $A_{surface}$ is the exposed radiating surface.

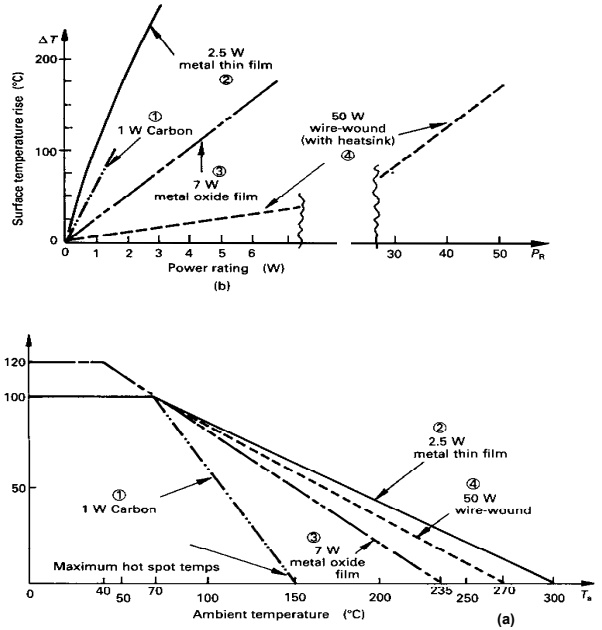


Figure 25.7. *Power resistor thermal properties: (a) power derating with increased temperature and (b) surface temperature increase with increased power dissipation.*

The hot spot temperature is limited, thus as the ambient temperature, T_a, increases the allowable power dissipated decreases, as shown in figure 25.7a for four different elements. These curves show that:

- No power can be dissipated when the ambient temperature reaches the hot spot temperature.
- No derating is necessary below 70°C.
- Some resistors, usually those with higher power ratings, can dissipate higher power at temperatures below 70°C.

The typical linear derating for power and energy of an element in an ambient temperature T_a, can usually be expressed in the form

$$P(T_a) = P_{T\ reference} \times \frac{\hat{T} - T_a}{\hat{T} - T_{T\ reference}} \qquad (25.16)$$

for temperature T_a greater than the reference temperature and $P(\hat{T}) = 0$.
Figure 25.7b shows the resistor surface temperature rise above ambient, nominally 20°C, at different levels of power dissipation. The lead lengths can significantly affect the thermal dissipation properties of resistors and an increase in lead length

- decreases the end of the lead, or soldering spot temperature
- increases the body temperature.

These characteristics are shown for 5 W and 20 W 'cemented' wire-wound resistors in figure 25.8. Figure 25.8a shows how the soldering spot temperature is affected by lead length. Figure 25.8b, on the other hand, is based on the assumption that the soldering spot is represented by an infinite heatsink. Therefore the shorter the lead length, the lower the body temperature for a given power dissipation. It is important to limit the solder pad temperature in order to ensure the solder does not melt – a distinct possibility with continuously dissipating power resistors soldered on pcbs.

25.4.1 Resistors with heatsinking

Aluminium clad resistors suitable for heatsink mounting, as shown in figure 25.1c, are derated with any decrease in the heatsink area from that at which the element is rated. Figure 25.9 shows the derating necessary for a range of heatsink-mounted resistors. For a given heatsink area, further derating is necessary as the ambient temperature increases. Figure 25.7a, curve 4, describes the ambient temperature related power derating of the aluminium-clad resistors on the rated heatsink, characterised in figure 25.9. Figure 25.7a, curve 4, can be used to derate these resistors when operating in an ambient other than 20°C, with the rated heatsink area shown in figure 25.9. The same percentage derating is applicable to a heatsink area smaller than the nominal area.

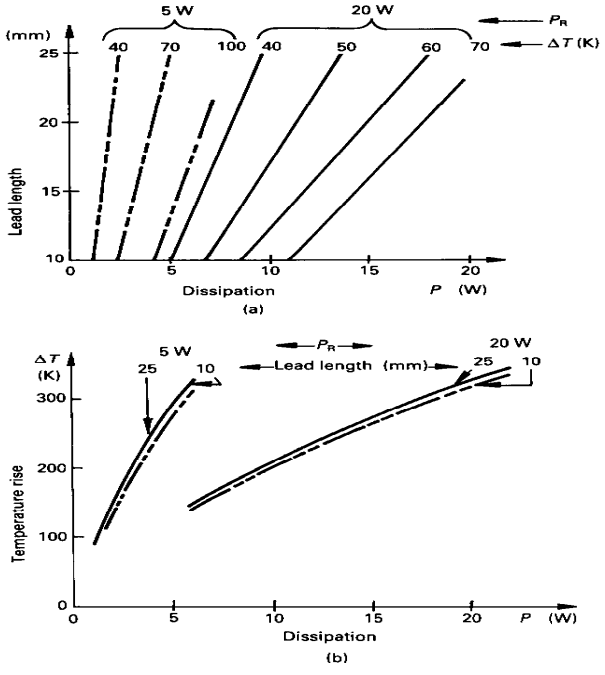


Figure 25.8. *Wire-wound 5 W and 20 W resistor dissipation as a function of: (a) lead length and temperature rise at the end of the lead (soldering spot) and (b) temperature rise of the resistor body, for two lead lengths.*

Example 25.4: ***Derating of a resistor mounted on a heatsink***

What power can be dissipated by an aluminium-clad, wire-wound resistor, (specified in figure 25.9), rated nominally at 50 W, in an ambient of 120°C with a heatsink reduced to 300 cm^2 and 1 mm thick?

Solution

The heatsink area has been reduced to 56 per cent, from 535 to 300 cm^2, hence from figure 25.9 the power rating below 70°C is reduced from 50W to 37.5 W.
From figure 25.7a, curve 4, at 120°C ambient, derating to 75 per cent of the relevant power rating is necessary. That is, 75 per cent of 37.5 W, 28.1 W, can be dissipated at an ambient of 120°C and with a heatsink area of 300 cm^2.

♣

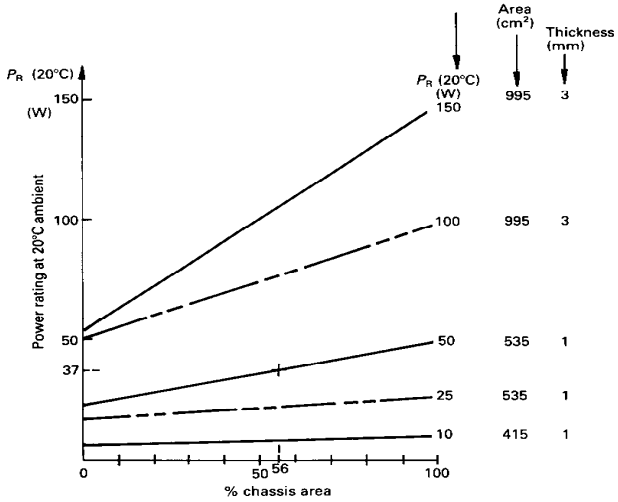


Figure 25.9. *Power dissipation of resistors mounted on a smaller heat sink than specified, right.*

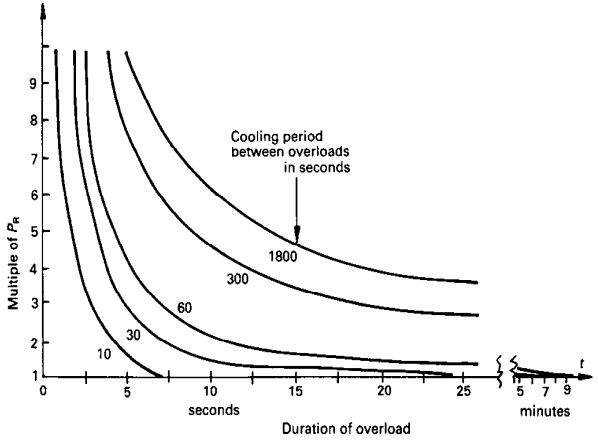


Figure 25.10. *Permissible short time overload ratings for heavy-duty tape wound power resistors.*

25.4.2 Short time or overload ratings

Resistors with power ratings greater than a watt are designed to handle short-term overloads, either continuously for minutes, or repetitively in short bursts of a few seconds. Figure 25.10 can be used to determine allowable short-duration, repetitive pulses. It can be seen that high, short-duration power pulses of a few seconds can be handled if the repetition rate is low. As the pulse duration increases, the overload capability reduces rapidly, with minimal overload allowable with power pulses over a few minutes in duration.

For power pulses of less than 100 ms, the power is absorbed by the thermal capacity of the resistive element and little heat is lost to the surroundings. The temperature rise ΔT of the resistive element in this adiabatic condition is given by (equation 5.2)

$$\Delta T = \frac{W}{mc} \quad \text{(K)} \tag{25.17}$$

where c is the specific heat capacity of the resistive element (J/kg/K)
W is the energy in the pulse of time t_p, (J)
m is the mass of the resistive element (kg).

Due to its homogeneous composition, the carbon ceramic rod resistor yields a good model for the relationship for temperature rise (°C) in free air due to an energy injection pulse. Using the data in appendix 25.8:

$$\Delta T = \frac{W}{mc} = \frac{W}{\gamma \times volume \times c} = \frac{W}{2250 \times \left(\frac{\pi D^2}{4} \times \ell\right) \times 8.89 \times 10^3} = \frac{W}{1.57 \times 10^6 \times D^2 \times \ell} \quad \text{(K)} \tag{25.18}$$

where the effective mass is calculated from the density, γ, and the active volume (all SI units).

Example 25.5: ***Non-repetitive pulse rating***

A 100 A rms, sine pulse with a period of 50 ms is conducted by a wire-wound resistor, constructed of 1 mm^2 cross-section *Ni-Cr* alloy (Nichrome).
Calculate the temperature rise. Assume for *Ni-Cr*

resistivity	$\rho = 1 \times 10^{-6}$ Ω m
specific heat	c = 500 J/kg/K
density	γ = 8000 kg/m^3

Solution

The mass m of the element of length ℓ and area A is given by

$$m = \gamma \ell A \quad \text{(kg)}$$

Resistance R of the wire is given by

$$R = \rho \frac{\ell}{A} \quad (\Omega)$$

The pulse energy is given by

$$W = \int_o^{t_p} i^2 R\, dt = \int_o^{50\text{ms}} \left(\sqrt{2} \times 100 \sin(\omega t)\right)^2 R\, dt = 500R \quad \text{(J)}$$

Substitution for R yields

$$W = 500 \times \rho \frac{\ell}{A} \quad \text{(J)}$$

The temperature rise ΔT from equation (25.17) is given by

$$\Delta T = \frac{W}{mc} = 500 \times \frac{\rho \ell}{A} \times \frac{1}{c} \times \frac{1}{\gamma \ell A} = \frac{500\rho}{\gamma c A^2} = \frac{500 \times 1 \times 10^{-6}}{8000 \times 500 \times \left(1 \times 10^{-6}\right)^2} = 125\text{K}$$

♣

25.5 Repetitive pulsed power resistor behaviour

A resistor may be used in an application where the power pulse experienced at a repetition rate of kilohertz is well beyond its power rating, yet the average power dissipated may be within the rated

power. The allowable square power pulse $\hat{P}$, of duration t_p and repetition time T, can be determined from figure 25.11a, which is typical for power film resistors, at a 70°C ambient. Within these bounds, any resistance change will be within the limits allowable at the continuous power rating. The pulse duration t_p, restricts the maximum allowable pulse voltage $\hat{V}$, impressed across a film resistor, as shown in figure 25.11b.

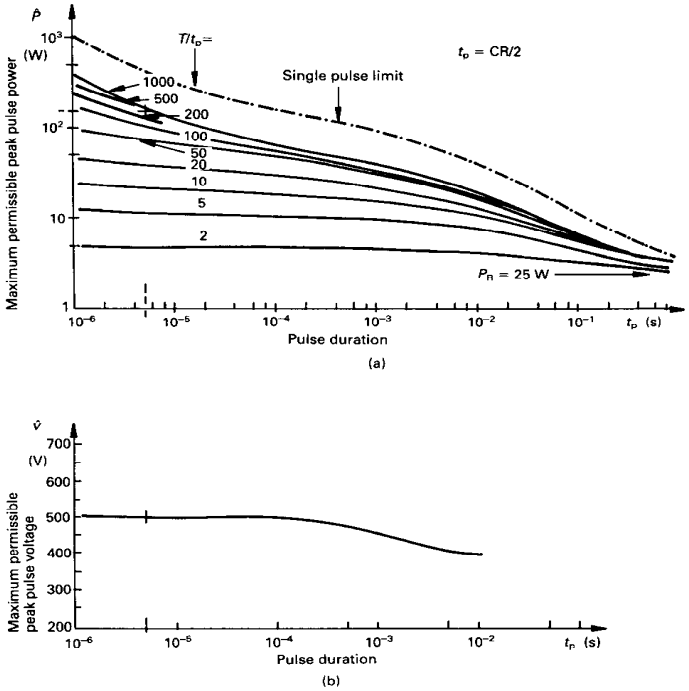


Figure 25.11. *Pulsed capabilities of a power metal film resistor, 2.5 W: (a) maximum permissible peak pulse power versus pulse duration and (b) maximum permissible peak pulse voltage versus pulse duration.*

Example 25.6: ***Pulsed power resistor design***

A 1 kΩ-10 nF, *R-C* snubber is used across a MOSFET which applies 340 V dc across a load at a switching frequency of 250 Hz. Determine the power resistor requirements.

Solution

The average power dissipated (charging plus discharging) in the resistor, which is independent of resistance, is:

$$\overline{P} = CV^2 f = (10 \times 10^{-9}) \times 340^2 \times 250\text{Hz} = 0.29\text{ W}$$

Figure 25.11 is applicable to a 2.5 W metal film resistor, when subjected to rectangular power pulses. The peak power $\hat{P}$ occurs at switching at the beginning of an *R-C* charging or discharging cycle.

$$\hat{P} = V_i^2 / R = (340\text{V})^2 / 1000\Omega = 116\text{ W}$$

where V_i, 340 V, is the maximum voltage experienced across the resistor. The 2.5W element has a power rating greater than the average to be dissipated, 0.29W. Assuming exponential pulses, then

$$t_p = \tfrac{1}{2}\tau = \tfrac{1}{2}CR = \tfrac{1}{2} \times 10 \times 10^{-9} \times 10^3 = 5\mu\text{s}$$

The average pulse repetition time, T, is 2 ms, therefore T / t_p = 400. From Figure 25.11, the peak allowable power is 150 W while the limiting voltage is 500 V. Both the experienced voltage, 340 V, and power, 116 W, are within the allowable limits. The proposed 2.5 W, 1 kΩ, metal thin film power

resistor is suitable. Furthermore from figure 25.7, curves 2, with an average power dissipation of 0.29 W, that is 11.6 per cent of P_R, the maximum allowable ambient temperature is 213°C, while the hot spot temperature is 40°C above ambient for an ambient below 70°C. In terms of average power dissipated, this resistive element is lightly stressed. On the other hand, the transient stress is relatively high.

♣

25.5.1 Empirical pulse power model

An empirical formula for the maximum pulse power may be given in the case of a metal film resistor. Typically, for a 1 W @ 70°C, 700 V rated, 60 K/W, metal film resistor

$$\hat{P} \le \sqrt{\frac{15}{t_p}} \quad \text{(W)} \qquad (25.19)$$

where 1 µs ≤ t_p ≤ 100 ms, such that the average dissipation is less than the rated dissipation.
Using t_p = 5 µs from example 25.2 in equation (25.19) indicates that this 1 W resistor is suitable for the application considered in example 25.4. In fact a ⅓W, 600 V rated, metal film resistor can fulfil the snubber function when using a quoted $\hat{P} = \sqrt{5/t_p}$.

25.5.2 Mathematical pulse power models

A more rigorous mathematical approach to power dissipation is possible for carbon ceramic resistive elements because of the symmetrical shapes and homogeneous composition. The thermal time constant in free air at 25°C is defined by the ratio of the maximum rated energy $\widehat{W}$ to maximum rated power $\widehat{P}$, that is

$$\tau = \frac{\widehat{W}}{\widehat{P}} \qquad (25.20)$$

The temperature rise for a single energy pulse is given by equation (25.17), that is

$$\Delta T = \frac{W}{mc} \quad \text{(K)}$$

If the temperature coefficient of resistance is assumed constant, then the peak temperature rise ΔT_p for *n* repetitive pulses of energy *W* and of period *t*, is

$$\Delta T_p = \Delta T \times \left[\left(1 - \left(e^{-\frac{t}{\tau}} \right)^n \right) + \left(1 - e^{-\frac{t}{\tau}} \right) \right] \qquad (25.21)$$

where *ΔT* is the temperature rise associated with each electrical pulse.
For continuous pulses, this equation asymptotes to

$$\Delta T_p = \Delta T \left[1 + \left(1 - e^{-\frac{t}{\tau}} \right) \right] \qquad (25.22)$$

Example 25.7: ***Solid carbon ceramic resistor power rating***

The resistor in example 25.3 has a maximum power rating of 4W at 25°C and a maximum energy rating of 1200J at 25°C. The element is subject to 2½ms, 1J energy pulses at a 20ms repetition rate.
Determine

i. the thermal time constant
ii. the maximum power and energy limits at 100°C, if the resistor is linearly derated to zero at 200°C, from its rating at 25°C. What is the new thermal time constant?
iii. the temperature rise due to
- one energy pulse
- after 1 second of pulses
- continuous pulses

Solution

i. The thermal time constant at 25°C is defined by equation (25.20)

$$\tau = \frac{\widehat{W}_{25°C}}{\widehat{P}_{25°C}} = \frac{1200\text{J}}{4\text{W}} = 300\text{ s}$$

ii. The power derating is given by

$$P(T) = P_{25°C} \times \left(\frac{200°\text{C} - \text{T}}{200°\text{C} - 25°\text{C}} \right)$$

$$P(100°\text{C}) = 4\text{W} \times \left(\frac{200°\text{C}-100°\text{C}}{175°\text{C}} \right) = 2.28\text{W}$$

Similarly, the energy derating is given by

$$W(T) = W_{25°C} \times \left(\frac{200°\text{C} - \text{T}}{175°\text{C}} \right)$$

$$W(100°\text{C}) = 1200\text{J} \times \left(\frac{200°\text{C}-100°\text{C}}{175°\text{C}} \right) = 686\text{J}$$

The thermal time constant remains unchanged after the two linear transformations

$$\tau = \frac{\widehat{W}_{100°C}}{\widehat{P}_{100°C}} = \frac{686\text{J}}{2.28\text{W}} = 300\text{ s}$$

iii. The temperature rise due to one energy pulse (20ms repetition rate) is given by equation

$$\Delta T = \frac{W}{mc} = \frac{W}{1.57\times10^6 \times D^2 \times \ell} \quad \text{(K)}$$

$$= \frac{1\text{J}}{1.57\times10^6 \times \frac{4}{\pi} \times 0.25\times10^{-4} \times 1\times10^{-2}} = 2°\text{C}$$

From equation (25.21) after 1 second, that is 50 pulses

$$\Delta T_{p-2} = \Delta T \times \left[\left(1 - \left(e^{-\frac{t}{\tau}} \right)^n \right) + \left(1 - e^{-\frac{t}{\tau}} \right) \right]$$

$$= 2°\text{C} \times \left[\left(1 - \left(e^{-\frac{20\text{ms}}{300\text{s}}} \right)^{50} \right) + \left(1 - e^{-\frac{20\text{ms}}{300\text{s}}} \right) \right] = 2°\text{C} \times \left[0.003328 + 0.0000666 \right] = 0.0007°\text{C}$$

In steady state, from equation (25.21), the peak temperature rise is

$$\Delta T_p = \Delta T \times \left[1 + \left(1 - e^{-\frac{t}{\tau}} \right) \right]$$

$$= 2°\text{C} \times \left[1 + \left(1 - e^{-\frac{20ms}{300s}} \right) \right] = 2°\text{C} \times \left[1 + 0.0000666 \right] = 2.0°\text{C}$$

♣

All resistors are thermally derated, starting at about 70°C, linearly to zero power dissipation at a maximum operating temperature, which is shown in Table 25.5 for the various resistor types.

Table 25.2: Zero rated power for thermally derated resistor types

Type (derating usually starts at 70°C)	Maximum temperature °C
Wire wound - power alumina former/ power beryllia former / precision	275 / 350 / 175
Metal foil	125
Metal foil – power / precision	170 / 175
Nickel film	150
Tantalum nitride film	150
Cermet thin film	200
Metal oxide film	125
Carbon film – power / precision	100 / 120
Metal glazed – power / precision	200 / 200
Thick film - metal oxide / metal oxide power	150 / 275
Conducting plastic – power / precision	120 / 120
Carbon composition / ceramic	130 / 220

25.6 Stability and endurance

The resistance stability of a resistor is dependent on power dissipation, ambient temperature, and resistance value. An endurance test gives the worst-case variation in resistance value or stability. It is the percentage resistance change at rated power and hot spot temperature after a specified time. An endurance specification is of the form:

- 1000 hours at recommended maximum dissipation P_R,
- which will limit the hot spot temperature to 375°C:
- ΔR less than 5 per cent of R

The time, percentage change in R, and temperature are varied with resistor type and physical size.

At power levels below rated dissipation, better stability than that for the endurance test is attainable, for the same duration. The stability period can be extended by the following empirical formula

$$\left.\frac{\Delta R}{R}\right|_t \approx X^{\log\frac{t}{t_1}} \times \left.\frac{\Delta R}{R}\right|_{t_1} \qquad (25.23)$$

which is valid for $10^{-3} \le t \le 10^5$ hours. The base χ depends on the resistor type and is between 1.1 and 5. Performance monograms as shown in figure 25.12 may be provided to enable a given resistor to be used at dissipation levels which will result in the stability required for that application. The first quadrant in figure 25.12 satisfies the thermal equation (25.14), while the third quadrant satisfies equation (25.19), with χ = 3.17. The following example illustrates many of the features of the stability performance monogram of figure 25.12.

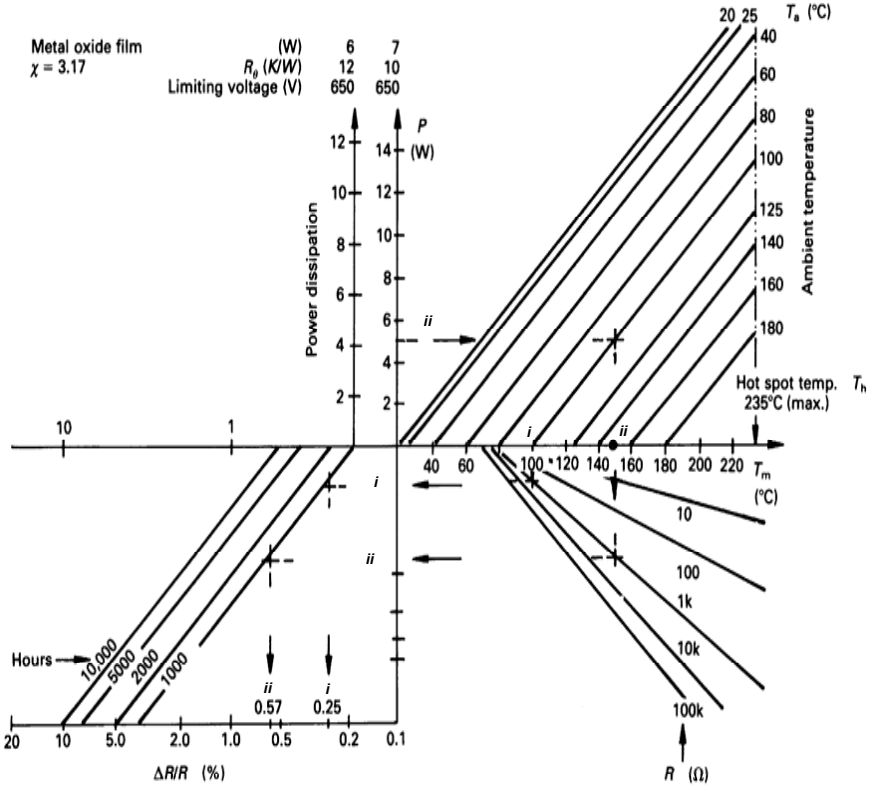


Figure 25.12. *Performance monogram for power resistors, showing the relationship between power dissipation, ambient temperature, hot spot temperature, and maximum resistance drift in time.*

Example 25.8: *Power resistor stability*

A 1 kΩ, 7 W, power metal oxide film resistor dissipates 5 W. If the maximum ambient is 100°C, use the monograph in figure 25.12 to find

i. the stability at 100°C while in circuit for 1000 hours but in a standby mode, that is, P = 0 W
ii. the hot spot temperature when dissipating 5 W
iii. the maximum expected resistance drift after 10^3 and 10^5 hours
iv. lifetime (5% = $\Delta R/R$) dissipating 5W in a 60°C ambient.

Solution

i. The resistance change given from the monogram for P = 0W at a 100°C ambient is indicative of the shelf-life stability of the resistor when stored in an 100°C ambient.
The stability is determined by performing the following operations. Find the intersection of P = 0 and the diagonal for T_a = 100°C. Then project perpendicularly to the 1 kΩ diagonal. The intersection is projected horizontally to the 1000 hour diagonal. This intersection is projected perpendicularly to the stability axis.
For example, from projections on figure 25.12, after 1000 hours, in a 100°C ambient, a 0.25 per cent change is predicted. For the 1 kΩ resistor there is a 95 per cent probability that after 1000 hours the actual change will be less than 2.5 Ω (0.25% of 1kΩ).

ii. The 5 W load line is shown in figure 25.12. A hot spot temperature, T_m, of 150°C is predicted (100°C + 5 W × 10°C/W).

iii. With a 1 kΩ resistor after 1000 hours, a $\Delta R/R$ of 0.57 per cent is predicted, as shown on figure 25.12. There is a 95 % probability that the actual change will be less than 5.7Ω.
The monogram does not show stability lines beyond 10,000 hours. Equation (25.23), with χ=3.17, can be used to predict stability at 100,000 hours

$$\left.\frac{\Delta R}{R}\right|_{10^5} \approx 3.17^{\log\frac{10^5}{10^3}} \times 0.57\% = 18.25\%$$

After 100,000 hours, it is 95 per cent probable that the actual resistance change of the 1 kΩ resistor will be less than 182.5 Ω (18¼% of 1kΩ).

iv. from the first quadrant in figure 25.12, the hot spot temperature is 110°C when dissipating 5W in a 60°C ambient. Solving equation (25.23) for 5% = $\Delta R/R$ = 5%:

$$\left.\frac{\Delta R}{R}\right|_{10^x} = 3.17^{\log\frac{10^x}{10^3}} \times 0.05$$

gives x = 5. That is, there is a 95% probability that the resistor will survive 10^5 hours with a resistance variation of less that 5%.

♣

25.7 Special function power resistors

Film and wire-wound resistors are available which have properties allowing them to perform the following functions

- fusing
- circuit breaking
- temperature sensing
- current sensing.

Table 25.3: Fusible resistor characteristics

parameter		condition	units	Metal alloy film	Wire wound
Power	P_R	@ 70°C	W	0.25 - 4.5	1 - 2
Resistance range	R		Ω	0.22 - 10k	0.1 - 1k
Tolerance			%	2	5
Temperature coefficient (resistance dependent)	α	x 10^{-6}	/K	± 500	-400 to + 1000
Stability		@ P_R T_a = 70°C, 1000 hours	%	2	10
Working voltage	V_m		V	$\sqrt{P_R R}$	$\sqrt{P_R R}$

25.7.1 Fusible resistors

Resistors up to 2 W are available which fuse when subjected to an overload current. The resistive element fused is generally metal alloy film, although only wire-wound elements are suitable at low resistance levels. The power load and interruption time characteristic shown in figure 25.13a shows that rated power can be dissipated indefinitely, while as the power increases significantly above the rated dissipation, the interruption time decreases rapidly. Interruption generally means that the nominal resistance has increased at least 10 times. Irreversible resistance changes can be caused by overloads which raise the change in hot spot temperature beyond 150°C, for the elements illustrated by figure 25.13b. The nature of the resistive element makes it unsuitable for repetitive power pulse applications. Typical fusible resistors are summarised in table 25.3.

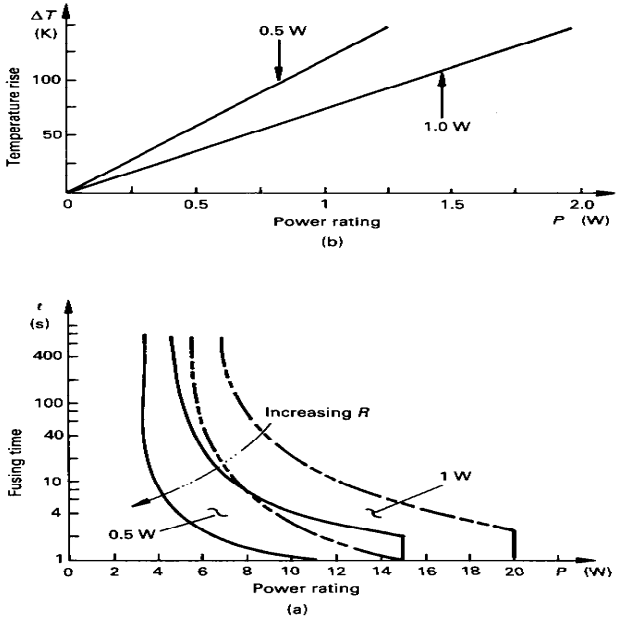


Figure 25.13. *Fusible resistor characteristics: (a) time to interruption (10 x R_R as a function of overload power and (b) temperature rise above ambient as a function of power dissipated.*

25.7.2 Circuit breaker resistors

The construction of two types of wire-wound circuit breaker resistors is shown in figure 25.14. Under overload conditions the solder joint melts, producing an open circuit. After fusing, the solder joint can be resoldered with lead free solder.
The joint melts at a specified temperature, and to ensure reliable operation the solder joint should not normally exceed 150°C. This allowable temperature rise is shown in figure 25.15a, while the circuit breaking time and load characteristics for both constructions are shown in figure 25.15b. This characteristic is similar to that of fusible resistors. A typical power range is 1 to 6 W at 70°C, with a resistance range of 75 mΩ to 82 kΩ and temperature coefficient of -80 to +500 × 10^{-6}/K depending on the resistance values. The maximum continuous rms working voltage tends to be limited by the power, P_R, according to $\sqrt{P_R R}$.

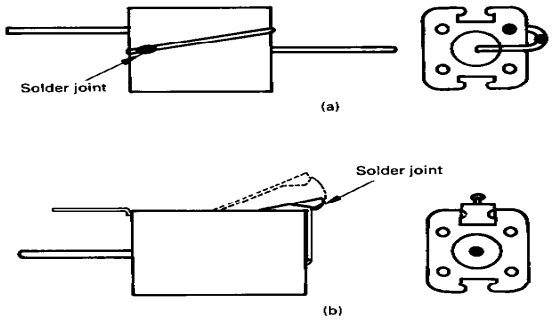


Figure 25.14. *Two circuit breaker resistors construction: (a) type 1 and (b) type 2.*

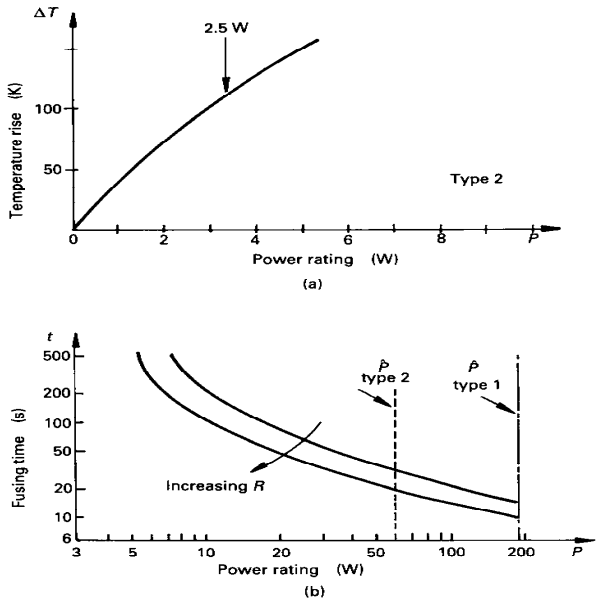


Figure 25.15. *Circuit breaker characteristics: (a) solder joint temperature rise versus power dissipated for resistor type 2 and (b) fusing times versus load for resistor types 1 and 2.*

25.7.3 Temperature sensing resistors

The temperature dependence of a resistive element can be exploited to measure temperature indirectly. Unlike normal resistors, temperature-sensing resistors require a large temperature coefficient to increase resistance variation sensitivity with temperature.
Both metal film and wire-wound temperature sensing elements are available with a temperature coefficient of over +5000 × 10^{-6}/K with 1 per cent linearity over the typical operating range of -55°C to 175°C. The high temperature sensitivity gives a 57 per cent increase in resistance between 25°C and 125°C.
The low power elements, up to ¼W @ 70°C, tend to be metal oxide, with a typical resistance range at 25°C of 10 Ω to 10 kΩ. A conformal encapsulation is used to minimise thermal resistance, hence ensuring an extremely fast response. For example, a 1/20 W temperature sensing resistor can have a thermal time constant to a step temperature, in still air, of 3.7 s. This time constant increases to 31 s for a larger mass, ¼ W rated resistor.

At powers commencing at 1 W, wire-wound elements are employed which utilise positive temperature coefficient alloy resistance wire. The nominal resistance range is lower than the film types; typically from 0.1 Ω to 300 Ω at 25°C. Response is slower than film types, but can be improved if oil immersed.
Glazed thick film temperature sensing resistors can be used to provide a negative temperature coefficient, -3000 × 10^{-6}/K at 25°C. Power is limited to ¼ W with a dissipation constant of up to 8.1 mW/K at 25°C in still air. Low thermal time constants of only 2.9 s are possible with 1/20 W elements.

The allowable working voltage for all types is power limited, $V=\sqrt{P_R R}$.

25.7.4 Current sense resistors

The resistive element consists of a flat metal band, with spot-welded terminals, and a ceramic encapsulation. The flat-band construction results in a non-inductive resistor of both high stability and overload capacity. Low current third and fourth terminal voltage sensing (Kelvin) leads may be incorporated; alternatively a mΩ/cm correction factor for lead length is given.
Power ratings of up to 20 W at 70°C and 20 A maximum, with a resistance range of 10 mΩ to 10 Ω are available. At these low resistance levels, the maximum continuous working voltage is power limited. The resistance temperature coefficient is typical of a wire-wound resistor, 100 to 600 × 10^{-6}/K depending on the resistance level.

25.7.5 Thermistors

A *thermistor* is a type of resistor with resistance varying predictably and rapidly according to its temperature. The word is a portmanteau of thermal and resistor.
Thermistors are widely used as inrush current limiters, temperature sensors, self-resetting overcurrent protectors, and self-regulating heating elements.
The resistance of a thermistor is solely a function of its absolute body temperature. When testing for resistance accuracy it is essential that the surrounding environmental temperature is held at a constant, and power dissipated in the thermistor is low enough to insure no 'self-heating'.
With transition metals, the relationship between resistance and temperature is linear, first-order, that is:

$$\Delta R = k \Delta T \qquad (25.24)$$

where:
ΔR = change in resistance
ΔT = change in temperature
k = first-order temperature coefficient of resistance

If k is positive, the resistance increases with increasing temperature, and the metal is termed a positive temperature coefficient (PTC) material. If k is negative, as with some sintered compounds, the resistance decreases with increasing temperature, and the compound is termed a negative temperature coefficient (NTC) material.
Thermistors can be classified into two types depending on the sign of k, namely NTC and PTC thermistors. Resistors that are not thermistors are designed to have a k as close to zero as possible, so that their resistance remains near constant over a wide temperature range.

NTC thermistors are made from chemically stabilised metallic oxides (such as manganese, iron, cobalt, nickel, copper and zinc), compressed and sintered at high temperatures between 1000°C and 1400°C to produce the polycrystalline NTC thermistor, which at a finally stage of manufacture is aged for stability.

Most PTC thermistors are of the 'switching' type, which means that their resistance rises suddenly at a certain critical temperature. The devices are made of a doped polycrystalline ceramic containing barium titanate ($BaTiO_3$) and other compounds. The dielectric constant of this ferroelectric material varies with temperature. Below the Curie point temperature, the high dielectric constant prevents the formation of potential barriers between the crystal grains, leading to a low resistance. In this region the device has a small negative temperature coefficient. At the Curie point temperature, the dielectric constant drops sufficiently to allow the formation of potential barriers at the grain boundaries, and the resistance increases sharply. At even higher temperatures, the material reverts to NTC behaviour.
Another type of PTC thermistor is the polymer PTC. This consists of a slice of plastic with carbon grains embedded in it. When the plastic is cool, the carbon grains are all in contact with each other, forming a conductive path through the device. When the plastic heats up, it expands, forcing the carbon grains apart, and causing the resistance of the device to rise rapidly. Like the $BaTiO_3$ thermistor, this device has a highly nonlinear resistance/temperature response and is used for switching, not for proportional temperature measurement.

Another type of thermistor is a Silistor - a thermally sensitive silicon resistor. Silistors are similarly constructed and operate on the same principles as other thermistors, but employ silicon as the semiconductive component material. Over small changes in temperature, if the right semiconductor is used, the resistance of the material is linearly proportional to the temperature. There are many different semiconducting thermistors with a range from 0.01 degree Kelvin to 2,000 K; -273.14°C to 1,700°C).

The NTC thermistor is best suited for precision temperature measurement. The PTC thermistor is best suited for temperature compensation and current limiting, as considered, in detail, in chapter 10.3.2.

Negative temperature coefficient (NTC) thermistors

In practice, the linear approximation in equation (25.24) is only applicable over a small temperature range. For accurate temperature measurements, the resistance/temperature curve of the device is be described in more detail using the Steinhart-Hart third-order approximation equation:

$$\frac{1}{T} = a + b \ln R + c \ln^3 R$$

where a, b and c are called the Steinhart-Hart parameters, and must be specified for each device. T is the temperature in degree Kelvin and R is the resistance in Ohms. To give resistance as a function of temperature, the above equation can be rearranged into:

$$R = e^{(\beta - \frac{1}{2}\alpha)^{\frac{1}{3}} - (\beta + \frac{1}{2}\alpha)^{\frac{1}{3}}}$$

where

$$\alpha = \frac{a - \frac{1}{T}}{c} \text{ and } \beta = \left(\left(\frac{b}{3c} \right)^3 + \tfrac{1}{4}\alpha^2 \right)^{\frac{1}{2}}$$

The error in the Steinhart-Hart equation is generally less than 0.02°C in the measurement of temperature. As an example, typical values for a thermistor with a resistance of 3kΩ at room temperature (25°C = 298.15 K) are:

$$a = 1.40 \times 10^{-3}$$
$$b = 2.37 \times 10^{-3}$$
$$c = 9.90 \times 10^{-8}$$

Temperature coefficients valid over a small temperature range.
i. NTC thermistors can also be characterised with the B parameter equation, which is essentially the Steinhart Hart equation with c=0.

$$\frac{1}{T} = \frac{1}{T_o} + \frac{1}{B} \ln \frac{R}{R_o} \qquad \left(\text{or } T = \left(\frac{1}{B} \ln \left(\frac{R}{R_o} \right) + \frac{1}{T_o} \right)^{-1} \qquad (\text{K}) \right)$$

where the temperatures are in degree Kelvin. Rearranging yields:

$$R = R_o e^{B\left(\frac{1}{T} - \frac{1}{T_o}\right)} \qquad (25.25)$$

R_o and T_o are rated temperature and resistance at that temperature (usually 25°C=298.15K), or

$$B = \frac{T\,T_o}{T_o - T} \ln \frac{R}{R_o} = \frac{T\,T_o}{T_o - T} \times 2.3026 \times \log_{10} \frac{R}{R_o}$$

The B-parameter equation can also be arranged in the form $y = mx + c$ by taking natural logs of both sides of the equation to give $\ell n R = Bx + 1 / T$. This can be used to convert the function of resistance against temperature of a thermistor, into a linear function of which the gradient can be found to give the B value (typically 2500K to 5000K. A quadratic approximation for B can be highly accurate, where

$$B = a + bT + cT^2$$

The quadratic coefficients are device material dependant, typically $2000 < a < 4000$, $1 < b < 8$, and $-0.0001 < c < -0.01$.

ii. Also, in conjunction with equation (25.25), the temperature coefficient of resistance or alpha, α, of an NTC thermistor is defined as:

$$\alpha = \frac{1}{R} \frac{dR}{dT} = -\frac{B}{T^2}$$

Conduction model

Many NTC thermistors are made from a pressed disc or cast chip of a semiconductor such as a sintered metal oxide. They work because raising the temperature of a semiconductor increases the number of electrons able to move about and carry charge - it promotes them into the conduction band. The more charge carriers that are available, the more current a material can conduct. This is described in the formula:

$$I = qnAv \qquad (25.26)$$

I = electric current (ampere)
n = density of charge carriers (count/m³)
A = cross-sectional area of the material (m²)
v = velocity of charge carriers (m/s)
q = charge of an electron q = 1.602x10^{-19} coulomb

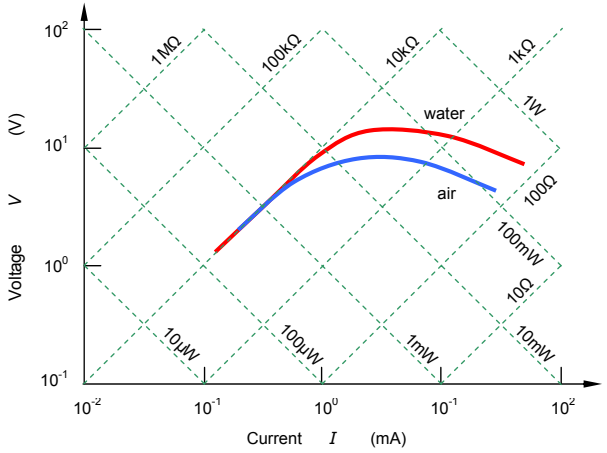


Figure 25.16. *I-V characteristics of a 10kΩ NTC thermistor at 25°C.*

The NTC thermistor has three regions

- The straight line section is for negligible self heating, thus V and I are proportional, thus the resistance is determined only by the temperature. This is the temperature sensing region with $dV/dI = R$=constant
- The non-linear rise and drop is due to self heating. At maximum voltage the relative decrease in resistance $\Delta R/R$ resulting from self-heating is equal to the relative increase in current, $\Delta I/I$, implying $dV/dI = R$= constant.
- The falling edge is when the decrease in resistance is greater than the relative increase in current. This self heating region is used for current inrush control and liquid level sensor, that when dV/dI <0.

Self-heating effects

Though commonly used, 'self-heating' is a misnomer. Thermistors are passive devices and thus cannot heat themselves. It is the external circuit that supplies the energy that causes the heating. 'Resistive heating' or Joule heating are more accurate terms.

When a current flows through a thermistor, it generates heat which raises the temperature of the thermistor above that of its environment. If the thermistor is being used to measure the temperature of the environment, this electrical heating may introduce a significant error if a correction is not made.

The electrical power input to the thermistor is

$$\frac{dH}{dt} = P_e = IV = I^2R = U(T - T_A) + C_{th}\frac{dT}{dt}$$

where I is current and V is the voltage drop across the thermistor.
dH/dt is the change of stored thermal energy with time
U is the dissipation factor of the thermistor = dP/dT, $U = h \times A$, W/K
T is the instantaneous NTC temperature

T_A ambient temperature
C_{th} heat capacity of the NTC thermistor = $\Delta H/\Delta T = U \times \tau_c$
dT/dt change of temperature with time

The solution to the thermal differential equation is

$$\Delta T = T - T_A = \frac{P_e}{U}\left(1 - e^{-\frac{t}{\tau_c}}\right) \quad \text{where} \quad \tau_c = \frac{C_{th}}{U} = \frac{C_{th}}{h \times A}$$

where τ_c is the thermal cooling time constant (63.2% decrease in temperature time, in still air at 25°C), from 85°C to 47.1°C. The thermal time constant τ_a is time it takes for an unload thermistor to increase its body temperature from 25°C to 62.9°C in a fluid, typically water, at 85°C.

This power is converted to heat, and this heat energy is transferred to the surrounding environment. At thermal equilibrium, when dT/dt=0 or $t >> \tau_c$, the rate of transfer is well described by Newton's law of cooling:

$$P_T = U \times (T(R) - T_o) = U\,\Delta T$$

where $T(R)$ is the temperature of the thermistor as a function of its resistance R, T_o is the temperature of the surroundings, and δ is the dissipation constant, usually expressed in units of milliwatts per °C. At equilibrium, the two rates must be equal.

$$P_T (= U\,\Delta T) = P_e (= I^2R)$$

The current and voltage across the thermistor will depend on the particular circuit configuration. As a simple example, if the voltage across the thermistor is held fixed, then by Ohm's Law we have $I = V / R$ and the equilibrium equation can be solved for the ambient temperature as a function of the measured resistance of the thermistor:

$$T_o = T(R) - \frac{V^2}{U \times R}$$

or

$$I = \left(\frac{U \times \Delta T}{R(T)}\right)^{1/2} \quad \text{and} \quad V = \left(U \times \Delta T \times R(T)\right)^{1/2}$$

The dissipation constant is a measure of the thermal connection of the thermistor to its surroundings. It is generally given for the thermistor in still air, and in well-stirred oil. Typical values for a small glass bead thermistor are 1.5 mW/°C in still air and 6.0 mW/°C in stirred oil. If the temperature of the environment is known beforehand, then a thermistor may be used to measure the value of the dissipation constant. For example, the thermistor may be used as a flow rate sensor, since the dissipation constant increases with the rate of flow of a fluid past the thermistor.

The NTC thermistor is used in three different modes which services a variety of applications.

i. Resistance versus Temperature Mode - the most prevalent. These circuits perform precision temperature measurement, control and compensation. Unlike the other applications this method depends on the thermistor being operated in a 'zero-power' condition. This condition implies that there is no self-heating.

The resistance across the sensor is relatively high in compensation to an RTD element, which is usually in the hundreds of ohms range. Typically, the 25°C rating for thermistors is from 10Ω to 10,000,000 Ω. The housing of the thermistor varies as the requirements for a hermetic seal and ruggedness, but in most cases, there are only two wires going to the element. This is possible because of the resistance of the wire over temperature is considerably lower than the thermistor element. And typically does not require compensation because of the overall

ii. Current-Over-Time Mode – This depends on the dissipation constant of the thermistor package as well as element's heat capacity. As current is applied to a thermistor, the package will begin to self-heat. If the current is continuous, the resistance of the thermistor will start to lessen. The thermistor current-time characteristics can be used to slow down the affects of a high voltage spike, which could be for a short duration. In this manner, a time delay from the thermistor is used to prevent false triggering of relays.

This type of time response is relatively fast as compared to diodes or silicon based temperature sensors. In contrast, thermocouples and RTD's are equally as fast as the thermistor, but they don't have the equivalent high level outputs.

iii. Voltage versus Current Mode - Voltage-versus-current applications use one or more thermistors that are operated in a self-heated condition. An example of this would be a flow meter. The thermistor would be in an ambient self-heated condition. The thermistor's resistance is changed by the amount of

heat generated by the power dissipated by the element. Any change in the media (gas/liquid) across the device changes the power dissipation factor of the thermistor. The small size of the thermistor allows for this type of application to be implemented with minimal interference to the system.

Aging affects on thermistor stability

'Thermometric drift', the main cause for NTC thermistor drift, is fixed temperature drift at all temperatures of exposure. For example, a thermistor that exhibits a -0.02°C shift at 0°, 40° and 70°C (even though this is a different percentage change in resistance in each case) would be exhibiting thermometric drift. Thermometric drift:

(1) occurs over time at varying rates, based on thermistor type and exposure temperature, and
(2) generally, increases as the exposure temperature increases.

Thermistor failure

i. *Silver Migration* This failure can occur due to one or more of: constant direct current bias, high humidity, and electrolytes (disc/chip contamination). Moisture finds its way into the thermistor and reacts. Silver (on the thermistor electrodes) turns to solution, and the direct current bias stimulates silver crystal growth across the thermistor element. The thermistor resistance decreases, eventually producing a short circuit.

ii. *Micro Cracks* Thermistors can crack due to improper potting materials if a temperature change causes contraction of the potting material, crushing the thermistor. The result is a thermistor that has erratic resistance readings and is electrically 'noisy'.

iii. *Fracture of Glass Envelope* Typically caused by mishandling of thermistor leads, this failure mechanism induces fractures in the glass coating at the lead/thermistor interface. These cracks may propagate around the thermistor bead resulting in increased resistance. Mismatching of epoxies or other bonding materials may also cause fracture.

iv. *Aging out of Resistance Tolerance* If thermistors are exposed to high temperatures over time, termed 'aging', their resistivity can change. Generally the resistivity increases, which results in a deceased temperature. Temperature cycling may be thought of as a form of aging. It is the cumulative exposure to high temperature that has the greatest influence on a thermistor component, not the actual temperature cycling. Temperature cycling can induce shifts if the component has been built into an assembly with epoxies or adhesives, which do not match the temperature expansion characteristics of the thermistor.

25.7.6 Other specialised resistors

Table 25.4: Other resistor types and uses

variable	type	material	application
temperature	thermistors	NTC - semiconductor oxide PTC - barium titanate	temperature control amplifier gain control voltage regulation – NTC current regulation – PTC flow control
strain	strain gauges	metal wire or foil	low pressure sensing
voltage	varistors (see chapter 10)	metal oxide silicon carbide	transient voltage protection
pressure	transducers microphones	carbon, ceramic	control application
humidity	-	carbon, metal film, thick film metal oxide	humidity sensors

25.8 Appendix: Carbon ceramic electrical and mechanical data and formula

Electrical parameter	units	value	mechanical	units	value
resistivity	Ω cm	3 to 30000	density	g/cm^3	2.25
Temperature coefficient	%per °C	-0.05 to -15	Coefficient of linear expansion	/ °C	+4×10^{-6} to +10×10^{-6}
Voltage coefficient	% /kV/cm	-0.5 to -7.5	Bending strength	kg.m for 15cm diameter	30 to 60
Dielectric constant		5	Youngs modulus	N.cm^{-2}	3×10^6
inductance		negligible	Crushing strength	N cm^{-2}	12000
Thermal conductivity	W/cm^2 °C/cm	0.04			
Specific heat capacity	J cm^3 / °C	2.0			

$$\alpha = +1600 \times e^{-\log \rho / 1.4} - 1350 \qquad \left(10^{-6}/°C \text{ or } ppm/°C\right)$$

$$\phi = -0.62 \times \rho^{0.22} \qquad (\%/kV/cm)$$

$$P_d = 2.6 \times (\Delta T)^{1.4} \times A_{surface} \qquad (W)$$

25.9 Appendix: Characteristics of resistance wire

Alloy type[a]	Resistivity, Ω cmil ft^{-1} 20 °C	Resistivity,[b] Ω cm^2 cm^{-1} × 10^6 20 °C	Resistance temp coefficient,[c] ppm °C^{-1}	Linear expansion thermal coefficient, cm/cm/°C × 10^6, 20–100 °C	Min tensile strength lb in^{-2} 25 °C[d]	Melting temp (approx), °C	Relative magnetic attraction	Density, g cm^{-3}, 20 °C	Heat capacity[e] J g^{-1} °C^{-1}	Thermo-electric potential to copper[f] (approx), V °C^{-1} × 10^6
80–20 Ni-Cr	650–675	108–112	+60 to +90 ± 20	12–18	100,000	1400	None	8.41	0.435	+6.0
Constantan	294–300	49–50	0 ± 20	14.5	60,000	1350	None	8.90	0.393	–45
Manganin	230–290	38–48	0 ± 15[c]	18.7	40,000	1020	None	8.192–8.41	0.406	–3.0, +1
Alloy 180	180	29.9	+180 ± 30	15.7–17.5	50,000	1100	None	8.90	0.385	–37
Alloy 90	90	14.9	+450 ± 50	16–17.5	35,000	1100	None	8.90	0.385	–26
Alloy 60	60	9.97	+500 to +800 ± 200	16.2–16.3	50,000	1100	None	8.90	0.385	–22
Alloy 30	30	4.99	+1400 to +1500 ± 300	16.4–16.5	30,000	1100	None	8.91	0.385	–14
Linear TC[a]	120	19.9	+4500 ± 400	12–15	70,000	1100	Strong	8.46	0.523	–40
Nickel A	60	9.97	+4800	13	60,000	1450	Strong	8.90	0.544	–22
High purity Ni[a]	50	8.31	+6000	13.3–15	50,000	1400	Strong	8.90	0.544	–22
Iron	61.1	10.15	+5000 to +6200	11.7 (20 °C)	50,000	1535	Strong	7.86	0.445	+12.2
Copper	10.37	1.72	+3900 to +4300	16.5 (20 °C)	35,000	1083	None	8.90	0.385	0
Evanohm[g]	800	133	0 ± 5	12.6	100,000	1350	None	8.10	0.448	+3.0
Karma[h]	800	133	0 ± 20	13.3	180,000	1400	None	8.10	0.435	+3.0
Alloy 800[i]	800	133	0 ± 5	15	150,000	1260	None	7.95		+2.5
Chromel R[j]	800	133	0 ± 10	13.5	95,000	1398	None	8.1	0.448	+1.0
Moleculoy[k]	800	133	0 ± 5	13.3	130,000	1395	None	8.12	0.435	+3.0
Nikrothal LX[l]	800	133	0 ± 5	12.6	150,000	1410	None	8.1	0.460	+2.0
Kanthal DR[l]	812	135	0 ± 20	11.9	100,000	1505	Strong	7.2	0.494	–3.5
Mesoloy[k]	825	137.2	0 ± 10	13.5	100,000	1500	Strong	7.15	0.481	–3.3
Alloy 815[j]	815	135.5	+82	15.9	~115,000	1520	Strong	7.25	0.460	~–3.7
Alloy K-20[j]	815	135.5	0 ± 20	13	100,000	1530	Strong	7.25	~0.460	–3.5
Evanohm S[g]	825	137	0 ± 5	13	100,000	1350	None	7.13	0.460	+0.2

Metal	nΩm Resistivity
Aluminum	22.7
Copper	17.0
Gold	23.0
Iron	105.0
Nickel	78.0
Platinum	106
Silver	16.0
Tin	115
Tungsten	55.0
Zinc	62.0
Carbon-steel	180
Brass	60
Constantan	450
Invar	100
Manganin	430
Nichroma	1105
Nickel-silver	272
Monel metal	473
Kovar	483
Phosphor-bronze	93
18/8 stainless steel	897

25.10 Appendix: Preferred resistance values of resistors (and capacitors)

Specific resistance values have been standardised based on logarithmic values $\sqrt[E]{10}$ or $10^{\frac{1}{E}}$ where E is the number of resistors (logarithmic steps) within each decade. If the first resistance is one per unit, then the geometric progression for the E resistors is given by $1 \times \sqrt[E]{10}^{\,n-1}$, for $n = 1$ to E. Common E values (IEC 60063) are *3*, 6, 12, 24, 48, 96, and 192. The E3 range is rare for resistors but not uncommon with capacitors, especially electrolytic capacitors which tend to have wider tolerances due to manufacturing processing limitations and constraints.

The E3 range has decade values based on 10, 22 (21.54), and 47 (46.4), which give a 50% error band.
The E6 range of 10, 15, 22, 33, 47, and 68 preferred values have a maximum error of 20%.
The E12 range of 10, 12, 15, 18, 22, 27, 33, 39, 470, 56, 68, 75, and 82 has a 10% error band.
The E24, E48, E96, and E192 ranges have resistance tolerance error bands of 5%, 2%, 1% and better than 1%, respectively. Resistors with a tolerance of better than ±½% are termed precision resistors.

Resistors are physically coded with 4, 5, or 6 colours bands, with numerical values (plus tolerance and temperature coefficient) assigned to the colours as shown below and in figure 25.1. Capacitors are not colour coded.

0 black		
1 brown	±1%	100ppm
2 red	±2%	50ppm
3 orange		15ppm
4 yellow		25ppm
5 green	±½%	
6 blue	±¼%	
7 violet	$\pm\frac{1}{10}$%	
8 grey	$\pm\frac{1}{20}$ %	
9 white		
$\times\frac{1}{10}$ gold	±5%	
$\times\frac{1}{100}$ silver	±10%	

EIA-RS-279

Reading list

Philips, Book 3, Part 1c, Fixed Resistors, 1985.

Philips, Book 3, Part 1f, Varistors, Thermistors and Sensors, 1986.

Siemens, Components, 1986.

Siemens, NCT Thermistors, Data Book, 1986/87.

Rohm.com

Vishay.com

Hvrint.com

Problems

25.1. The resistance of a temperature dependent negative temperature coefficient resistor is given by

$$R(T) = Ae^{\frac{B}{T}}$$

where B = 3300 K and R = 10 Ω at 25°C. The resistive element has a thermal resistance to ambient of 45.5 K/W. Assume the maximum resistor temperature is 1000°C and absolute zero is -273°C.

(1) Calculate the temperature coefficient at 25°C.
[-0.037 or -3.7 per cent/K]

(2) When the non-linear resistor is dissipating power, what is the maximum attainable terminal voltage? At what temperature and current does this voltage occur at? Assume the ambient temperature is 25°C.
[58.3°C, 3.29 Ω, 0.73 W, 1.55 V, 0.47A]

(3) The non-linear resistance dependence on temperature can be 'linearised' by the parallel connection of a resistor. The resultant characteristic has an inflexion point, which is set at the mid-point of the required temperature operating range. Calculate the required parallel connected resistor if the mid-temperature point of operation is 58.3°C.
[2.2 Ω, 1 W]

(4) What series resistor must be added such that the maximum voltage condition occurs at 5 V across the series plus parallel combination?
[3.45 Ω, 6 W]

25.2. Derive a series of general formulae for the parts of problem 25.1.

$$\left[\alpha = -B/T^2;\quad \hat{T} = \tfrac{1}{2}B\left(1-\sqrt{1-4T_a/B}\right);\quad R_p = R_{T_m}\left(\frac{B-2T_m}{B+2T_m}\right)\right]$$

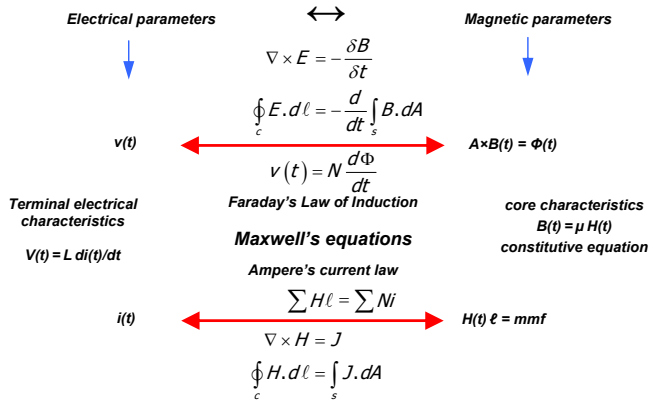
25.3. Resistors coming off the production line are selected according to value and allocated to one of 12 bins for each decade.
Find the nominal centre value for each bin to give a range with equal ratios, and compare with the 'E12' series (1, 1.2, 1.5, 1.8, 2.2, 2.7, 3.3, 3.9, 4.7, 5.6, 6.8, 8.2, *10*).
Find the maximum percentage difference between the nominal value and the E12 value.
Show that the 12-step series corresponds to a ±10 per cent tolerance for practical resistors.
[4.5 per cent between 3.16 and 3.3]

blank

26

Soft Magnetic Materials - Inductors and Transformers

Maxwell's equations form the basis of the relationships between the electrical and magnetic equations that are commonly used in the design of magnetic components such as inductors and transformers. The monogram below succinctly shows those relationships necessary for the design of magnetic components for power electronics applications.



Soft magnetic materials (remove the field and the material returns to the non-magnetised state) are used extensively in power electronic circuits, as voltage and current transformers, saturable reactors, magnetic amplifiers, inductors, and chokes. These magnetic devices may be required to operate at only 50/60 Hz, or at frequencies down to dc or over 1 MHz. For example, a steel laminated ac mains voltage transformer operates at 50/60 Hz, while its ferrite switch-mode power supply counterpart may operate at 500 kHz. Soft magnetic materials have been utilised in other chapters for the following applications:

switching aid circuits	
- linear inductor	(8.3.3)
- saturable inductor	(8.3.4)
- snubber discharge	(figure 8.5)
- unified energy recovery	(9.2.1)
- thyristor *di/dt* control	(figure 8.5)
pulse transformers	(figures 7.7f)
current transformer	(Example 26.7)
turn-on snubber energy recovery	(figures 9.2a)
L-C resonator circuits	(figure 9.5c)
transient current sharing	(figure 10.8)
rfi filtering	(10.4.2)
single and three phase transformers	(11, 12)
cycloconverter intergroup reactors	(13.5)
phase shifting transformers	(15.1.3ii)
current source inductors	(15.2)
smps inductance and transformers	(17, 18)

Hard magnetic material devices, such as those materials used for permanent magnets and ferrite beads for rfi suppression, are considered in chapter 27.

26.1 Inductor and transformer electrical characteristics

Magnetic cores with windings form inductors (a single winding) and coupled circuits (more than one winding), namely transformers, as shown in figure 26.1. Typical *B-H* curve characteristics are shown in figure 26.2 for the different soft magnetic materials shown in table 26.1.

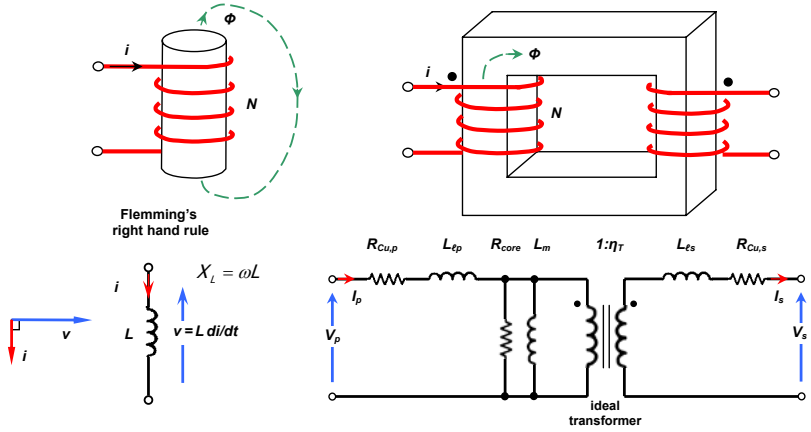


Figure 26.1. *Inductor and transformer models and circuits.*

26.1.1 Inductors

From Faraday's Law:

$$v = N\frac{d\phi}{dt} = NA\frac{dB}{dt} \quad \text{(V)} \tag{26.1}$$

whence for sinusoidal flux

$$v = 4.44NB_sA_ef \quad \text{(V)} \tag{26.2}$$

Inductance (specifically self-inductance) is specified from equation (26.1) and from

$$v = L\frac{di}{dt} \quad \text{(V)} \tag{26.3}$$

$$L = N\frac{d\phi}{di} = N\frac{\Phi}{I} \quad \text{(H)} \tag{26.4}$$

Using $\phi = BA_e$ and $H\ell_e = Ni$, equation (26.4) becomes

$$L = \frac{N^2A_e}{\ell_e}\,dB/dH \quad \text{(H)} \tag{26.5}$$

where *dB/dH* is the slope of the *B-H* curve, according to $B = \mu_o\mu_i H$. Therefore, before core saturation

$$L = \frac{N^2A_e}{\ell_e}\mu_o\mu_i = \frac{N^2}{\Re} \quad \left(=\frac{N\Phi}{I}\right) \quad \text{(H)} \tag{26.6}$$

where $\Re$ is the magnetic circuit reluctance.
The subscript **e** is used to denote the effective core parameter, as shown in table 26.2.
The energy stored in an inductor is given by

$$W = \tfrac{1}{2}LI^2 \quad \text{(J)} \tag{26.7}$$

The energy stored in an air gap volume is

$$W = \tfrac{1}{2}B^2\frac{A\ell_{ag}}{\mu_o} = \tfrac{1}{2}BH\ A\ell_{ag} \quad \text{(J)} \tag{26.8}$$

while the force between the two magnetic surfaces bounding that air gap is

$$F = \tfrac{1}{2}B^2\frac{A}{\mu_o} = \tfrac{1}{2}BH\ A \qquad \text{(N)} \tag{26.9}$$

The effective inductance of uncoupled series, L_s, and parallel, L_p, connected inductors are

$$\begin{aligned} L_s &= L_1 + L_2 + L_3 + \ldots \\ \frac{1}{L_p} &= \frac{1}{L_1} + \frac{1}{L_2} + \frac{1}{L_3} + \ldots \end{aligned} \tag{26.10}$$

26.1.2 Transformers or magnetically coupled circuits

The ideal transformer shown in figure 26.1, with a primary and a secondary winding, in the turns ratio $1{:}\eta_T$, is wound so as to produce the shown voltages according to the usual flux dot convention. Any two of the following electrical equations can be used to derive the third equation.

$$\begin{aligned} P_{in} &= P_{out} \\ I_p \times 1 &= I_s \times \eta_T \\ \frac{V_s}{V_p} &= \frac{\eta_T}{1} \end{aligned} \tag{26.11}$$

Impedance in one winding can be referred (transferred) to the other winding in the turns ratio – squared. Series impedance Z_s in the secondary becomes $Z_p = Z_s / \eta_T^2$ in the primary circuit. Specifically

$$Z_s = R + j\omega L - \frac{1}{j\omega C} \quad \text{becomes} \quad Z_p = \frac{R}{\eta_T^2} + j\omega\frac{L}{\eta_T^2} - \frac{1}{j\omega\eta_T^2 C} \tag{26.12}$$

Note that secondary resistance and inductance referred to the primary are divided by the turns ratio squared, while capacitance is multiplied by η_T^2.
For mutually coupled circuits (transformers), the relationships between the primary and secondary electrical parameters are

$$\begin{aligned} v_p &= L_p\frac{di_p}{dt} \pm M\frac{di_s}{dt} \\ v_s &= L_s\frac{di_s}{dt} \pm M\frac{di_p}{dt} \end{aligned} \tag{26.13}$$

where L_s and L_p are the primary and secondary self inductances given by any of equations (26.4) to (26.6).

M is the mutual inductance, $M = k\sqrt{L_pL_s} = k\ N_s\frac{\Delta\phi}{\Delta i_p}$ for a coupling factor k ($0 \le k \le 1$).

The stored magnetic energy in the core with current in the primary and secondary is

$$\begin{aligned} W &= \tfrac{1}{2}L_pi_p^2 \pm Mi_pi_s + \tfrac{1}{2}L_si_s^2 \\ &= \tfrac{1}{2}\left(\sqrt{L_p}i_p \pm \sqrt{L_s}i_s\right) \end{aligned} \tag{26.14}$$

In equation (26.14), it is required for a transformer, that no energy is stored in the core whence the negative sign is applicable and for energy $W = 0$, $1\times i_p = \eta_T \times i_s$. Faraday's Law, equation (26.2), is applicable to transformers. In the case of a transformer, this equation shows that the advantage of a high core flux density is that more volts v, per turn N, for a given frequency f, results.
When the primary and secondary coupled coils are series connected

$$L_{series} = \left(L_p \pm M\right) + \left(L_s \pm M\right) = L_p + L_s \pm 2M \quad = \left(N_p + N_s\right)^2 / \Re \tag{26.15}$$

When the primary and secondary coils are parallel connected

$$L_{parallel} = \frac{1}{L_p \pm M} + \frac{1}{L_s \pm M} = \frac{L_p \times L_s}{L_p + L_s \pm 2M} \tag{26.16}$$

Note the extra mutual coupling terms, when compared to equation (26.10) for the uncoupled cases.
Figure 26.1 shows how the coupled circuit model of the ideal transformer, is extended to give the usual transformer model, which includes copper winding resistance R_{Cu}, leakage inductance L_ℓ, magnetising inductance L_m, and core losses (eddy current and hysteresis) R_{core}.

26.2 Magnetic material types

Diamagnetic ($\mu_r < 1$) and paramagnetic ($\mu_r \gtrsim 1$) materials are not considered. Two basic types of soft magnetic materials are common, depending on the application requirements. These materials are:

- Ferromagnetic materials based on iron and nickel, which are for lower frequencies, < 2kHz, while
- Ferrimagnetic materials (a subgroup of ferromagnetic materials), which are based on ceramic oxides of metals (ferrites), are applicable to frequencies from a few kilohertz to well over 80 MHz.

26.2.1 Ferromagnetic materials

26.2.1i - Steel

Cold-rolled grain-oriented steel is a 3-4 per cent silicon iron, cold reduced to develop a high degree of grain orientation, which gives

- increased flux for a given magnetising force and
- decreased size for a given rating, hence reduced weight.

Normally cores are produced in a number of material lamination thicknesses

- 0.3 mm for frequencies up to 200 Hz
- 0.1 mm for frequencies between 200 Hz to 2 kHz and
- 0.05 mm for higher frequencies and pulse applications.

Steel laminations for low frequency applications are available in different shapes. E and I laminations or strip C cores or toroids are extensively used for mains transformers and ac line inductors. Non-orientated silicon steels are extensively used for machine laminations.

26.2.1ii - Iron powders

Two general forms of iron powder cores are employed

- Cores are made by highly compacting insulated high quality spongy iron powder.
- High resistivity is required to reduce eddy current losses, so the iron powder is subjected to an acid treatment to produce an insulating oxide layer on the surface of each individual particle. This fine carbonyl iron is mixed with a bonding material and highly compressed. The bonding material used limits the maximum core temperature. Minute gaps appear between the particles, severely reducing the permeability. It is difficult to saturate such materials.

26.2.1iii - Alloy powders

These cores are made by highly compacting insulated alloy powder. The alloy is usually 50-75 per cent nickel, the remainder being iron with a small percentage of copper and molybdenum. The higher the iron percentage, the higher the saturation flux density and the higher the core losses.
Powder iron and alloy cores are available in toroidal or ring shapes, cylindrical and hollow cylindrical cores, as well as cup cores, bobbins, pot cores, and beads.

26.2.1iv - Nanocrystalline

Nanocrystalline soft magnetic alloys are brittle, thin ribbon, 18µm, materials based on iron Fe, silicon Si and boron B with small additions of niobium Nb and copper Cu. They are produced via a rapid solidification technique, being initially in a precursor amorphous (non-crystalline) state and then crystallized into a precise mix of amorphous and nanocrystalline phases when subsequently heat annealed at around 500 to 600°C.
An amorphous magnetic metal has high permeability due to no crystalline magnetic anisotropy. However when applying heat treatment on a typical amorphous metal at temperatures higher than its crystalline temperature, magnetic properties deteriorate for a rapid crystal growth of grains up to 1um. But if the recrystalline grains are restricted to the nano-order, about 10 nm in size, the soft magnetic crystal grains have good magnetic properties. Thus the suppressed grain growth recrystalization during annealing due to the enriching Nb and Cu gives the material it's unique magnetic properties. This extremely fine-grained microstructure with grain sizes of 10 nanometres is termed nanocrystalline.
Nanocrystalline alloys combine low magnetic anisotropy and low magnetostriction, both prerequisites for high magnetic permeability, with high magnetic flux density B_s and good thermal stability. Due to the exchange coupling of randomly oriented grains of Fe, the magnetocrystalline anisotropy averages out to zero. Magnetostriction can also be cancelled by a combination of positive values for the crystalline phase and negative values for the remaining amorphous phase, resulting in zero magnetostriction.

Figure 26.2a shows permeability and saturation magnetic flux densities of representative nanocrystalline soft magnetic materials. Since the compositions of nonmagnetic elements can be reduced in the alloy design, higher saturation magnetic flux density can be obtained in the nanocrystalline soft magnetic materials compared to the existing bulk soft magnetic and amorphous materials.

26.2.2 *Ferrimagnetic materials - soft ferrites*

Ferrites are grey/black, hard, brittle, chemically inert ceramic materials, which have a magnetic cubic (spinel) structure.
The most general ferrites are polycrystalline magnetic isotropic (grains non-aligned) ceramic oxides, which are compounds of iron oxide, Fe_20_3, about 50%, mixed with one or more oxides of bivalent transition metals such as Fe0, Ni0, Zn0, Mn0, Cu0, Ba0, Co0, and Mg0, to give the general compositional form $MeFe_20_4$. At lower frequencies, below a few MHz, a Mn-Zn combination is added to iron oxide, while for higher frequencies, above a MHz, Ni-Zn is the additive.
The raw pure oxide materials are mixed with organic binders, pre-sintered at 1000°C, a process called Calcining and then the partially formed ferrite structure pellets are wet ground by milling, to form a submicron particle slurry with water. After spray drying, the powder material is shaped by means of pressing and sintering at between 1150°C and 1300°C, which cause densification and substantial shrinkage. The sintering process involves raising the temperature to 1300°C in about 3 h, with 15 per cent oxygen present. The cores are cooled slowly without oxygen present to about 200°C in 20 h after entry. In producing the ferrite crystal structure, a 15 per cent linear, and 40 per cent by volume shrinkage occurs during sintering.
A diverse range of ferrite core shapes is available, which include, E, I, U, toroid, drum, pot, rod, tube, and screw. Where appropriate, diamond-wheel-ground air gaps are available on the centre pole. Manufacturing yields limit the physical component in size. Toroid cores of 152 mm outside diameter are not uncommon, and exotic shapes such as motor stators are made for special applications.

26.3 Comparison of material types

Table 26.1 shows typical comparative data for the main classes of soft ferro and ferri magnetic materials. Generally, those materials with higher saturating flux densities, B_s, have higher initial permeability μ_i, and hence offer higher inductance but at the expense of higher core eddy current and hysteresis losses.

Table 26.1: Typical comparative data of soft magnetic materials

Material	thickness	B_s	B_r/B_s	H_c	μ_r	μ_r	P_{cv}	λ_s	T_c
	µm	T	%	A/m	$\times10^3$ 1kHz	$\times10^3$ 100kHz	kW/m^3	10^{-6}	°C
		H_m=800A/m 25°C	H_m=800A/m 25°C	H_m=800A/m 25°C	H_m=0.5A/m 25°C	H_m=0.5A/m 25°C	100kHz B_m=0.2T 25°C		
Nanocrystalline Square Fe-Si	18	1.23	89	0.6	30	5	600	0	570
Nanocrystalline hi-μ_r Fe-Si	18	1.23	5	0.6	50	16	250	0	570
Fe based amorphous	25	1.56	83	2.4	5	5	2200	+27	415
Co-based hi-μ_r amorphous	20	0.55	5	0.3	115	18	280	0	180
Co-based square amorphous	20	0.60	85	0.3	30	10	460	0	210
3% Si steel	50	1.9	85	6.0	2.7	0.8	8400	-0.8	750
6½% Si steel	50	1.3	63	45	1.2	0.8	5800	-0.1	700
50% Ni Permalloy	25	1.5	95	12	-	-	3400	+25	500
80% Ni hi-μ_r Permalloy	25	0.74	55	0.5	50	5	1000	0	480
80% Ni square Permalloy	25	0.74	80	2.4	-	-	1200	0	460
Mn-Zn hi-μ_r ferrite	-	0.44	23	8.0	5.3	5.3	1200	-0.6	150
Mn-Zn low-loss ferrite	-	0.49	29	12	2.4	2.4	680	-0.6	220

In rfi suppression and filtering applications, silicon steel is not effective since the initial permeability, μ_i, falls rapidly with frequency hence at the high suppression frequency, inductance is small. Thus iron powder or a high iron alloy may be used, which have relatively high flux densities and high losses. For rfi suppression, a high core loss aids suppression.
At inaudible frequencies, >20 kHz, for a low core loss, ferrites are extensively used. Although ferrite flux densities are relatively low, typically 0.5 T for power application ferrites, eddy current and hysteresis losses are low. The low eddy current loss results from the high core material resistivity. With ferromagnetic materials, the eddy current loss is reduced by using thinner laminations or electrically isolated powder particles. A major disadvantage of a ferrite core is its poor temperature stability and low allowable core temperature. On the other hand, high initial permeabilities, >20,000, are obtainable.
Ferrite materials, application, and component design are specifically considered, although the concepts developed are generally applicable to ferromagnetic materials.

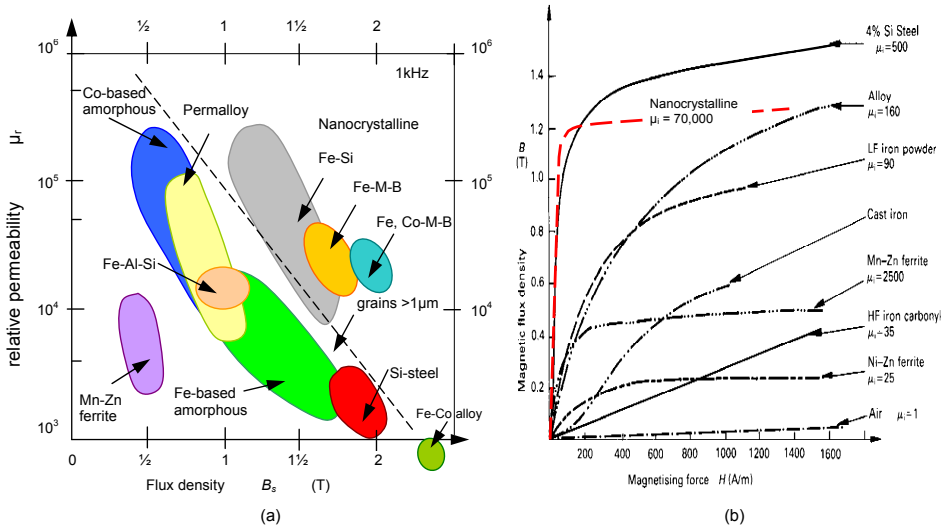


Figure 26.2. *Magnetic alloy characteristics:*
(a) permeability versus saturation magnetic flux densities and (b) B-H characteristics.

26.4 Ferrite characteristics

The definitions and explanations given are applicable to soft magnetic materials in general and are illustrated specifically by reference to ferrite materials.
General mechanical and thermal properties of power ferrites are given in Appendix 26.8, while typical magnetic properties are given in Appendix 26.9.

Table 26.2: Core effective magnetic dimensions and parameters

core factor	symbol	definition	units
form factor, ℓ_e/A_e	c_1	$\Sigma\ell/A$	m^{-1}
effective magnetic area	A_e	$c_1/\Sigma\ell/A^2$	m^2
Effective magnetic length	l_e	$A_e\,c_1$	m
effective magnetic volume	V_e	$\ell_e\,A_e$	m^3
core permeance	c	μ_o/c_1	H

26.4.1 *Dimensions and parameters*

The effective magnetic dimensions are constant for a given core and are defined in table 26.2. These effective constants are based on the length ℓ and area A of the individual limbs comprising the complete

core. These effective dimensions are used for magnetic component design, such as transformer core loss, which is given per unit effective volume, V_e.
From the parameters in table 26.2, inductance is calculated from equation (26.6) as

$$L = \mu_i c N^2 \qquad \text{(H)} \tag{26.17}$$

26.4.2 Permeability

Figure 26.2 shows that a non-linear relationship exists between B and H for magnetic materials, and is characterised by the dimensionless parameter μ_r - the relative permeability - according to $B = \mu_o \mu_r H$ (where $\mu_o = 4\pi\times10^{-7}$ H/m). Figure 26.3 shows a detailed B-H magnetising curve for a ferrite material along with its hysteresis loop. The case of an air core magnetic circuit, for which $\mu_r = 1$, is also shown.
Figure 26.3 illustrates various definitions for μ_r based on the ratio flux density to field strength, namely

$$\mu_r = \frac{1}{\mu_o}\frac{B}{H} \tag{26.18}$$

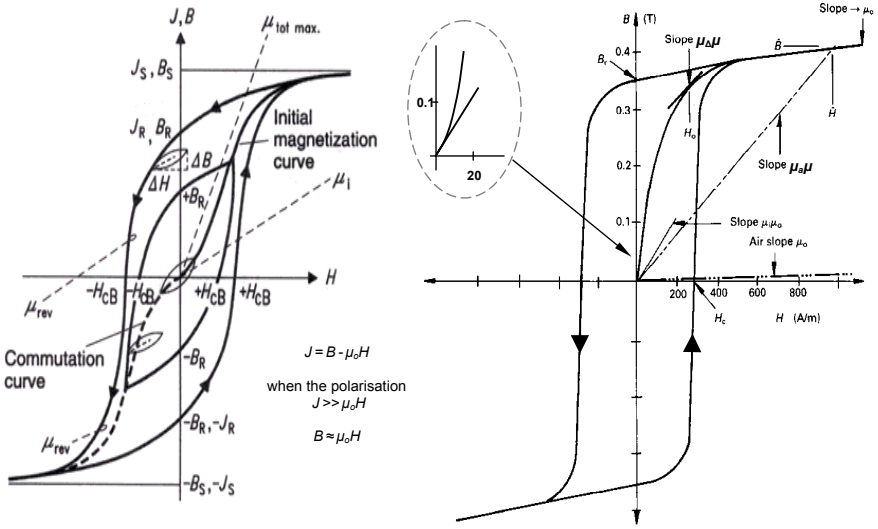


Figure 26.3. *Hysteresis loop illustrating permeability definitions, remanence B_r, and coercive force H_c.*

26.4.2i - *Initial or intrinsic permeability, μ_i*

The initial permeability, which is dependant on temperature and frequency, is the permeability at weak field strengths at $H = 0$ and ΔH tends to zero, that is

$$\mu_i = \left[\frac{1}{\mu_o}\frac{\Delta B}{\Delta H}\right]_{H\,=\,0,\,\Delta H\,\to\,0} \tag{26.19}$$

26.4.2ii - *Amplitude permeability, μ_a and maximum permeability, $\hat{\mu}$*

The amplitude permeability applies to large magnitude sinusoids (high excitation), with no dc field (offset) applied, and is the ratio of the sinusoid peak B and H

$$\mu_a = \left[\frac{1}{\mu_o}\frac{\hat{B}}{\hat{H}}\right]_{H\,=\,0} \tag{26.20}$$

The maximum permeability $\hat{\mu}$ is the maximum μ_a obtainable for any H, that is, $\hat{\mu}$ = max [μ_a] for all values of H. The variation of amplitude permeability with magnetising force or flux density is shown in figure 26.4. Because of the non-linear nature of the B-H curve loop, the amplitude permeability is highly dependant of the applied field strength magnitude. The figure 26.4 is representative of a ferrite material suitable for a wide range of power electronic applications. More technical data for this material is presented in Appendix 26.9 and in the figures that follow.

26.4.2iii - *Reversible or incremental permeability, μ_{rev}, μ_Δ*

When a core is magnetised with a polarising dc offset field upon which a small ac field is superimposed, the ac H field produces a small lancet-shape hysteresis loop which reduces to a straight line as the ac H field is reduced. The slope of this line, shown in figure 26.3, is called the incremental or reversible permeability

$$\mu_\Delta = \frac{1}{\mu_o}\lim_{\Delta H\to 0}\left[\frac{\Delta B}{\Delta H}\right]_{H\,=\,\text{constant}} \tag{26.21}$$

The incremental permeability, μ_Δ is a function of the dc magnetic bias, as shown in figure 26.5. It is usually a maximum when no dc field is present, while for a toroid it is identical with the initial permeability, μ_i. With increased current, μ_Δ, hence inductance, decreases.

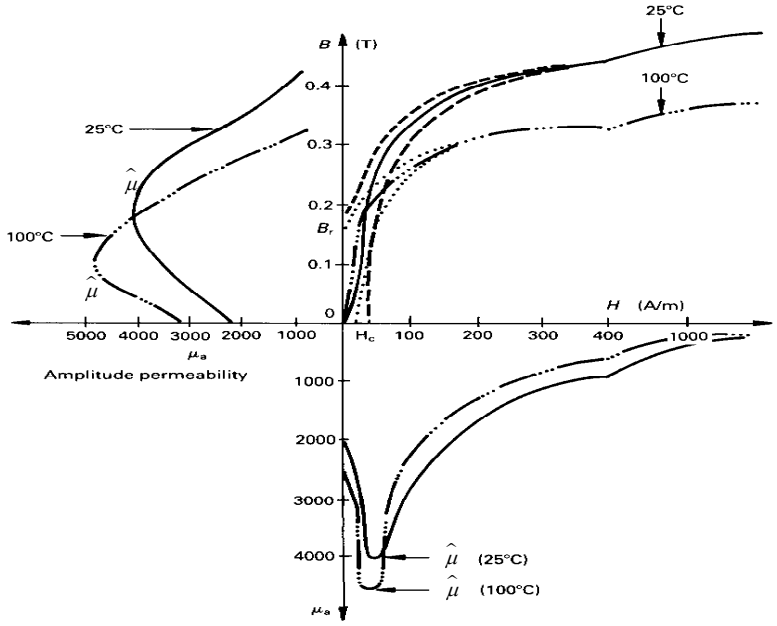


Figure 26.4. *Temperature dependence of flux density B and amplitude permeability, μ_a.*

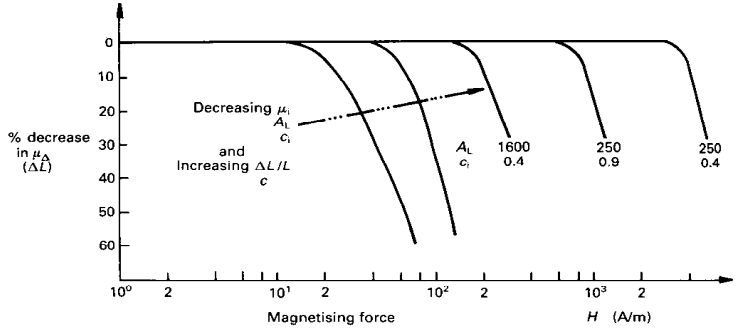


Figure 26.5. *Variation of permeability with field strength.*

26.4.2iv *- Effective permeability, μ_e*

The inductance of a coil with a (air) gapped core of effective (or apparent) permeability μ_e is given by

$$L = \frac{\mu_o \mu_e N^2}{\sum \ell / A} = \mu_e c N^2 = A_L N^2 = \mu_e L_o \qquad \text{(H)} \tag{26.22}$$

hence

$$\mu_e = \frac{L}{cN^2} = \frac{L}{L_o} = \frac{1}{\mu_o}\frac{L}{N^2}\sum \ell / A \tag{26.23}$$

where L_o is the coil inductance if the core is removed (air, μ_r=1), whence the permeability drops.
The term A_L is the inductance factor and is equal to $\mu_e c$. Conversely

$$N = \alpha\sqrt{L} \tag{26.24}$$

where $\alpha = 1/\sqrt{A_L}$ and is termed the turns factor.
If the air gap width, ε, is small compared with the core of effective length, ℓ_e, such that $\varepsilon << \ell_e$, the effective permeability approximates to

$$\frac{1}{\mu_e} = \frac{1}{\mu} + \frac{\varepsilon}{\ell_e} \tag{26.25}$$

The introduction of an air gap is equivalent to connecting two inductors in parallel: one without an air gap, μL_o; the other also without a gap but having an inductance $(\ell_e/\varepsilon)\ L_o$. The effective permeability of a gapped core at low flux levels is specified by the initial permeability, μ_i, and is given by

$$\frac{1}{\mu_e} = \frac{1}{\mu_i} + \frac{\varepsilon}{\ell_e} \tag{26.26}$$

The effective permeability for high flux densities is expressed in terms of the amplitude permeability, μ_a, that is

$$\frac{1}{\mu_e} = \frac{1}{\mu_a} + \frac{\varepsilon}{\ell_e} \tag{26.27}$$

That is

$$\mu_e = \frac{\mu_a}{1 + \frac{\varepsilon \mu_a}{\ell_e}} \tag{26.28}$$

If the magnetic circuit is not homogeneous, has an air gap for example, the effective permeability is the permeability of an equivalent homogeneous non-gapped structure of the same shape, dimensions, and reluctance that would give the inductance equivalent to the gapped structure.
A fringing factor, ε/β, must be introduced for significant gap widths, to account for the effective increase in permeability due to the fringing flux effect. The bulging flux in the gap results in a reduced gap flux density since the effective gap area is increased.

26.4.2v *- Complex permeability, $\bar{\mu}$*

Because of core losses, a coil can be represented by

- a series L_s - R_s circuit for an inductor, figure 26.6
- a parallel R_p // L_p circuit for a transformer (magnetically coupled circuit), figure 26.6.

Core losses are modelled by the inclusion of resistance and the associated losses can be accounted for by considering the coil permeability as a complex variable, $\bar{\mu}$. For the inductor series equivalent circuit

$$Z = R_s + j\omega L_s = j\omega\bar{\mu} c N^2 = j\omega\bar{\mu} L_o = \omega L_o \mu_s'' + j\omega L_o \mu_s' \qquad (\Omega) \tag{26.29}$$

where $L_o = c N^2$ (μ_s'' and μ_s' are the real (inductance) and imaginary (loss) components of $\bar{\mu}$) such that

$$\begin{aligned} \bar{\mu} &= \mu_s' - j\mu_s'' \\ &= \frac{L_s}{cN^2} - j\frac{R_s}{\omega c N^2} \end{aligned} \tag{26.30}$$

while for the transformer parallel equivalent circuit (figure 26.6)

$$\frac{1}{Z} = \frac{1}{R_p} + \frac{1}{j\omega L_p} = \frac{1}{\omega\mu_p'' L_o} - \frac{1}{j\omega\mu_p' L_o} = \frac{1}{j\omega\bar{\mu} L_o} \qquad \text{(S)} \tag{26.31}$$

such that

$$\begin{aligned} \frac{1}{\bar{\mu}} &= \frac{1}{\mu_p'} - \frac{1}{j\mu_p''} \\ &= \frac{1}{L_p / cN^2} - \frac{1}{jR_p / \omega c N^2} \end{aligned} \tag{26.32}$$

Since the parallel and series circuits are equivalent

$$\frac{\mu_s''}{\mu_s'} = \frac{\mu_p'}{\mu_p''} = \tan\delta \tag{26.33}$$

where tan δ is the core loss factor (deviation from the ideal phase angle of 90°)

$$\tan\delta = \frac{R_s}{\omega L_s} = \frac{\omega L_p}{R_p} = \frac{1}{Q} \tag{26.34}$$

Q is a measure of the efficiency of the magnetic component.
The complex permeability components are related according to

$$\begin{aligned} \mu_p' &= \mu_s'\left(1 + \tan^2\delta\right) = \mu_s'\left(1 + 1/Q^2\right) \\ \mu_p'' &= \mu_s''\left(1 + 1/\tan^2\delta\right) = \mu_s''\left(1 + Q^2\right) \end{aligned} \tag{26.35}$$

The parallel and series equivalent components are related by

$$\begin{aligned} R_s &= \frac{R_p}{1+Q^2} = \frac{R_p}{1 + 1/\tan^2\delta} \\ L_s &= \frac{L_p}{1 + 1/Q^2} = \frac{R_p}{1 + \tan^2\delta} \end{aligned} \tag{26.36}$$

The vector diagram for the equivalent parallel and series inductor models are shown in figure 26.6.

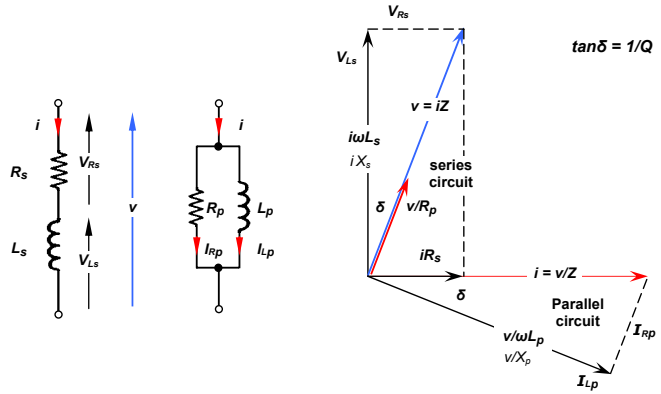


Figure 26.6. *Inductor parallel and series equivalent circuit vector diagrams.*

For low losses, namely at low frequencies, $\tan^2\delta \to 0$ in equation (26.35), whence $\mu_p' = \mu_s'$, while at high losses, at high frequencies, $\mu_p'' = \mu_s''$ since $\tan^2\delta \to \infty$ in equation (26.35). Complex permeability characteristics are shown in figure 26.7. The cut-off frequency, f_c is defined as the frequency at which the permeability is half the initial permeability, μ_s', at low frequency. At 25°C, f_c for Mn-Zn materials is approximated by $f_c \approx 4000/\mu_i$(MHz), for μ_i at low frequency.
The complex permeability components are measured at low flux densities. Mn-Zn ferrites applicable to power application usually have high permeability, low resistivity, and a high dielectric constant. In such cases, the complex permeability is highly dependent on the core dimensions, as shown in figure 26.7, which characterises stacked toroids. Because of the associated large volume, volume resonance occurs where eddy currents dominate losses.

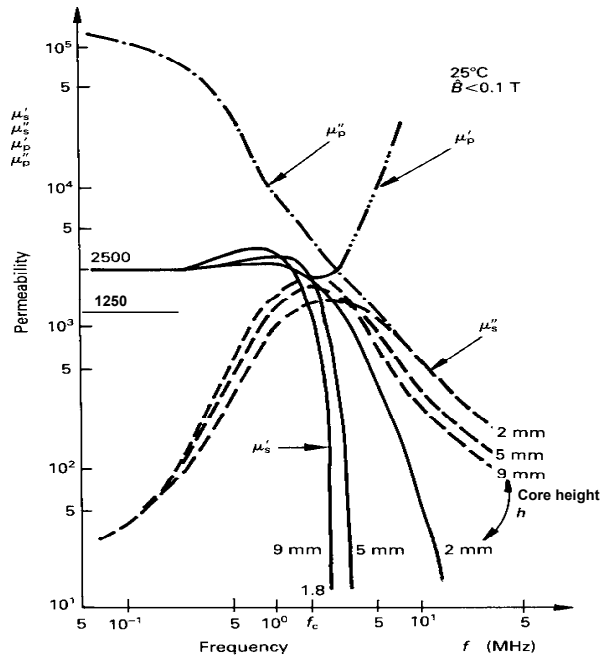


Figure 26.7. *Influence of core height (h) on the frequency characteristics of the complex permeability for a toroid.*

26.4.3 Coercive force and remanence

The coercive force H_c is the field strength at which the hysteresis loop cuts the *H*-axis as shown in figures 26.3 and 26.4. It is the reverse magnetic field needed to reduce a magnetically saturated structure from remanence to zero magnetic induction. It is representative of the static hysteresis loss of the material. The point where the hysteresis loop intersects the *B*-axis is called the remanence (flux density), B_r. Where a core is operated with a magnetic field strength bias, for example, as with an inductor carrying dc current, the value of flux density is reduced to B_s - B_r for calculations. The area within the hysteresis loop represents core hysteresis loss, in Joules per unit volume.

26.4.4 Core losses

26.4.4i *- Core losses at low H*

At low magnetising forces, the total losses, represented by R_t, can be separated into three core components (magnetic components, R_m) and a copper turns component, R_{Cu}. The components are

- frequency dependent eddy currents, R_F
- frequency dependent hysteresis, R_h
- magnetic drag, remanence loss, or residual loss, R_r
- copper winding loss including both dc and ac components, R_{Cu}, where

$$R_t = R_m + R_{Cu}$$
$$R_t = R_F + R_h + R_r + R_{Cu} \quad (\Omega) \tag{26.37}$$

The coil is represented by the series R_t - L_s circuit where L_s is the lossless self-inductance. Empirical formulae, called *Jordan formulae* can be used to calculate R_t at low magnetic forces.

The series coil model impedance is given by (equation (26.29))

$$Z = R_t + j\omega L_s \quad (\Omega) \tag{26.38}$$

whence

$$\tan\delta_t = \frac{R_m + R_{Cu}}{\omega L_s} = \frac{R_t}{\omega L_s} = \frac{R_F}{\omega L_s} + \frac{R_h}{\omega L_s} + \frac{R_r}{\omega L_s} + \frac{R_{Cu}}{\omega L_s} = \tan\delta_F + \tan\delta_h + \tan\delta_r + \tan\delta_{Cu} \tag{26.39}$$

where tan δ_t is the loss factor for the coil. The reciprocal of the loss factor is the inductor quality factor, namely

$$Q = \frac{1}{\tan\delta_t} = \frac{\omega L_s}{R_t} \tag{26.40}$$

The copper loss (R_{Cu}) is usually excluded so as to characterise the core material specifically, whence

$$Q = \frac{1}{\tan\delta} = \frac{\mu_s''}{\mu_s'} = \frac{\mu_p'}{\mu_p''} \tag{26.41}$$

An alternative relative core loss factor is tanδ /μ_i or 1/ $\mu_i Q$. This particular figure of merit factor is generally characterised only for high frequency Ni-Zn ferrites. The relative loss factor for a gapped core, tan δ_e, can be found by multiplying the core loss factor by the gapped core effective permeability, μ_e, that is

$$\tan\delta_e = \frac{\mu_e}{\mu}\tan\delta \tag{26.42}$$

The term μ_e/μ_i is extensively used in equations to account for an air gap introduced into the core length. In transformers, the hysteresis component R_h increases as the hysteresis loop opens up at higher flux densities, and the hysteresis loss resistance R_h, in terms of the hysteresis loss factor tanδ_h, is

$$R_h = \omega L \times \tan\delta_h$$

The hysteresis material constant η_B, which characterises hysteresis losses of a specific material, is

$$\eta_B = \frac{\tan\delta_h}{\mu_e \times \Delta\hat{B}}$$

26.4.4ii *- Core losses at high H*

1 - Ferrites

Core losses, P_V, with high flux densities in Mn-Zn ferrites are applicable to power electronic application. Empirical formulae are not practical, and ferrites used for choke and transformer cores are provided with experimentally characterised total core loss per unit volume data, as indicated in figure 26.8. This loss, for a power Mn-Zn ferrite, is given as a function of frequency, temperature, and flux density. The general loss term for hysteresis and eddy current losses is of the form

$$P_v = k f^a B^b \quad (\text{W/m}^3) \tag{26.43}$$

where $1.2 < a < 1.6$ for hysteresis loss and $1.9 < a < 2.2$ for eddy current loss
$2.1 < b < 2.6$ for hysteresis loss and $1.8 < b < 2.3$ for eddy current loss
k is a function of temperature

For a specified and limited operating range, core losses in figure 26.8 can be approximated by

$$P_V(25°\text{C}) = P_h + P_F = 5.8\times10^{-5}\times f^{1.2}\times\hat{B}^{2.11} + 3.32\times10^{-7}\times f^2\times\hat{B}^2 \quad (\text{mW/cm}^3) \tag{26.44}$$

where f is in kHz for 10 kHz ≤ f ≤ 500 kHz
and $\hat{B}$ is the peak flux density in mT for 50mT ≤ $\hat{B}$ ≤ 250mT.

Temperature dependence is modelled according to

$$P_V = \kappa \times P_V(25°\text{C}) \tag{26.45}$$

where

$$\kappa = 1.48\times10^{-4}\times T^2 - 21.2\times10^{-3}\times T + 1.44$$

for f < 200 kHz and $\hat{B}$ ≥ 100 mT

$$\kappa = 1.2\times10^{-4}\times T^2 - 17.8\times10^{-3}\times T + 1.38$$

for f ≥ 200 kHz and $\hat{B}$ ≤ 100 mT

The temperature T is with respect of 0°C.
The per unit volume loss P_V *(T)* is applicable to a square wave. For a half wave sine, power losses are reduced by 0.7 - 0.8 while for a full wave rectified sine wave, losses are increased by 1.8 to 2.2.

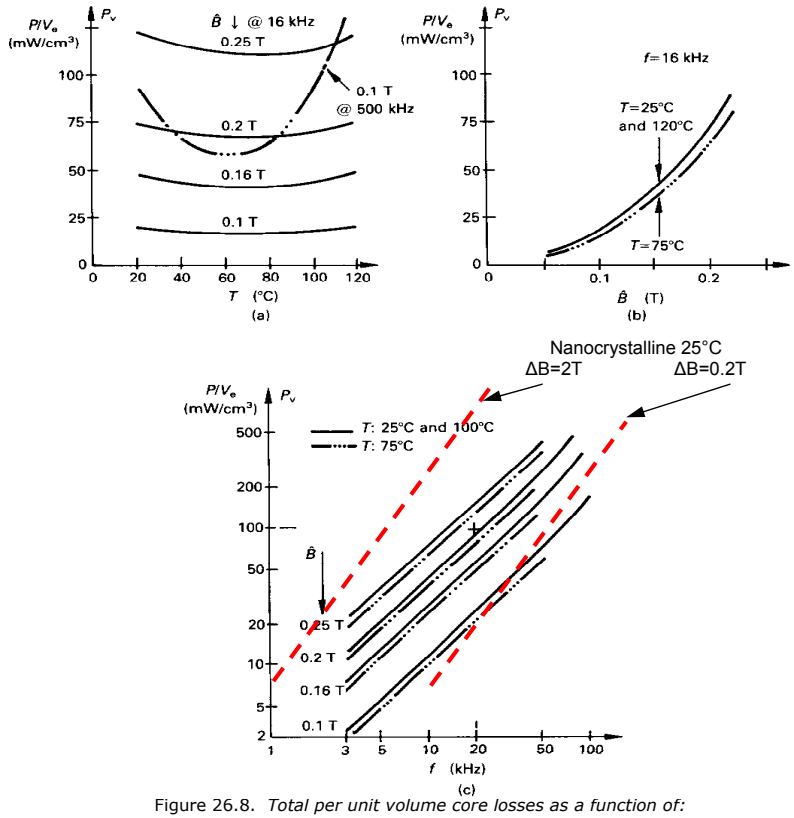


Figure 26.8. *Total per unit volume core losses as a function of: (a) core temperature, T; (b) maximum flux density, $\hat{B}$; and (c) frequency, f.*

2 - Nanocrystalline alloys

The magnetic reversal loss characteristics of nanocrystalline material is similar to that of ferrite, but with lower losses and higher allowable flux swings. Based on equation (26.43), core losses are approximated by

$$P_v = 3.09 \times \Delta B^{1.5} \times f^{1.5} \quad \left(\text{mW/cm}^3\right) \tag{26.46}$$

Since the density is 7.35 g/cc, that is 1W/kg=7.35mW/cm^3

$$P_v = 0.42 \times \Delta B^{1.5} \times f^{1.5} \quad \left(\text{W/kg}\right) \tag{26.47}$$

These empirical formula, shown in figure 26.8c and 28.23a, are valid for flux swings of up to $\Delta B = 2$T and $f \leq 200$kHz, at 25°C.

3 - Laminated silicon steel

Hysteresis and eddy current losses for silicon steel can be calculated by using well established, classical empirical formulae.

(a) Hysteresis loss

Steinmetz equation predicts hysteresis loss according to

$$P_h = \lambda_h \hat{B}^n f V_e \quad \text{(W)} \tag{26.48}$$

where λ_h and n are characteristics of the core material:

$n = 1.7$

$\lambda_h = 500$ for 4 per cent silicon steel

$= 3000$ for cast iron

(b) Eddy current loss

Eddy current loss is predicted by

$$P_e = \frac{\left(\pi \hat{B} f t\right)^2}{6\rho} V_e \quad \text{(W)} \tag{26.49}$$

where t is the thickness of the lamination, parallel to the flux path, and ρ is the magnetic material resistivity. This formula illustrates why high resistivity ferrites have low eddy current loss, even at high frequencies. In the case of iron, the addition of 3-4 per cent silicon increases the resistivity by about four times, reducing both eddy current and hysteresis losses.

Eddy currents produce magnetic fields in the magnetic material, by Lenz's law, which will oppose the applied field. This reduces the flux density in the core centre such that most of the flux is confined to a thin layer or skin near the surface, termed *skin effect*.

Within a magnetic material with an ac flux, the flux density distribution is given by

$$B(x) = B(0) e^{\frac{-x}{\delta}} \quad \text{(T)} \tag{26.50}$$

where x is the distance from the surface

$\delta = \sqrt{(\rho/\mu_o \pi f)}$ is called the skin depth.

Laminations should be less than ½δ thick. The skin effect in metals can be used to absorb radiated and conducted rfi by using laminations > 2δ thick.

A similar effect occurs within conductors carrying ac current, where the current is minimal at the conductor centre. The current density, J, is given by

$$J(x) = J(0) e^{\frac{-x}{\delta}} \quad \left(\text{A/m}^3\right) \tag{26.51}$$

Below 20-50 kHz and above a few megahertz, solid wire is preferred. In between these frequencies, individually insulated stranded wire, *Litz wire* (after *Litzendraht*) is preferred; decreasing from 0.07 mm to 0.03 mm in strand diameter as the frequency increases and interwinding capacitance dominates. Copper foil can also be employed.

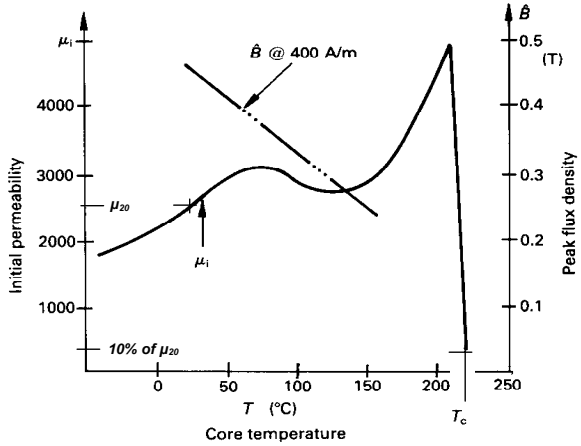


Figure 26.9. *Permeability, μ_i, and maximum density, $\hat{B}$, as a function of core temperature, T.*

26.4.5 Temperature effects on core characteristics

Generally ferrites have poor characteristic temperature stability. At higher temperatures, at the *Curie point*, core materials lose their ferromagnetic magnetic properties abruptly and become paramagnetic ($\mu_e \approx 1$). The temperature causes disruption of the magnet ordering in the crystalline lattice due to molecular thermal motion. The phenomenon is reversible and below the Curie temperature, T_c, the material becomes magnetic again. Typical magnetic material Curie temperatures are:

Fe	770°C
Co	1130°C
Ni	358°C
$Nd_2Fe_{14}B$ (N54)	120°C
Ferrite Mn-Zn	180°C
Nanocrystalline Fe-Si	570°C

The temperature effect on initial permeability in figure 26.9 illustrates the sudden loss of permeability at 212°C, whence the permeability falls to 1, to that of air. The Curie temperature is usually defined as that temperature where the initial permeability falls to 10% of that permeability at 20°C. Generally, Curie temperature is inversely proportional to the initial permeability, μ_i. For most ferrites the initial permeability increases with temperature, and reaches a maximum just below the Curie temperature, as shown in figure 26.9.
Other ferrite parameters are also affected by temperature. Increased temperature decreases flux density and hysteresis loss as shown in figures 26.4 and 26.9. The effects of temperature on total core loss per unit volume are shown in figure 26.8a.

26.4.6 *Inductance stability*

Three factors affect inductance core stability:

- *Parameter effects*
- *Time effects*
- *Temperature effects*

26.4.6i *- Parameter effects*
From the differential of equation (26.22)

$$\frac{dL}{L} = \frac{d\mu_e}{\mu_e} \tag{26.52}$$

while differentiating equation (26.26) yields

$$\frac{d\mu_e}{\mu_e^2} = \frac{d\mu_i}{\mu_i^2} \tag{26.53}$$

Substituting equation (26.53) into equation (26.52) gives

$$\frac{dL}{L} = \frac{L_1 - L_2}{L_1} = \frac{d\mu_i}{\mu_i^2}\frac{A_L}{c} \tag{26.54}$$

The factor $d\mu_i / \mu_i^2$ is constant for a given temperature, hence any change in inductance is due to variations in A_L and c. Thus in order to increase the stability of an inductor in a given material with $\varepsilon << \ell_e$, it is necessary to increase the magnetic circuit air gap (to reduce the inductance factor A_L) or to select a bigger core (to increase the core permeance factor c).

26.4.6ii *- Time effects*
Initial permeability of a ferrite decreases with time under constant operating conditions, including constant temperature. Alternatively, after a ferrite is subjected to shock (thermal, mechanical or magnetic) the permeability increases abruptly, then gradually drifts down over a long period. A *disaccommodation factor*, *df*, independent of effective permeability, is introduced, which characterises the material such that the change in inductance is defined by

$$\frac{dL}{L} = \frac{L_1 - L_2}{L_1} = df\ \mu_e \log_{10}\frac{t_2}{t_1} \quad \text{for} \quad t_2 > t_1 \tag{26.55}$$

The disaccommodation factor is defined by

$$df = \frac{\mu_2 - \mu_1}{\mu_2^2} \tag{26.56}$$

with units of ppm ($/10^{-6}$).
This expression is based on the fact that permeability is proportional to the logarithm of time. The *df* increases slightly with temperature. Generally the *df* decreases, as shown in table 26.3:

- as the initial permeability increases for a given resistivity
- as resistivity decreases.

The effective disaccommodation factor $df_e = df \times \mu_e$ is the actual disaccommodation of a magnetic circuit where material permeability has been reduced to μ_e by gapping.

Table 26.3: Factors affecting the disaccommodation factor for ferrites

		ρ(Ω cm)		
		10^5	500	≈ 20
μ_i		11-250	800-2000	4000
T_c	°C	450-300	250-170	145
df	$\times 10^{-6}$	50-10	20-2	3

Example 26.1: ***Inductance variation with time***

A pot ferrite core with an effective permeability of 100 ($A_L = 250$) and a disaccommodation factor $df < 35 \times 10^{-6}$ has been in satisfactory operation for five weeks after production. What is the expected inductance variation after 10 years?

Solution

From equation (26.55)

$$\frac{dL}{L} = df\ \mu_e \log_{10}\frac{t_2}{t_1} < 35\times10^{-6}\times100\times\log\frac{520\text{ weeks}}{5\text{ weeks}}$$

that is, $dL < 0.7$ per cent can be expected.

♣

26.3.6iii *- Temperature effects*
Figure 26.9 shows that between +5°C and +55°C the permeability μ_i variation as a function of temperature is approximately linear for this ferrite. The temperature coefficient of permeability α (or *TC*) is given by

$$TC = \alpha = \frac{1}{\mu_{i1}}\frac{\mu_{i2} - \mu_{i1}}{T_2 - T_1} = \frac{1}{\mu_{i1}}\frac{\Delta\mu_i}{\Delta T} \qquad (\text{K}^{-1}) \tag{26.57}$$

where $\Delta\mu_i = \mu_{i2} - \mu_{i1}$ is the initial permeability variation over the temperature range $\Delta T = T_2 - T_1$.
The relative temperature factor *tf* is defined in terms of the temperature coefficient and intrinsic permeability by

$$\alpha_F = \text{temperature factor} = tf = \frac{\alpha}{\mu_{i1}} = \frac{1}{\mu_{i1}^2}\frac{\Delta\mu_i}{\Delta T} \qquad (\text{K}^{-1}) \tag{26.58}$$

If the permeability variation is large, the temperature factor is modified to

$$\alpha_F = \text{temperature factor} = tf = \frac{1}{\mu_{i1}\mu_{i2}}\frac{\Delta\mu_i}{\Delta T} = \frac{\text{relative temperature coefficient}}{\Delta T} \qquad (\text{K}^{-1}) \tag{26.59}$$

In a magnetic circuit with an air gap and effective permeability, μ_e, the actual effective temperature coefficient of the core is reduced with gapping according to (see equation (26.42))

$$\alpha_e = \alpha\frac{\mu_e}{\mu_i} = \alpha_F\mu_e = tf \times \mu_e \qquad (\text{K}^{-1}) \tag{26.60}$$

The term $\alpha_F = \alpha_i / \mu_i$ is called the relative temperature coefficient (per unit of permeability). The relative inductance change between two temperatures can be determined by

$$\frac{dL}{L} = \frac{L_2 - L_1}{L_1} = \alpha_F\mu_e\Delta T = \frac{\mu_{i1} - \mu_{i2}}{\mu_{i1}\mu_{i2}} \times \mu_e \tag{26.61}$$

For effective permeability $\mu_e < 80$, (that is, a less dominant air gap), the temperature coefficient $\alpha_e = \mu_e\alpha_F$ should be increased by 10 to 30 × 10^6/K to account for the temperature influence of the winding.

Example 26.2: ***Temperature effect on inductance***

The gapped pot core in example 26.1 is specified by a relative temperature coefficient of 1 x 10^6/K. What is the expected inductance variation over the temperature range 25 to 55°C?

Solution

From example 26.1

$$\mu_e = \frac{A_L}{c} = \frac{250}{2.5} = 100$$

$$\Delta T = 55 - 25 = 30°\text{C}$$

From equation (26.61)

$$\frac{dL}{L} = \alpha_r \mu_e \Delta T$$
$$= 1\times10^{-6}\times100\times30 = 0.3 \text{ per cent inductance variation}$$

♣

Whilst permeability decreasing with aging for ferrite, and significantly so for Cobalt based amorphous metals, nanocrystalline and non-alloyed metals are comparatively stable.

26.4.7 Stored energy in inductors

The energy stored in the magnetic field in the core (before saturation) is given by

$$E = \tfrac{1}{2}BH\breve{V}_e \quad \text{(J)} \tag{26.62}$$

where $\breve{V}_e$ is the effective minimum core volume. It can be shown that before saturation, the stored magnetic energy is equivalent to the stored electrical energy, whence

$$E = \tfrac{1}{2}BH\breve{V}_e = \tfrac{1}{2}Li^2 \quad \text{(J)} \tag{26.63}$$

For un-gapped cores, like a toroid, the ferrite effective volume, V_e, is equal to the minimum effective volume, $\breve{V}_e$. Inductors meeting this requirement may necessitate an excessive core size. However the introduction of an air gap, ε, can reduce the core size significantly, since a significant amount of the energy can be stored in the gap volume. The minimum effective volume is now larger than the ferrite core effective volume, and is given by

$$\begin{aligned}\breve{V}_e &= V_e + A_e\mu_i\varepsilon \\ &= A_e\left(l_e + \mu_i\varepsilon\right) \quad (\text{m}^3)\end{aligned} \tag{26.64}$$

Figure 26.10 shows modified *B-H* characteristics (no hysteresis) for air gap inductors. The line curve o-b represents the core without an air gap, which results in the largest inductance. The energy stored in the core for a flux BA_e, in the linear portion of the curve, is the area of the shaded triangle 0-a-b, as defined by equation (26.62). When an air gap is introduced, the effective permeability falls as shown by the decreased slope of line o-c. The figure shows that as the air gap increases, the inductance decreases. It can be shown that the stored energy in the air gap and core is represented by the shaded area o-a-c, for a given flux, BA_e. It can be seen that the energy stored in the gap of length ε, o-b-c, although its length is much shorter than the core length, ℓ_c, is much greater than that stored in the core, 0-a-b.
The total energy stored o-a-c, E_T, in the magnetic circuit comprises the energy stored in the air gap, o-b-c, E_ε, plus the energy stored in the magnetic core material, o-a-b, E_{core}. These two energies are equal to the areas of the shaded triangles, o-b-c and o-a-b, respectively in figure 26.10. That is

$$\begin{aligned}E_T &= E_{core} + E_\varepsilon \\ &= \tfrac{1}{2}B_cH_cV_c + \tfrac{1}{2}B_\varepsilon H_\varepsilon V_\varepsilon\end{aligned} \tag{26.65}$$

If leakage is neglected, then the air gap flux is the same as the core flux. If fringing is neglected then the area of the core at the air gap is the same as the area of the gap. Then

$$\begin{aligned}E_T &= \tfrac{1}{2}BH_cA_e\ell_c + \tfrac{1}{2}BH_\varepsilon A_e\varepsilon \\ &= \tfrac{1}{2}BA_e\left(H_c\ell_c + H_\varepsilon\varepsilon\right)\end{aligned} \tag{26.66}$$

For a gapped core, as shown in figure 26.13, from Ampere's current law (horizontal axis in figure 26.10)

$$Ni = H_c\ell_c + H_\varepsilon\varepsilon \tag{26.67}$$

Therefore, substituting equation (26.67) into equation (26.66) gives the total stored energy as

$$\begin{aligned}E_T &= \tfrac{1}{2}BA_e\left(H_c\ell_c + H_\varepsilon\varepsilon\right) \\ &= \tfrac{1}{2}BA_e\left(Ni\right)\end{aligned} \tag{26.68}$$

which is equal to the area of the shaded triangle o-a-c.

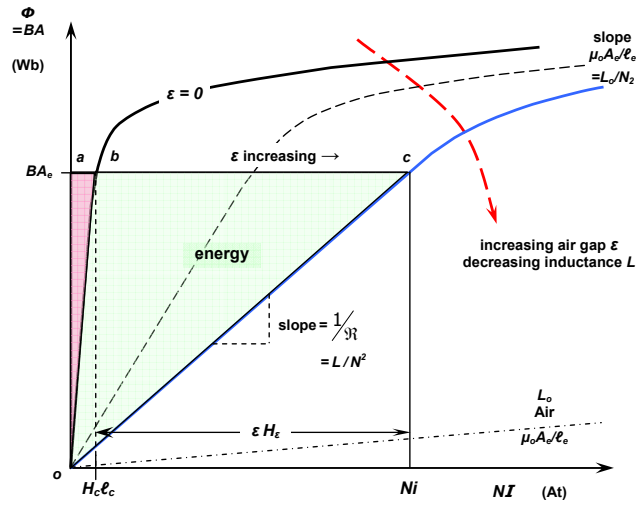


Figure 26.10. *Effects of an air gap on inductance and stored energy.*

The inductance *L* is given by equation (26.4), that is

$$L = \frac{N\phi}{i} \tag{26.69}$$

Substitution of equation (26.67) for the current *i* gives

$$\begin{aligned}L &= \frac{N^2\phi}{H_c\ell_c + H_\varepsilon\varepsilon} \\ &= \frac{N^2A_e\mu_o}{\dfrac{\ell_c}{\mu_r}+\varepsilon} = \frac{N^2A_e\mu_o}{\ell_c+\varepsilon}\left[\frac{\left(\ell_c+\varepsilon\right)\mu_r}{\ell_c+\mu_r\varepsilon}\right] = \frac{N^2A_e\mu_o\mu_e}{\ell_{total}}\end{aligned} \tag{26.70}$$

where $\ell_{total} = \ell_c + \varepsilon$ and

$$\mu_e = \frac{\left(\ell_c+\varepsilon\right)\mu_r}{\ell_c+\mu_r\varepsilon} \tag{26.71}$$

Making the usual assumption that the length of the core is much greater than the length of the air gap, $\ell_c >> \varepsilon$, yields equation (26.28) for the effective permeability.

26.5 Ferrite inductor and choke design, when carrying dc current

Air gaps in magnetic circuits are introduced in order to reduce the influence of a superimposed dc current, manufacturing dispersion and/or to improve parameter stability. Saturable inductors for a semiconductor switch turn-on snubber normally do not employ an air gap, in order to reduce the stored energy, which may be subsequently dissipated, and to minimise the magnetising current magnitude.
Empirical equations have been derived for cylindrical inductors with a cylindrical core, which give an inductor with a large air gap. Design equations and examples are given in Appendix 26.11.

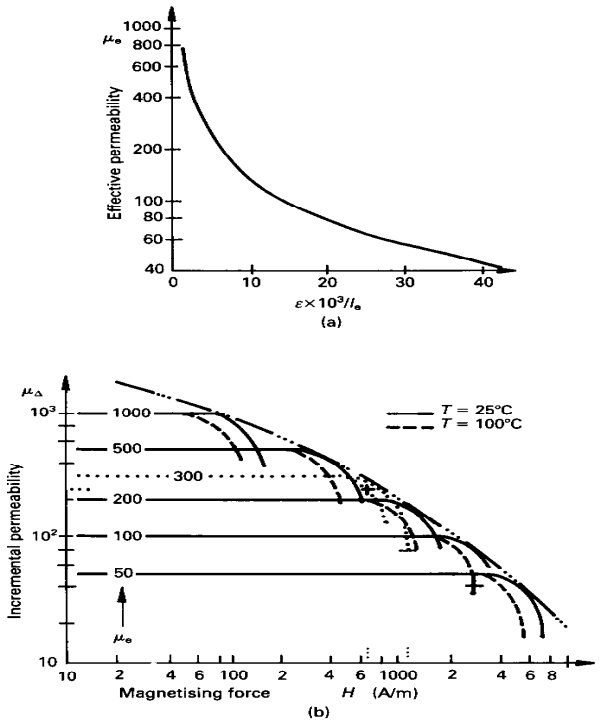


Figure 26.11. *Permeability as a function of: (a) air gap, ε and (b) superimposed dc field and air gap.*

26.5.1 Linear inductors and chokes

The introduction of an air gap reduces the effective permeability, μ_e such that the coil inductance is given by the equation (26.22)

$$L = \mu_e c N^2 = \mu_e L_o = A_l N^2 \qquad \text{(H)} \tag{26.72}$$

Figure 26.11a shows the variation of the effective permeability, μ_e at both low flux levels and without a dc bias, as a function of the relative air gap, ε/ℓ_e as specified by equation (26.27). As the air gap and the superimposed dc field are varied, the incremental permeability, μ_Δ varies as shown in figure 26.11b. This figure indicates how inductance varies with dc bias current (H).
Figure 26.11 does not specify the optimum inductor design since for a given inductance and dc current the optimum air gap and number of turns are not specified. The minimum number of turns and air gap requirements, for a dc current, can be determined by means of the Hanna curves in figure 26.12.

This figure shows experimental families of curves of per unit core energy against magnetomotive force per unit length, for different air gap widths. The resultant curves are ferrite type dependent and dimensionally independent. Hanna curves therefore allow the determination of minimum turns N and air gap ε, from the required inductance L and dc current I.

Three distinct energy levels are shown in the Hanna curves in figure 26.12.

i. At low dc currents (H) the per unit energy increases linearly with H. This region corresponds to the horizontal regions in figure 26.11b, where

$$L = \mu_\Delta c N^2 \qquad \text{(H)} \tag{26.73}$$

and as H varies, μ_Δ is constant.

ii In the mid energy region, the per unit energy can decrease with increased H. The incremental permeability decreases, causing L to decrease at a greater rate than the increase in the dc current squared, I^2. This region is characterised by the fall off in μ_Δ, hence inductance, as H increases as shown in figure 26.11b.

iii. At high dc currents, the core material saturates, and μ_Δ tends to unity. Air core inductance results, where

$$L = L_o = c N^2 \tag{26.74}$$

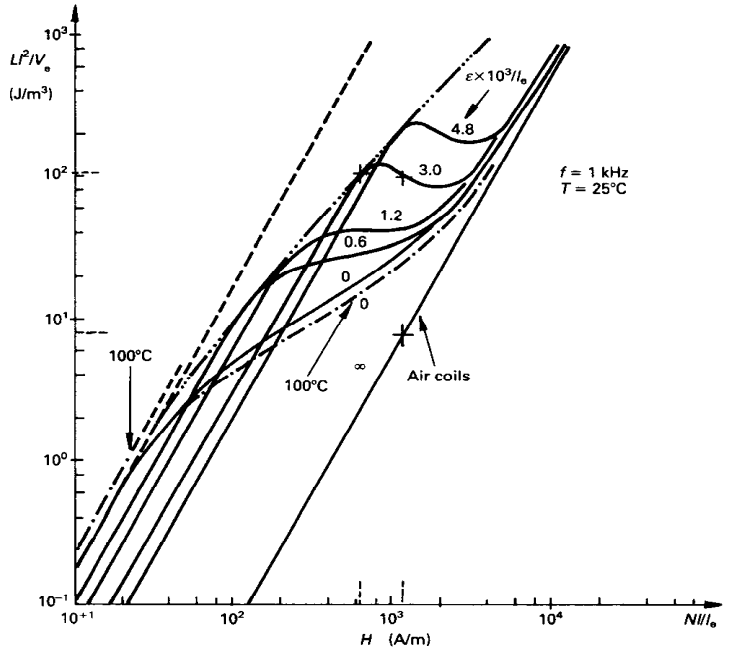


Figure 26.12. *Hanna curves, showing trajectories for different air gaps.*

Example 26.3: ***Inductor design (ferrite) with Hanna curves***

A 20 μH, 10 A choke is required for a forward converter. The inductance must be constant for unidirectional currents to 10 A. An available E-core pair has the following effective parameters

ℓ_e = 0.11 m, A_e = 175 × 10^{-6} m^2, V_e = 19.3 × 10^{-6} m^3 and
μ_i = 2500 @ 25°C and 3000 @ 100°C (from figure 26.9)

i. At a core temperature of 25°C, determine the required air gap and turns. Allow a 5 per cent decrease in inductance at rated conditions.
ii. Estimate the inductance at 20A dc.
iii. Calculate the inductance at 10A and 20A dc, both at 100°C.

Solution

i. Evaluate $\dfrac{LI^2}{V_e} = \dfrac{20\times10^{-6}\times10^2}{19.3\times10^{-6}} = 104 \text{ J/m}^3$

From figure 26.12, restricted to the constant-L region, 104 J/m^3 corresponds to
(a) $\varepsilon/\ell_e = 3 \times 10^{-3}$
whence $\varepsilon = 3 \times 10^{-3} \times \ell_e = 3 \times 10^{-3} \times 0.11$
The required total air gap is 0.33 mm
(b) H = 650A/m
Since $H = NI/\ell_e$
$N = H\ell_e/I$ = 650 × 0.11/10 = 7.15 turns
Use 7 turns and a 0.33 mm total air gap.

ii. At 20 A, 25°C
$H = NI/\ell_e$ = 7 × 20/0.11 = 1270A/m

Two alternative design approaches may be used to estimate the inductance.

(a) The effective permeability, μ_e, before saturation can be evaluated from equation (26.27)

$$\frac{1}{\mu_e} = \frac{1}{\mu_a} + \frac{\varepsilon}{l_e} = \frac{1}{2500} + 3\times10^{-3}$$

that is

$$\mu_e \approx 300$$

From figure 26.11b, for μ_e = 300 it can be seen that the incremental permeability μ_Δ is constant, as required to 650A/m, then μ_Δ decreases as saturation commences. At H = 1270A/m, μ_Δ has fallen to 75, from 300. The incremental inductance at 20 A is about ¼ of 20μH, namely 5.0μH.
(b) Alternatively, a simpler approach uses only figure 26.12. H = 1270A/m projects 100 J/m^3. Solving 100.0 = $L_{20A} I^2 / V_e$ with I = 20A yields L_{20A} = 5μH.

iii. The effective permeability at 100°C is

$$\frac{1}{\mu_e} = \frac{1}{\mu_i} + \frac{\varepsilon}{l_e} = \frac{1}{3000} + 3\times10^{-3}$$

that is

$$\mu_e \approx 300$$

It is seen that, although the initial permeability varies significantly with temperature, here the effective permeability is dominated by the air gap, hence is essentially temperature independent. Figure 26.11b, with H = 640A/m, projects μ_Δ = 220 at 100°C. Using $L \propto \mu_\Delta$, the inductance falls to about 15μH at 100°C, 10 A.
At 20A, 100°C, the effects of saturation are highly significant, and figure 26.11b indicates that the incremental permeability is low. The best approximation is to use the air coil curve in figure 26.12. Hence H = 1270A/m projects 9J/m^3. From 9 = $L_{20A} I^2 / V_e$, at 20A, 100°C, an inductance of at least 0.43μH can be expected.

♣

Figure 26.5 shows how μ_Δ and hence the inductance, falls off as H, and hence the current, increases for ferrite core materials. The larger the air gap, and hence the lower A_L and the lower L, the higher H before inductance rolls off. Inductance rolls off faster, the wider the air gap, hence the higher the magnetic field strength, H. The decrease in effective permeability, μ_e and inductance factor, A_L, with increase of air gap, ε, is shown in figure 26.13 for two E-cores.
Figure 26.14 shows typical curves for the decrease in μ_Δ, hence inductance, with increased H, hence current, for both ferrites and alloy or iron powder cores. Because power ferrites have a squarer *B-H* curve than powder cores, the inductance of ferrites falls off faster. By increasing the core volume, the fall off rate of inductance can be reduced. Depending on core loss for a given volume, a powder core may be more effective than a ferrite; and would have better utilisation of the copper window area. The design approach previously considered in example 26.3 in fact neglects the optimisation of core size and copper I^2R loss.

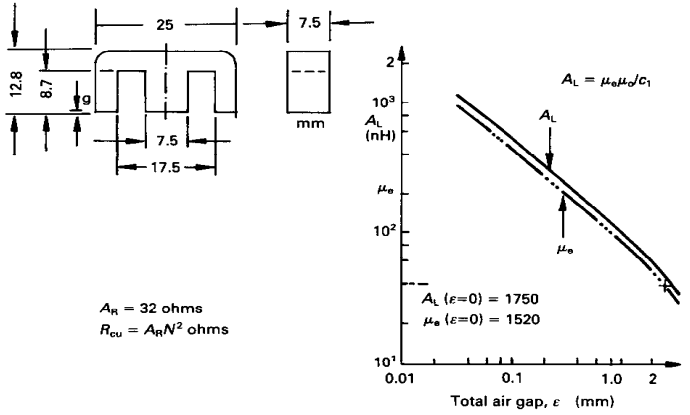


Figure 26.13. *Characteristics of a pair of gapped E-cores. Core dimensional parameters are given in table 26.5.*

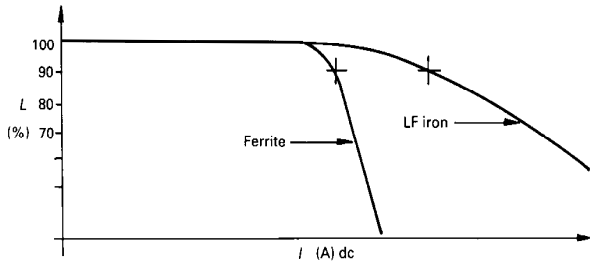


Figure 26.14. *Comparison of inductance characteristics illustrating how inductance falls off faster with ferrite cores than with iron cores, at higher currents.*

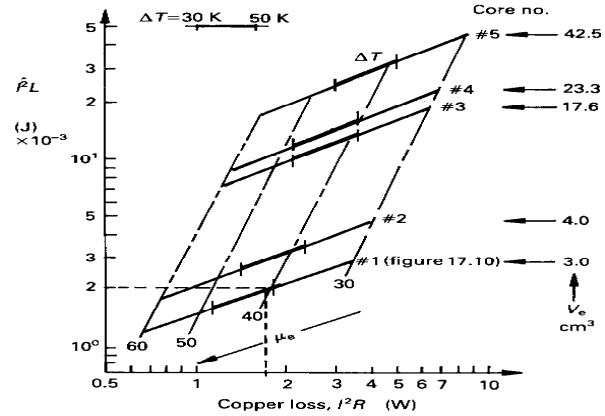


Figure 26.15. *Magnetic biasing capability $\hat{I}^2L$, copper loss I^2R, effective permeability μ_e and over-temperature ΔT of five different effective volume V_e ferrite cores.*

26.5.1i - Core temperature and size considerations

Figure 26.15 relates stored energy, LI^2, and copper loss, I^2R, for different cores of the same ferrite type. Once L and I are fixed, figure 26.15 can be used to determine the optimum core size and air gap. This figure shows that with increasing air gap (decreasing μ_e), the magnetic biasing capability increases along with the associated copper loss, I^2R. A flowchart is shown in figure 26.16, which outlines the inductor iterative design procedure to be used in conjunction with figure 26.15.

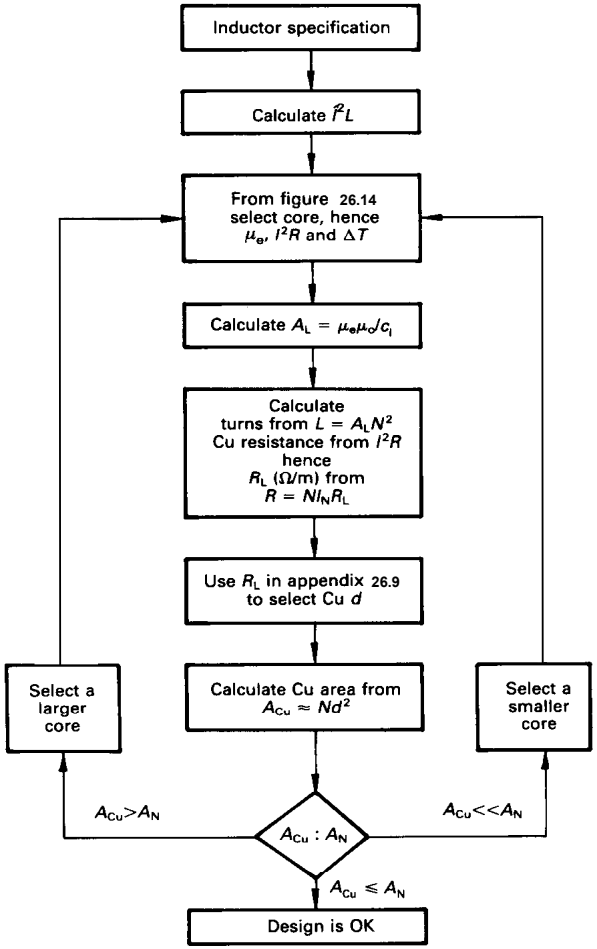


Figure 26.16. *Linear inductor design flowchart.*

Example 26.4: ***Inductor design including copper loss***

With the aid of figure 26.15, design a 20 μH, 10 A dc inductor, calculating the copper loss and temperature rise for the predicted optimum air gap and number of turns.

Solution

Following the procedure outline in the flowchart of figure 26.16
Evaluating $LI^2 = 20 \times 10^{-6} \times 10^2 = 2$ mJ
From the monogram in figure 26.15 use core # 1, with $\mu_e = 40$ and $I^2R = 1.8$ W. This copper loss will produce a 50°C temperature rise above ambient on the core surface, beneath the winding. The thick bars in figure 26.15 represent a 30-50°C temperature increase range.
The core type # 1 has A_L and μ_e values versus total air gap, and effective parameters as shown in figure 26.13 and table 26.5. For $\mu_e = 40$, $A_L = 45$ nH, a total air gap of 2.7 mm is required.
From $L = A_L N^2$

$$N = \sqrt{20 \times 10^3/45} = 21 \text{ turns}$$

For $I = 10$ A dc, $I^2R_{Cu} = 1.8$ W, then $R_{Cu} = 18$ mΩ. The copper turns diameter is determined from

$$R_{Cu} = N\,\ell_N\,R_L \quad (\Omega) \tag{26.75}$$

where R_L is the resistance per meter, Ω/m.
ℓ_N is the mean turn length which is either provided for a given former or may be estimated from core physical dimensions. From table 26.5:

$\ell_N = 52$ mm
$R_L = R_{Cu} / N\ell_N$
$= 18 \times 10^{-3}/21 \times 52 \times 10^{-3} = 0.165$ Ω/m

Using standard wire tables, Appendix 26.12 for 0.165 Ω/m, use 28 SWG (0.154 Ω/m) which has a diameter of 0.36 mm and 0.434 when enamelled. The resultant copper current density is 77 A/mm^2.

In many applications 4 A/mm^2 is used for finer gauge wires up to 20 A/mm^2 for heavier gauge wires. These current densities represent about 5 per cent of the fusing current, I_{fusing} which is approximated by

$$I_{\text{fusing}} = 80\,d^{1.5}$$

The diameter d is in mm.
This constraint is unrealistic and inductor and transformer design is based on temperature rise.
The approximate copper area is

$$A_{Cu} = N \times d^2$$
$$= 21 \times 0.434^2$$
$$= 3.88\text{mm}^2$$

From table 26.5, the useful winding cross-section is

$A_N = 56$ mm^2

Only 8 per cent of the former window area is filled, hence the actual copper length is overestimated and I^2R loss, hence temperature rise, will be less than the allowed 1.8 W and 50°C respectively.

♣

Comparing the design of examples 26.3 and 26.4, it will be seen that the same design specification can be fulfilled with the latter core of 20 per cent the volume of the former. The bigger core required an 0.33 mm air gap to give $\mu_e = 300$, while the smaller core required a larger gap of 2.7 mm to give $\mu_e = 40$. Both cores are of the same ferrite type. The incremental inductance of the smaller core will fall off with current, much faster than with the larger core, as indicated by figure 26.5.
For a switch mode power supply application, the rms value of current is less than the peak current at which the inductance is specified. The copper loss, hence temperature rise, is then based on an rms current basis.

26.5.2 Saturable inductors

Saturable inductors are used in series with semiconductor switching devices in order to delay the rise of current, thereby reducing switch turn-on stress and loss. In the case of a power transistor, the collector current is delayed until the collector voltage has fallen (see 8.3.4). For thyristor devices, the delay time allows the gate activated cathode area to spread hence giving a high initial *di/dt* capability. In each case the inductor supports the supply voltage, then after a finite time saturates to a very low inductance, supporting little voltage, and does not influence the switch current.
Ferrites are ideal as the core of a saturable inductor because of their low magnetic field strength, H_s, at the onset of flux density saturation, B_s. While the inductor supports voltage, v, the flux density increases, moving up the *B-H* curve at a rate according to Faraday's law

$$v = NA_e \frac{dB}{dt} \tag{26.76}$$

A low magnetising current results. After a finite time the flux density reaches the knee of the *B-H* curve (B_s, H_s), the core saturates and the incremental permeability falls from an initially high value to that of air, $\mu_\Delta = 1$. The high initial permeability, hence high inductance, limits the initial current. The time t_s, for the core to saturate should be equal to the switch turn-on voltage fall time, t_{fv}. The low saturation inductance allows the switch current rapidly to build up to a level dictated by the load.
If the switch voltage fall is assumed linear then the inductor voltage rise is $V_s\,t / t_{fv}$. The time t_s, taken to reach core saturation (B_s, H_s) from integration of Faraday's law is (see Chapter 8.3.4)

$$t_s = \frac{2NA_eB_s}{V_s} \tag{26.77}$$

for $t_s \leq t_{fv}$.

The flux density, hence H, and current increase quadratically with time, $I_s\left(t/t_{fv}\right)^2$. At saturation the magnetising current magnitude (hence switch current) is

$$I_s = \frac{H_s \ell_e}{N} \quad \left(= \frac{2B_s H_s V_e}{t_s V_s}\right) \quad \text{(A)} \tag{26.78}$$

which should be small compared with the switch on-state full-load current magnitude.

The inductance before saturation is given by

$$L = A_L N^2 \quad \text{(H)} \tag{26.79}$$

and falls to that of an air-cored inductor, viz.:

$$L_{sat} = c\,N^2 \quad \text{(H)} \tag{26.80}$$

after saturation, when leakage and lead length will, in practice, dominate inductance.
The energy stored in the core (pre-saturation) and subsequently dissipated at core reset is given by

$$\begin{aligned} E &= \tfrac{1}{2} B_s H_s V_e \quad \left(= \tfrac{1}{4} I_s V_s t_s\right) \\ &= \tfrac{1}{2} B_s H_s A_e \ell_e \quad \text{(J)} \end{aligned} \tag{26.81}$$

which must be minimised.

Table 26.4 summarises saturable inductor requirements based on equations (26.77) to (26.81).

Table 26.4: Design requirement of a saturable inductor

Given V_s and t_{fv}			Material dependent		Core shape dependent		
			H_s	B_s	A_e	ℓ_e	N
Minimise energy $E = \frac{1}{2} B_s H_s A_e \ell_e$	E	(J)	low	low	low	low	×
Maximise time $t_s = 2NA_eB_s/V_s$	t_s	(s)	×	high	high	×	high
Minimise mag current $I_s = H_s \ell_e / N$	I_s	(A)	low	×	×	low	high
Maximise inductance $L_e = N^2 A_e B_s / \ell_e H_s$	L	(H)	low	high	high	low	high
Requirement			low H_s (high μ_r)	-	-	short ℓ_e	high N
Compromise			-	high B_s if H_s is low	high A_e if ℓ_e is short		

26.5.3 Saturable inductor design

Figure 26.17 shows a saturable inductor iterative design flowchart. The design starting point is the type of ferrite. The desired ferrite should have minimal high frequency loss, associated with a low magnetic field intensity, H_s, at saturation. These features would be associated with ferrites having a low coercive force, H_c and remanence, B_r. The ferrite material shown in figure 26.4 fulfils these requirements with

$$H_c = 12\text{A/m} \qquad B_r = 0.18\text{T}$$
$$H_s = 200\text{A/m} \qquad B_s = 0.4\text{T}$$

Ferrites with lower magnetic field strengths are available but tend to be limited in size. A material with a high initial permeability is one indicator of a suitable ferrite type.
The next considerations are core shape and effective core parameters such as effective length, ℓ_e and area, A_e. The core should have a short effective length, ℓ_e. The area and length are traded in maintaining sufficient copper window area, A_N.
A core shape without an air gap will produce the highest possible initial, hence effective, permeability. Example 26.5, which follows, illustrates that a toroid (or tube) core offers a good solution.
A high number of turns, N, is desirable, and preferred to an increase in area, A_e.
Design should be based on the maximum core temperature. An increase in temperature decreases H_s at a faster rate than B_s, as shown in figure 26.4. From equation (26.77), many turns are required which, in combination with decreased H_s, advantageously decrease the magnetising current, I_s.

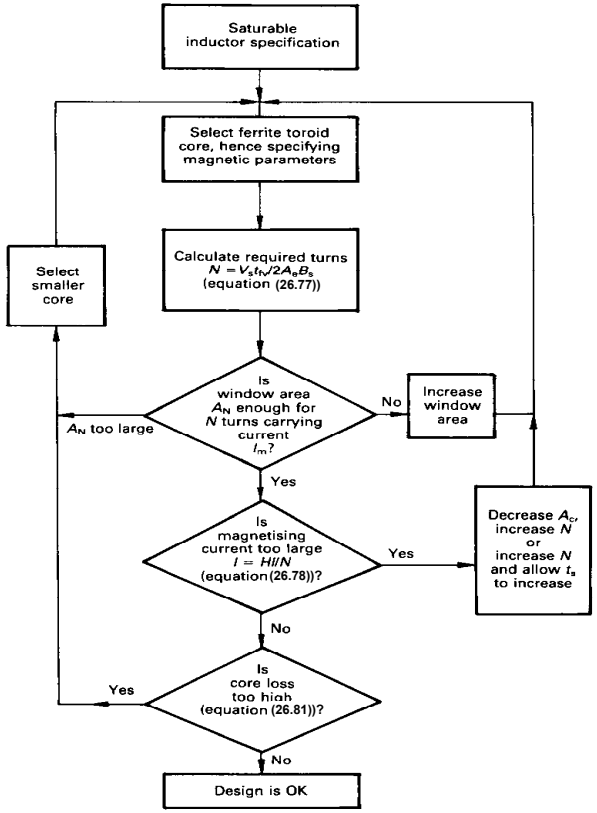


Figure 26.17. *Saturable inductor design flowchart.*

Table 26.5: Pot, toroid, and E-core design data.
Applicable magnetic data are presented in appendix 26.9

		Physical dimensions (mm)		
		Pot core d_o = 25 h = 16	Toroid d_o = 39 d_i = 24.77 h = 6.61	E-core (pair) See figure 26.13
V_e	cm^3	3.63	3.86	3.02
A_e	cm^2	0.999	0.398	0.525
ℓ_e	cm	3.64	9.71	5.75
c_1	cm^1	3.64	24.4	10.9
A_{min}	cm^2	0.95	0.398	0.45
A_L	nH	4300	1540	(ε = 0)1750
μ_e		1245	(μ_i) 3000	(ε = 0)1500
c	nH	3.45	0.51	1.15
A_N	cm^2	0.357 (0.266)	4.75	0.56
ℓ_N	cm	5.3 (5.35)	7.6	5.2
S_A	cm^2	18.4	48.7/58	20
Weight	g	23.4	19.3	2 × 8

Example 26.5: ***Saturable inductor design*** (also see example 8.6)

A pot, toroid, and E-shaped core of the same Mn-Zn ferrite as characterised in appendix 26.9, and of similar volume, have characteristics and parameters as shown in table 26.5, with H_s = 200At/m.
Design a saturable inductor for each core shape, for a switch having a 200 ns linear voltage fall time at turn-on when switching on a V_s = 600 V dc supply rail.
The core is to saturate when the switch voltage reaches saturation (0 V), after 200ns.
Estimate the core power removed at reset if the switching frequency is 20 kHz.

Solution

		Pot	Toroid	E-cores (ε = 0)
From equation (26.77) $N = V_s t_{fv} / 2A_e B_s = 1.875/A_e$ A_e is in cm^2		2	5	4
From equation (26.78) $I_s = H_s \ell_e / N = 2\ell_e / N$ ℓ_e is in cm	(A)	3.64	3.85	2.83
From equation (26.79) $L = A_L N^2 \times 10^{-3}$	(μH)	20.2	38.5	28.0
From equation (26.80) $L_{sat} = c N^2$	(nH)	13.8	12.75	18.4
From equation (26.81) $P_d = V_e \times 8 \times 10^{-1}$ V_e is in cm^3	(W)	2.90	3.09	2.42

Based on the available copper window area, A_N and number of turns, the cores would be applicable to switching currents in excess of 100 A. Smaller cores could be used for lower current levels, although window area A_N tends to dictate the required core.
From equation (26.81), the power dissipated at 20 kHz (last row in the table) is given by $P_d = \tfrac{1}{2} B_s H_s V_e f_s$.

♣

26.6 Power ferrite transformer design

Above a few kilohertz, Mn-Zn ferrite material is almost exclusively used for power transformer cores, and has been optimised by manufacturers for a wide frequency range. Specific core shapes have also been developed to cover a wide power range. In the case of voltage transformers, at 20 kHz and below 100 W, pot cores are used, or when low flux leakage and low emi are important. Such cores can be processed on automatic machines which wind and assemble the whole unit. At powers above 100 W, E-E and E-I cores are extensively used.
The usable power range of the pot core is increased by increasing frequency and at 500 kHz no alternative exists, because of the low leakage flux, low self-capacitance, and good shielding offered by pot cores.

26.6.1 Ferrite voltage transformer design

To simplify ferrite core selection, manufacturers provide the characteristic curves given in figure 26.18 which show the power that can be transmitted by various core shapes. Specifically, these curves show power for the modes of operation commonly used in switch-mode power supplies; such as push-pull, forward, and flyback, as considered in chapter 17, versus the core plus copper volume.
A formal transformer design approach based on copper and core losses is shown in the flowchart in figure 26.19 and is applicable to all smps types.

Stage 1 and stage 2

The transformer, primary and secondary voltages, currents, and powers, hence efficiency, must be specified or determined. Other requirements are switching frequency, ambient temperature, and allowable temperature rise at the core to copper interface. The final specification should include

$$\begin{matrix} V_p & V_s \\ I_p & I_s \\ P_p & P_s \end{matrix} \qquad \eta, f, T_a, \Delta T$$

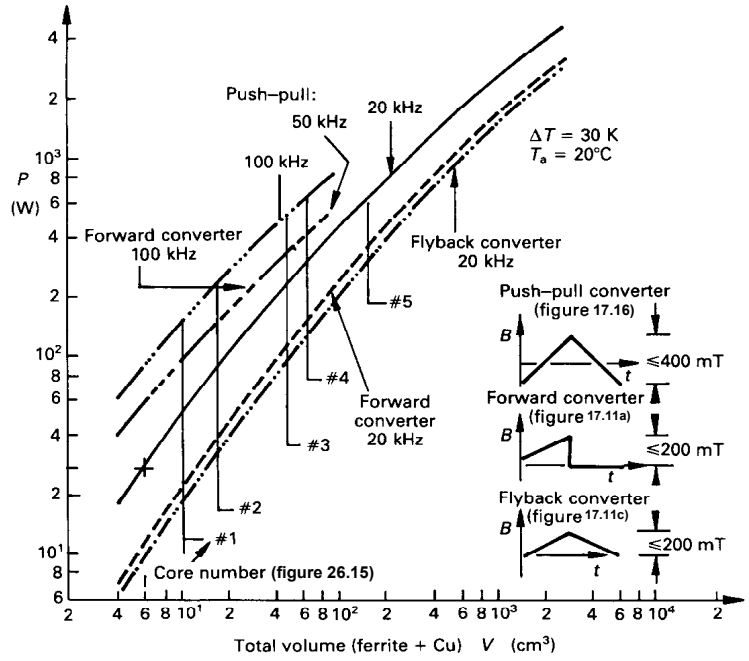


Figure 26.18. *Transmissible power, P, versus volume (ferrite plus copper), V, of transformers with ferrite Mn-Zn cores.*

Stage 3

The difference between input power and output power is the total power loss, P_L, which comprises copper and core losses. The maximum efficiency is obtained when the copper loss equals the core loss.

Stage 4

The total power loss, P_L, ambient temperature, T_a, and temperature rise, ΔT, specify the exposed copper and core surface area requirements, S_A, according to (see equation 5.4)

$$S_A = \frac{P_L}{\Delta T \, S_d} \quad (\text{m}^2) \qquad (26.82)$$

where S_d is a surface dissipation factor.
Empirical equations are commonly provided for S_d. Based on the assumption that thermal stability is reached half by convection and half by radiation, the surface area requirement can be approximated by

$$S_A = 145 \times \left[\frac{1000}{T_a + 273}\right]^{2.06} \frac{P_L}{\Delta T^{1.22}} \quad (\text{cm}^2) \qquad (26.83)$$

Stage 5

A core with the minimum surface area, S_A, is selected using manufacturers' data, ensuring that the ferrite type is appropriate to the operating frequency and that the core shape meets any engineering, cost or other special requirements. The manufacturers' data required include the effective dimensional parameters, copper winding area, A_N, and average turn length, ℓ_N.
Some manufacturers provide transformer design data for each core. This specific data can be employed, rather than the general procedure that follows.

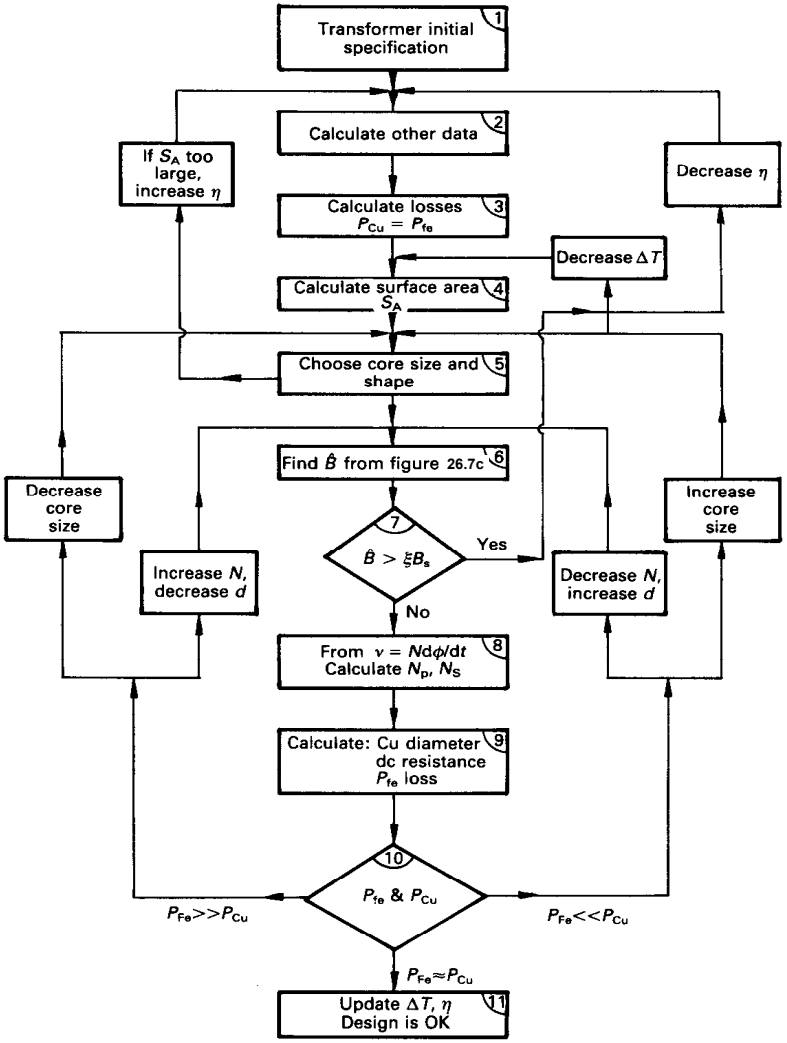


Figure 26.19. *Voltage transformer design flowchart.*

Stage 6

Using the core volume, V_e, and core loss $P_c = ½P_L$, whence core loss per cm^3, $P_w = P_c/V_e$ the maximum allowable operating flux density, B_{op}, for the specified frequency can be determined from the power loss curves in figure 26.8c.

Stage 7

The rated saturation flux density, B_s, cannot safely be used. For a transformer using both quadrants of the *B-H* characteristics, for example, a push-pull smps transformer

$$B_{op} \le 0.8\, B_s$$

while for a core used with a flux bias

$$B_{op} \le 0.4\, B_s$$

These limits avoid operational saturation of the core in one direction. If the working flux density, B_{op}, is too high, either

- reduce the efficiency and go to stage 1/2, or
- reduce the allowable temperature rise and go to stage 4.

Stage 8

The required number of primary turns, N_p, can be calculated from Faraday's law, which yields

$$N_p = \frac{V_p}{k\, B_{op} A_e f} \tag{26.84}$$

where $k = 4$ for a square-wave voltage
$k = 4.44$ for a sine wave.
If $B_{op} > 100$ mT, the effective area, A_e, in equation (26.84) is replaced by the core minimum area section, A_{min}, since that portion experiences the highest flux density.

The number of secondary turns is calculated according to

$$N_s = N_p \frac{V_s}{V_p} \tag{26.85}$$

Stage 9

The winding diameter, d_p, for the allotted primary for a window area, A_p, is calculated according to

$$d_p = 2\sqrt{\frac{A_p k_w}{\pi N_p}} \quad \text{(m)} \tag{26.86}$$

where k_w is a winding space factor, 0.7, which accounts for insulation, winding taps, shielding, air space, etc. A similar expression for the diameter of the secondary, d_p, involves the number of secondary turns, N_s, and allotted area, A_N. The total winding area $A_p + A_s$ must not exceed the available winding area, A_N.

Standard copper wire tables, Appendix 26.12, provide the resistance per meter, R_L, for the calculated diameters. From equation (26.75), the dc resistance of the primary can be calculated according to

$$R_p = N_p\, \ell_N\, R_{Lp} \quad (\Omega) \tag{26.87}$$

Similarly for calculating the secondary dc resistance, R_s. The total copper loss can be calculated as

$$P_{Cu} = I_p^2 R_p + I_s^2 R_s \quad \text{(W)} \tag{26.88}$$

Stage 10

The core loss, P_c, and the copper loss, P_{Cu} are compared. If

(i) $P_{Cu} > P_c$

Either decrease the number of turns and increase the copper diameter. This will reduce the copper loss and increase B_{op}, and hence P_c. Recalculate from stage 6.
or select a larger core, which will increase the copper window area, A_N, hence increasing the allowable wire diameter. Recalculate from stage 5.

(ii) $P_{Cu} < P_c$

Either increase the number of turns which will reduce diameter d, B_{op} hence P_c, and then recalculate from stage 6.
or select a smaller core, which will require d to be reduced, and then recalculate from stage 5.

Proceed if $P_{Cu} \approx P_c$.

Stage 11

Update the value of total losses, P_L, and hence recalculate the power requirements and resultant efficiency.
Calculate the actual core temperature rise from equation (26.83), rearranged

$$\Delta T = 59\left(\frac{1000}{T_a + 273}\right)^{1.69} \times \left(\frac{P_L}{S_A}\right)^{0.82} \quad \text{(K)} \tag{26.89}$$

where S_A is the heat dissipating area in cm^2 of the chosen core.

Example 26.6: ***Ferrite voltage transformer design***

Consider the design requirements for the split dc rail push-pull smps shown in figure 15.15b, which is specified as follows

v_o = 5 V	V_s = 48 V ± 15 per cent
i_o = 4A	f = 20kHz
P_o = 20 W	T_a = 25°C, $\Delta T \le$ 35 K
	η = 97 per cent (excluding secondary stage losses)

Solution

Based on the flowchart in figure 26.19 and the eleven stages outlined, design proceeds as follows.

Stage 1

The transformer must deliver 20 W plus losses associated with an output inductor and the pair of Schottky diodes in the output rectifier. The inductor loss is estimated at 4 per cent of the output power, 0.8 W, while the diode total loss is 0.6 V × 4 A = 2.4 W. Thus the transformer output power requirement P_s is 23.2 W (20 W + 0.8 W + 2.4 W). With a 97 per cent efficiency, the transformer input power, P_p, requirement is 1/97 per cent of 23.2 W, namely 23.9 W. The nominal primary current, I_p, at the nominal voltage, 24 V is

$$I_p = \frac{P_p}{V_{pn}} = \frac{23.9\text{W}}{24\text{V}} = 1\text{A}$$

The maximum primary voltage, V_p, is ½ ×1.15×V_{sec} = 27.6 V, since the 48 V supply is centre tapped and has + 15 per cent regulation. For worst case, it is assumed that the voltage drop across the switches is zero.
The transformer secondary voltage, V_{sec}, for the centre tapped full-wave rectifier circuit, must be large enough to overcome the diode voltage drop, V_d, and must allow for averaging of the nominal low duty cycle switching action of the primary input power. With pwm regulation each input switch operates for approximately 25 per cent of the time, thus

$$½V_s = 2\times(V_o + V_d)$$

where the ½ indicates that half of the secondary winding conducts at any one time, while the 2 approximates the pwm average on-time. Thus for V_d = 0.6 V and V_o = 5 V

$$V_{sec} = 4\times(5+0.6) = 22.4\text{V}$$

Stage 2

Extracting the transformer data from stage 1

I_p = 1A	I_s = 4A
V_p = 27.6 V	V_{sec} = 22.4 V
P_p= 23.9 W	P_s = 23.2 W

η = 97 per cent, f = 20 kHz, T_a = 25°C, ΔT = 35 K

Stage 3

The total transformer power loss, P_L, from $P_s - P_p$, is 0.7 W.
Thus $P_c = P_{Cu}$ = ½ ×0.7 = 0.35 W each.

Stage 4

The surface area requirement is calculated from equation (26.83)

$$S_A = 145\times\left[\frac{1000}{25+273}\right]^{2.06}\times\frac{0.7\text{W}}{35^{1.22}}$$

= 16.1 cm^2 for a 35 K temperature rise

Stage 5

Either the pot core in table 26.5 or the pair of E-cores in figure 26.13 have sufficient surface area, 18.4 and 20 cm^2 respectively, and both are of a ferrite material suitable for a 10 to 100 kHz operating frequency range.
At the low power level of 23.9 W (<100W), choose the pot core.

Stage 6

Using the technical data given in table 26.5, the core loss per unit volume is calculated

$$P_w = \frac{P}{V_e} = \frac{0.35\text{W}}{3.63\text{V}} = 0.096\text{W/cm}^3$$

Stage 7

From figure 26.8c, an operating flux density of 0.21 T at 20 kHz will result in the allowable core loss of 0.1 W/cm^3. For push-pull operation, the maximum allowable flux density is about 80 per cent of B_s, that is, 80 per cent of 0.48 T, namely 0.38 T.

Since 0.21 T < 0.38 T, a working flux density of 0.21 T is acceptable.

Stage 8

Since the operating flux density is greater than 100 mT, the pot core minimum area, A_{min} (0.95 cm^2) is used for calculations, rather than the effective area, A_e (0.999 cm^2). From equation (26.84), the required number of primary turns is given by

$$N_p = \frac{27.6\text{V}}{4\times0.21\text{T}\times0.95\times10^{-4}\times20\times10^3} = 17.3$$

Use 17 turns.

The number of secondary turns is given by equation (26.85)

$$N_s = \frac{V_s}{V_p}N_p = \frac{22.4}{24}\times17 = 15.7$$

where the nominal primary voltage is used. Use 16 turns per secondary winding.

Stage 9

From table 26.5, the available winding area, A_N, is either 0.357 cm^2 for a one-section former or 0.266 cm^2 for a two-section former. Since the primary and secondary voltages are relatively low, insulation and isolation present few difficulties, hence single enamel copper wire and a single section former can be used. The available copper area, 0.357 cm^2, is divided between the primary and secondary so as to provide a uniform current density within the winding area. The primary to secondary currents are in the ratio of 1:4, hence 0.285 cm^2 is allocated to the secondary (approximately 80 per cent) while 0.072 cm^2 is allocated to the primary winding. The copper wire diameter is calculated using equation (26.86)

$$d_p = 2\sqrt{\frac{A_p k_w}{\pi N_p}} = 2\times\sqrt{\frac{0.072\times10^{-4}\times0.8}{17\pi}} = 0.66\text{ mm}$$

$$d_s = 2\sqrt{\frac{A_s k_w}{\pi N_s}} = 2\times\sqrt{\frac{0.285\times10^{-4}\times0.8}{32\pi}} = 0.95\text{ mm}$$

Using the standard wire tables in Appendix 26.12 and equation (26.75) to calculate the winding resistance

		Primary	Secondary	
d_{Cu}	mm	0.6	0.95	bare Cu
d_{Cu+en}	mm	0.65	1.017	single enamel
R_L	Ω/m	0.06098	0.02432	
R_{Cu}	Ω	0.055	0.0206/16 turns	

The total power copper loss is given by equation (26.88)

$$P_{Cu} = I_p^2 R_p + I_s^2 R_s$$
$$= 1^2\times0.055 + 4^2\times0.0206$$
$$= 0.055 + 0.330 = 0.385\text{W}$$

Stage 10

The core loss is 0.35 W while the copper loss is only slightly higher at 0.385 W. No iterative change is necessary. The updated total loss, P_L, is 0.735 W.

Stage 11

The secondary power requirement remains 23.2 W while the primary requirement has increased to 23.94 W. The efficiency has been reduced to

$$\eta = \frac{23.3\text{W}}{23.94\text{W}} \equiv 96.9\,\text{per cent}$$

from 97 per cent.

Using the actual core surface area, 18.4 cm^2, and loss, 0.735 W, the core temperature rise can be calculated from equation (26.89)

$$\Delta T = 59 \times \left(\frac{1000}{25+273}\right)^{1.69} \times \left(\frac{0.735}{18.4}\right)^{0.82}$$
$$= 32.6\ \text{K}$$

which is less than the 35 K allowable temperature rise limit.

♣

The transformer design of example 26.7 could be based on figure 26.18. The volume of the core plus copper, for the pot core in table 26.5, can be estimated from its diameter of 25 mm and height of 16 mm. This yields a total volume of 6 cm^3, after allowing for slots.

Using figure 26.18, for a total volume of 6 cm^3, at 20 kHz, for a push-pull converter, 28 watts can be transmitted in a 20°C ambient, producing a 30 K core temperature rise. These results and those from example 26.6 compare as follows.

		Figure 26.18	Example 26.6
P	W	28	23.2
ΔT	K	30	32.6
T_a	°C	20	25

All other operating conditions are identical. Any design discrepancy is accounted for by

- the higher ambient temperature
- the poorer winding slot utilisation.

A centre tapped secondary represents poorer slot utilisation compared with using a single winding, which requires four rectifying diodes since, because of a limited core size range, the same core would be used independent of the type of secondary circuit. A centred tapped secondary would result in the cost saving associated with two fewer Schottky diodes.

Transformer VA rating

The VA rating of a given transformer core need only be limited by Faraday's law, equation (26.84), and the current density in the coils. Faraday's equation uniquely specifies the number of turns, whence the VA rating is then only confined by the copper current density, *J*.

$$P_{trans} = V \times A = kN\Delta B_{op} A_e f \times JA_n k_w$$

where k_w is the copper winding or fill factor.

Transformer windings are usually designed to operate at less than 5% of the fusing current level of copper. The *VA* equation shows that a transformer can be exploited to transfer virtually limitless energy for short periods (a few seconds) by tolerating an extremely high copper current density, *J*. The penalty of briefly operating at a high percentage of the fusing current density of copper is the poor regulation due to the copper I^2R losses. The *VA* equation does not explicitly involve core length, which is a factor that determines magnetizing inductance. This design consideration is not a constraint when using nanocrystalline cores, at frequencies in excess of 1kHz, because of their extremely high relative permeability's, typically in excess of 30,000.

26.6.2 Ferrite current transformer

By adding a secondary winding, a linear inductor can be converted into a voltage transformer, while a saturable inductor can be converted into a current transformer. The linear inductor and voltage transformer (of E-I laminations) are characterised by a core with an air gap (inherent in transformers which use E-I laminations). The saturable inductor and current transformer generally use an un-gapped core.

A given transformer can operate either in the voltage mode or the current mode depending on the load impedance. The voltage transformer operates into a high impedance circuit, while the current transformer requires a low impedance load. Current and voltage transformer action both cease at core saturation, and air coupling results in a low coupling factor, $k \to 0$.

The equivalent circuit model is identical for each transformer mode, and the same basic equations apply in each case. A saturable inductor is required to support a large voltage for a short period, while a current transformer supports a low voltage for a long period. In each case, the primary voltage-time product is equal for a given core and primary turns.

26.6.3 Current transformer design requirements

The basic requirement of a current transformer is a fixed ratio between the primary and secondary currents according to

$$I_p N_p = I_s N_s \quad \text{(A)} \tag{26.90}$$

Ideally the load impedance is zero, hence zero secondary voltage is developed. Practically, a secondary voltage, V_{sec}, exists (at least due to winding resistance), whence from Faraday's law

$$V_{\text{sec}} = N_s \frac{d\phi}{dt} = N_s A \frac{dB}{dt}$$

For a constant secondary voltage (a short circuit), the core flux density increases linearly, effectively moving up the *B*-*H* curve at a constant rate and reaches saturation, B_s, in time

$$t_s = \frac{N_s B_s A_e}{V_{\text{sec}}} \quad \text{(s)} \tag{26.91}$$

Where a core is operated with an *H* offset, for example, as with an inductor carrying dc current or a unidirectional current transformer, the maximum value of flux density used for analysis should be reduced because of remanence to $B_s - B_r$, whence

$$t_s = \frac{N_s (B_s - B_r) A_e}{V_{\text{sec}}} \quad \text{(s)} \tag{26.92}$$

The lower the secondary voltage, V_{sec}, the longer the time before saturation, at which point current transformer action ceases. The core is fully reset by a negative voltage of sufficient duration for which the voltage-time product must equal that *v*-*t* product of the on-period. Fortunately, a high reset voltage can generally be employed, which produces a short reset time. Effectively, the reset voltage forces the magnetising current, or more accurately, flux whence the stored energy, to zero.

This magnetising current, $\breve{I}_p$ should be minimal and its presence modifies the ideal ampere-turns balance according to

$$\left(I_p - \breve{I}_p\right) N_p = N_s I_s \quad \text{(A)} \tag{26.93}$$

where the magnetising current $\breve{I}_p$ is given by

$$\breve{I}_p = \frac{H \ell_e}{N_p} \quad \text{(A)} \tag{26.94}$$

The initial magnetising current is zero, and for a constant secondary voltage, increases linearly with time. Low leakage core shapes should be used to minimise leakage inductance.

At all times the primary voltage is related to the secondary voltage according to

$$V_p = V_{\text{sec}} \left(\frac{N_p}{N_s}\right) = \frac{V_{\text{sec}}}{n_T} \quad \text{(V)} \tag{26.95}$$

The requirements of the previous equations are summarised in table 26.6 where the maximum on-time is $\breve{t}_{on}$, while the time available for reset is the minimum off-time $\breve{t}_{off}$. The secondary reset voltage V_{sr} requirement and associated dissipated energy *W* are also included. This table summarises key current transformer requirements as follows

- use a core material which has a low magnetising force, H_s, at saturation
- use a core with a short effective core path length, ℓ_e
- use a high number of turns, N_p and N_s, for a given turns ratio
- operate the transformer with a low secondary voltage, V_{sec}.

Table 26.6: Current transformer requirements showing how magnetic parameter variation affects electrical characteristics

		core parameters					circuit parameters		
		H_s	B_s	B_r	A_e	ℓ_e	N_s	N_p	V_{sec}
Equation (26.92) $t_s = \frac{N_s(B_s - B_r)A_e}{V_{sec}}$	s	*	↑	↓	↑	*	↑	↑	↓
Equation (26.94) $\check{I}_p = \frac{H_s \ell_e}{N_p} \frac{\hat{t}_{on}}{\check{t}_{off}}$	A	↓	*	*	*	↓	*	↑	*
Equation (26.98) $V_{sr} = V_p \frac{\hat{t}_{on}}{\check{t}_{off}}$	V	*	*	*	*	*	*	*	↓
$W = H_s(B_s - B_r)A_e\ell_e\left(\frac{\hat{t}_{on}}{t_s}\right)^2$	J	↓	↓	*	↓	↓	*	*	*
Design requirements		low H_s	—	low B_r	—	low ℓ_e	high turns		low V_{sec}

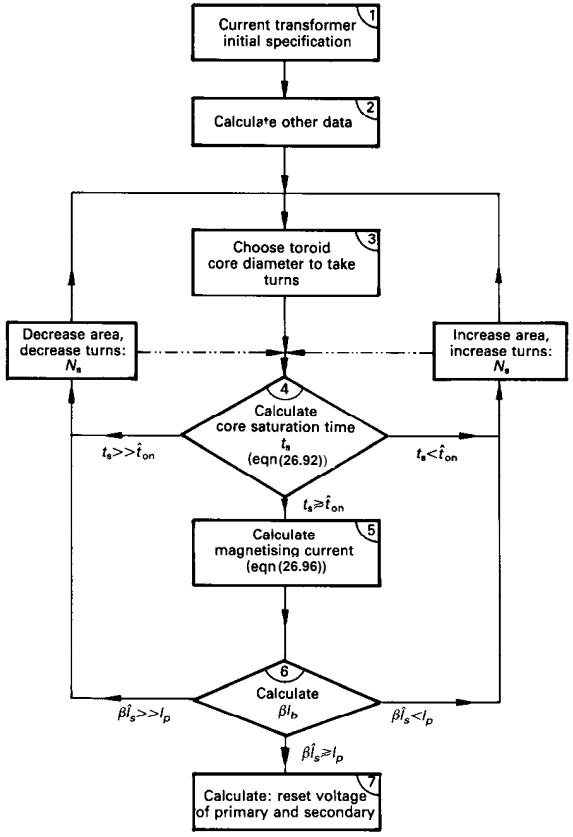


Figure 26.20. *Current transformer design flowchart.*

26.6.4 Current transformer design procedure

Figure 26.20 shows a flowchart design procedure for a current transformer and the design stages are as follows.

Stage 1/stage 2

The current transformer primary and secondary currents, hence turns ratio $n_T = N_s/N_p$, must be specified with the limits on duty cycle times, $\hat{t}_{on}$ and $\check{t}_{off}$. The expected secondary voltage V_{sec} must be specified.

Stage 3

Select a ferrite toroid with an internal diameter, hence window area A_N, sufficient to accommodate the required minimum turns, $n_T + 1$. The copper turns current ratings must be taken into account. The core specifies the effective parameters ℓ_e, A_e, and V_e. The ferrite type specifies B_s, B_r, and H_s.

Stage 4

Calculate the time t_s, before the core saturates from equation (26.92). This time must be greater than the required maximum output current pulse width, $\hat{t}_{on}$.

(i) If $\hat{t}_{on} > t_s$
Either increase the number of turns, using a core with a larger window A_N if necessary.
or increase the core area, A_e, which can be achieved with the same window area, A_N, either with a core of increased thickness or by using two stacked cores.
go to stage 3

(ii) If $t_s >> \hat{t}_{on}$
Either decrease the number of turns which may allow a smaller core size.
or decrease the core cross-sectional area.
go to stage 3

(iii) If $t_s \geq \hat{t}_{on}$, proceed to stage 5

Stage 5

Calculate the magnetising current at $\hat{t}_{on}$

$$\check{I}_p = \frac{H_s \ell_e \hat{t}_{on}}{N_p t_s} \quad \text{(A)} \tag{26.96}$$

Stage 6

Calculate the secondary current, taking the magnetising current $\check{I}_p$ into account

$$\hat{I}_s = \frac{\hat{I}_p - \check{I}_p}{n_T} \tag{26.97}$$

Is $\beta = \hat{I}_p / I_s$ sufficiently large?

(i) If $\hat{I}_p > \beta \hat{I}_s$
Either decrease the magnetising current by increasing core area.
or increase the turns ratio, n_T.
go to stage 3

(ii) If $\hat{I}_p \ll \beta \hat{I}_s$
Either decrease the turns ratio, n_T.
or decrease the core cross-sectional area.
go to stage 3

(iii) If $\hat{I}_p \lesssim \beta \hat{I}_s$, proceed to stage 7

Stage 7

Calculate the core reset voltage

$$V_{sr} = V_p \frac{\hat{t}_{on}}{\check{t}_{off}} \quad \text{(V)} \tag{26.98}$$

Calculate the reflected primary on-state voltage during core reset

$$e_p = \frac{V_{sr} N_p}{N_s} \qquad (26.99)$$

Example 26.7: ***Ferrite current transformer design***

A current transformer primary is used in the collector of a bipolar junction transistor switching circuit and the secondary is used to provide transistor base current as shown in figure 26.21. The maximum collector current is 100 A and the transistor has a gain of 8 at 100 A, in saturation ($v_{be\ sat}$ = 1.2V).

The transistor maximum on-time is 46 µs while the minimum off-time is 4 µs. Design a suitable current transformer using the toroid ferrite core, which has low flux leakage and is specified by the data in table 26.5 and appendix 26.9. Assume a core temperature of 25°C.

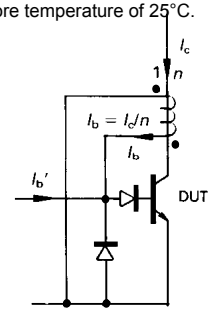


Figure 26.21. *Current transformer for BJT base drive.*

Solution

Based on the flowchart in figure 26.20 and the procedure previously outlined:

Stage 1

The required turns ratio factor is $n_T = N_p / N_s = \beta$ = 8/1. In allowing for the magnetising current component, choose n_T = 15/2.
The secondary winding voltage is the maximum transistor base to emitter voltage plus the maximum voltage drop across a series diode. Maximum voltage occurs at maximum current.

$$\begin{aligned} V_s &= V_{be_{sat}} + V_D \\ &= 1.2\text{V} + 1.2\text{V} \\ &= 2.4\text{ V} \end{aligned}$$

Stage 2

The current transformer requirements can be summarised as follows

$n_T = N_p / N_s$ = 15/2 $\quad \hat{t}_{on}$ = 46µs

V_{sec} = 2.4V $\quad \check{t}_{off}$ = 4µs

Stage 3

The ferrite toroid core specified in table 26.5, fulfils the following requirements

B_s = 0.4 T at H_s = 200 A/m

and A = 0.398 cm^2, ℓ_e = 9.71cm

while the available window area, A_N, is 4.75 cm^2. This window must accommodate two conductor turns of 100 A (plus magnetising current) each and fifteen conductor turns of 12 A each.

Stage 4

The time, t_s, before core saturation is given by equation (26.92), and assuming B_r = 0

$$t_s = \frac{15\times0.4\times0.4\text{T}\times10^{-4}}{2.4\text{V}} = 100\ \mu\text{s}$$

Since $t_s > \hat{t}_{on}$, that is 100 µs > 46 µs, proceed to stage 5.

Stage 5

The maximum primary magnetising current, $\hat{I}_p$ is specified by equation (26.96)

$$\hat{I}_p = \frac{46\mu\text{s}}{100\mu\text{s}} \times \frac{200\times9.71\times10^{-2}}{2} = 4.47\text{A}$$

Stage 6

The 4.47 A of magnetising current detracts from the primary current available for current transformer action. The maximum available secondary current under worst-case conditions is given by equation (26.97)

$$\hat{I}_s = \frac{100\text{A} - 4.47\text{A}}{15/2} = 12.7\text{A}$$

The maximum allowable collector current is this base current, 12.7 A, multiplied by the transistor gain, 8, which yields 102 A. This is larger than the specified maximum collector current of 100 A, hence the design is correct.

Stage 7

In the on-state, the secondary voltage is 2.4V and the reflected primary voltage is 0.32V.
The maximum secondary voltage, V_{sr}, required to reset the core is given by equation (26.98)

$$V_{sr} = 2.4\text{V} \times \frac{46\mu\text{s}}{4\mu\text{s}} = 27.6\text{V}$$

The reflected primary voltage is 3.7 V.
The available inherent circuit reset voltage is usually much larger, being clamped by a base circuit diode in avalanche. Therefore the core reset time will be shorter than 4 µs.
At currents much lower than 100 A, the secondary voltage is decreased, hence the magnetising current is reduced. This reduced magnetising current could consume the full collector current at collector currents of a few amperes. It is therefore necessary to add extra base current to compensate for this deficiency at low currents. The minimum secondary voltage, V_{sec}, specifies the extra requirement according to

$$I_b' = \frac{\check{V}_s}{V_s} \times \frac{\hat{I}_p}{n_T} \qquad (26.100)$$

For $\check{V}_s$ = 1.2 V, the extra base current requirement is

$$I_b' = \frac{1.2}{2.4} \times \frac{4.47}{7.5} = 300\text{mA}$$

This current can be delivered from an inductive circuit since zero extra current is initially required, and the requirement rises linearly to 300 mA in 46 µs.
A base start pulse of a few microseconds duration is required initially to turn the transistor on, whence collector current is established and current transformer action commences, and is self-sustaining.

♣

26.6.5 Current measurement: closed loop ferrite transformer

Figure 26.22 shows a ferrite current measurement transformer where a compensation winding maintains the air gap flux at zero, enabling dc (as well as ac) currents to be measured. Measurement bandwidth is typically dc to 200kHz. The current to be measured, primary current I_p, produces an mmf in the ferrite toroidal core. A Hall effect transducer detects the flux in the core air gap and an op amp compensation circuit drives current through the high turns winding in an attempt to zero the core flux. The current in the compensation winding is therefore proportional to the current being measured, according to

$$N_p I_p = N_s I_s \qquad (26.101)$$

The same transducer can be used to measure voltage by adding an external series resistor in the primary, which produces a current that is measured, which is proportional to the voltage. The number of

primary turns is usually large so as to minimize the resistor current. The resistance, in conjunction with the primary self inductance (and leakage), limit the measurement bandwidth, to the time constant *L/R*.

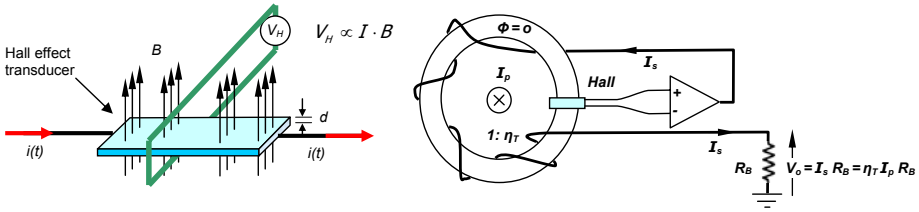


Figure 26.22. *Current measurement transducer using a flux compensated toroidal ferrite core.*

26.6.6 Current measurement: Rogowski Coil

Rogowski coils, shown in figure 26.23, are used for passive detection and versatile non-invasive measurement of alternating current (AC) or high speed current pulses (non-dc). It is typically wound on an air-core so in theory there are no effects due to hysteresis, saturation, or non-linearity.
The operating principle is that if a closely uniformly wound air-cored toroidal coil of *N* turns/m is placed axially around a straight conductor carrying current i in a closed path, the alternating magnetic field produced by the current in the conductor induces a coil output voltage *E* in the coil that is proportional to the rate of change of the cross section area *A* sq m which encircles any flux linked component produced by the current *i*, given by the expression:

$$E = -M\frac{di}{dt}$$

where *M* is the mutual inductance between the Rogowski coil and the conductor and *di/dt* is the rate of change of current in the conductor. If the coil outputs are connected to an integrator, the output signal reproduces the current waveform.
Instead of measuring the short circuit current through the coil directly, the measurement is instead the integral of the open circuit voltage.

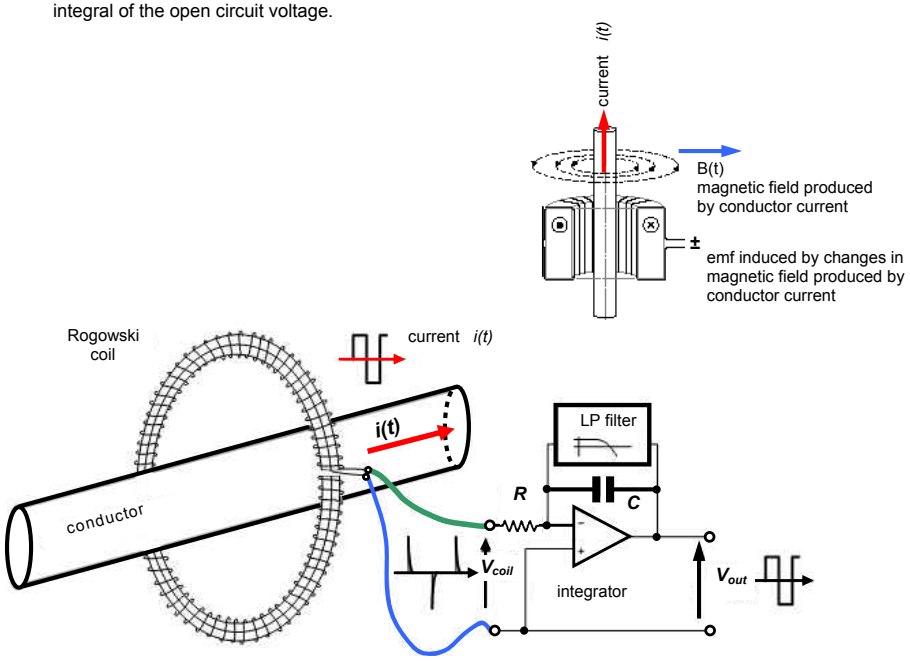


Figure 26.23. *Rogowski coil and active integrator basic operation.*

There are two advantages to the Rogowski coil.

One advantage of a Rogowski coil over other types of current transformers is that it can be made open-ended and flexible, allowing it to be wrapped around a live conductor without disturbing it.

A second advantage is that the Rogowski coil does not use a magnetically permeable core like a standard current transformer, making it of low inductance. Since it has no permeable core to saturate, it can respond linearly to extremely large currents. Being of low inductance it can also respond to very fast frequency pulses. A standard current transformer can have its core saturated at very high currents, and the inductance limits its frequency response. The closer in form to a perfectly symmetric toroidal uniform coil of wire, with equally spaced windings, the Rogowski coil is less susceptible to external electromagnetic interference.

Rogowski coil operating principle
A Rogowski coil works by sensing the magnetic field in the space around the conductor that carries the current. The relationship is given by the Ampere's Law. According to it, the line integral of the magnetic field around a closed loop is equal to the net current encircled by it, no matter what path the loop takes.

$$\oint H\cos\alpha\, d\ell = i(t)$$

The mathematical expression that shows this effect where $d\ell$ is a small element of length along the loop, *H* is the magnetic field in $d\ell$ and is the angle between the direction of the field and the direction of the element.
The magnetic field due to a long straight conductor carrying current *i*, in air, is

$$B = \mu_o H = \frac{\mu_o i}{2\pi R}$$

where $\mu_o = 4\pi\times10^{-7}$ and *R* is the perpendicular radial distance from the conductor to the point at which the magnetic field is calculated (the major radius of the toroid). The direction of the magnetic field being tangentially perpendicular to the current and to the radius *r*, and determined by use of the right hand rule.
Each turn of the Rogowski coil *N* turns produces a voltage proportional to the rate of change of the magnetic flux *B* through the turn. Assuming a uniform magnetic field density throughout the turn of area *A*, by Faraday's equation, the rate of change magnetic flux is equal to the rate of change of magnetic field density times the cross-sectional area of the turn $\pi\times r^2$ (toroid cross section radius, *r*).

$$V_{turn} = -\frac{d\Phi}{dt} = -A\frac{dB}{dt}$$

The output voltage from the coil with *N* turns, effectively series connected, is

$$V_{coil} = -NA\frac{dB}{dt}$$

Substitution of *B* gives

$$V_{coil} = -NA\frac{dB}{dt} = -\frac{\mu_o NA}{2\pi R}\frac{di}{dt} = -\frac{\mu_o NA}{S}\frac{di}{dt} = -M\frac{di}{dt} \qquad (26.102)$$

where *S* is the mean circumference of the toroid and *M* is the mutual inductance between the coil and the conductor and is independent of the frequency.
The self-inductance *L* of a coil uniformly wound with a toroidal shape toroid, which affects the output voltage frequency response, is

$$L = \mu_o N^2\left[R - \sqrt{R^2 - r^2}\right] \qquad (26.103)$$

If a rectangular cross section ring is used then the emf produced is given by

$$V_{coil} = -\frac{\mu_o NH}{2\pi}\ell n\frac{c}{b}\frac{di}{dt}$$

where *H* is the rectangular core height and *b* and *c* are the inner and outer diameter of the coil.

In order to get a voltage proportional to current an integrator - either active or passive - must be used. An active integrator, as shown in figure 26.23, using an operational amplifier is a common solution. The op-amp needs to have sufficient frequency response (both upper and lower cut off half-power points) and current sourcing and sinking capability to drive the capacitor at the expected frequency.

The integrator needs a resistor placed across the capacitor in order to be made into a leaky integrator. The resistance placed across the capacitor should be just small enough to leak off the capacitor and keep it zeroed but not so small that it interferes with the integration performance in the frequencies of interest.
Ignoring any leaky resistance added to the integrator of figure 26.23 the output is

$$V_{out} = -\frac{1}{RC}\int V_{coil}\,dt$$

After integrating the signal of equation (26.102) the total output voltage is

$$V_{out} = \frac{1}{RC}\int M\frac{di}{dt}dt = \frac{M}{\tau}i$$

The transducer/amplifier sensitivity, or transfer function, is:

$$\frac{V_{out}}{i} = \frac{M}{\tau}$$

where V_{out} is the output voltage of the integrator, $\tau = RC$ is its time constant and i is the conductor current. Changing τ, operation range can be modified and it is possible to operate from mA to MA.

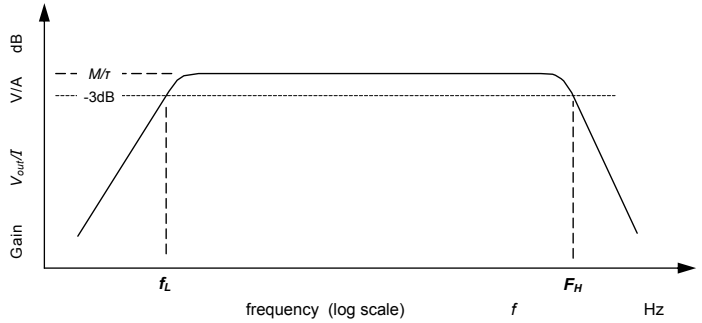


Figure 26.24. *Frequency response of a Rogowski coil and integrator*.

It is important to take into account linearity and bandwidth of integrator, and design it according to the kind of current to be measured. For high frequencies it is appropriate to use a passive integrator composed only of R and C, such that the mid-band gain is M/CR (V/A)
The relationship V_{out} proportional to i is constant across the transducer bandwidth. The bandwidth is defined as the range of frequencies from f_L to f_H for which sinusoidal currents can be measured to within 3dB of the specified sensitivity M/CR, as shown in figure 26.24.
At low frequencies, the integrator gain increases and theoretically becomes infinite as the frequency approaches zero. This would result in unacceptable dc drift and low frequency noise; hence the integrator gain is limited at low frequencies. This limitation is controlled by a low pass filter in parallel with the integrating capacitor C. The low pass filter sets the low frequency bandwidth f_L, typically less than 1Hz.
Furthermore, due to the distributed inductance, equation (26.103), and inter-turn capacitance of the Rogowski coil, there is a high frequency bandwidth f_H, (generally >1MHz) above which the measurement is attenuated and significant phase delay occurs.

Construction

There are a number of Rogowski coil current transformer types.

1 - Flexible Rogowski coil
The insulated winding is placed over a long and flexible plastic former typically between 3.5 and 15mm in diameter. The coil is fitted by wrapping it round the conductor to be measured and bringing the ends together. External insulation can be composed of one or several insulation layers (to increase the sensitivity), thermal shrinkable protection, electrostatic screen, which affects the flexibility of the coil. Electrostatic screen can be added to improve insulation of external influences.
Although less sensitive and less accurate than the rigid form, a flexible coil is better for high frequency measurements.
It is useful with large size or awkward shaped conductors or in places with limited access or where a lightweight transducer is needed which can be suspended on the conductor. As an open coil, it is not necessary to disconnect the conductor that carries the current to be measured and the user has only to unit the ends after the coil is placed around the conductor. Its form is compact and versatile.

Typical electrical features are:

Mutual inductance *M*:	between 30 and 300nH.
Maximum frequency:	between 100kHz and 1MHz depending on *M*.
Minimum frequency:	between 1 and 10Hz, depending on the integrator.
Current range:	from 1A to >1MA.
Accuracy:	1%.

2 - Rigid Rogowski coil
The rigid coils are wound on a solid plastic former, normally in a toroidal shape, and tend to be bulkier than flexible coils but have better stability. External insulation can be composed of one or several insulation layers. Alternatives are varnishing or encapsulated and potted. An electrostatic screen can be added to improve insulation to external influences.
The output voltage is stable and the accuracy is good. A rigid coil lower measurement frequency range is lower than with flexible coil, hence is more applicable for low current and low frequency measurements. Disadvantageously, being a continuous ring, the current-carrying conductor must be disconnected and placed through the core centre hole before the measurement. It can be used for high precision measurements or for permanent installation.

Typical electrical features are:

Mutual inductance *M*:	between 3 and 5μH.
Maximum frequency:	between 10kHz and 30kHz depending on *M*.
Minimum frequency:	down to 0.1Hz, depending on the integrator
Current range:	from 100mA to >100A.
Accuracy:	0.1%

3 - Planar coil
The sensor can be manufactured using a planar coil rather than a toroidal coil. In order to reject the influence of conductors outside the sensors measurement region, planar Rogowski current sensors use a concentric coil geometry instead of a toroidal geometry to limit the response to external fields. The main advantage of the planar Rogowski current sensor is that the coil winding precision that is a requirement for accuracy can be achieved using low cost printed circuit board manufacturing.

Features and applications

A Rogowski coil used as current sensor has numerous advantages:

- The air coil has no hysteresis, it does not saturate and is linear. The mutual inductance is independent of the current.
- Non intrusive.
- Good response to current transients, so they are appropriate for current pulse measurements or for protection systems.
- High bandwidth. The high-frequency limit is determined by the self-resonance of the coil and depends on the coil design. Although not applicable to DC measurement, with an accurate integrator design, it is possible to measure frequencies lower than 1Hz.
- Can measure ac signals superimposed on large dc currents.
- The same coil can measure a wide range of currents, from mA to MA, with a typical sensitivity of 1.0mV/A for ± 6000A peak, and a *di/dt* typically from 1 to 25kA/μs, over a coil temperature range from –20C° to 100°C..
- Easy calibration. Because of its linearity, coils may be calibrated at any current level.
- It is lightweight, compact and easy to install and to transport. Importantly, it is easy to use.
- Output variation with the temperature is low.
- Low power consumption, and can be totally passive. Low cost.
- High frequency bandwidth (3dB) decreases with coil length, for example, 100mm 12MHz, 200mm 8MHz.
- Rogowski Coil cross sectional diameter specifies electrical isolation, for example, 3.5mm for 2kV isolation and 4.5mm for 5kV isolation.

Another useful feature of a Rogowski sensor is immunity to far-field interference. EMF components induced by the same far field source will cancel each other. Two components of EMF induced by the current passing through the wire inside the coil will be added to each other.

One disadvantage is that the Rogowski coil produces an EMF proportional to *di/dt*. Therefore, at connecting or disconnecting instants, the EMF goes ‘infinite’. Transient voltage suppressors or other voltage protection is needed to prevent overloading the interfacing electronics. Also, accuracy is slightly dependant on position of the current carrying conductor in loop

Table 26.7 highlights the differences between five current measurement techniques, namely the conventional current transformer, Hall effect sensor, current transformer based on a Rogowski coil, flux gate (which is a Rogowski coil with a magnetic core) and a shunt resistor.

Table 26.7: Features of five current measurement techniques

Feature	Current transformer	Hall effect	Rogowski coil	Flux gate	Shunt resistor
Operating principle	$N_1I_1 = N_2I_2$	$V_H = I{\cdot}B$	$V \propto \mu_o dI/dt$	$V \propto \mu_o\mu_r dI/dt$	$V_S = I{\cdot}R_S$
bandwidth	low	medium	high	medium	low
isolation	high	high	high	high	low
linearity	good	medium	excellent	good	low
High current measurement capability	good	good	very good	very good	low
Saturation and hysteresis problems	yes	yes	no	no	no
Power dissipation	low	medium	low	medium	low
Temperature effects on output	low	medium	very low	low	medium
Transient response	medium	medium	very good	medium	low
Low frequency response	medium	very good	good	medium	low
Dc offset	no	yes	no	yes	no
Easy of installation	medium	medium	medium	medium	low
weight	medium	medium	low	medium	low
dimensions	medium	medium	low	medium	low
cost	medium	high	low	high	low

26.7 Auto-transformers

An autotransformer is a single winding electrical transformer that has at least three electrical connections called taps. The voltage source and the load are each connected to two taps. One tap at the end of the winding is a common connection to both circuits (source and load). Each tap corresponds to a different source or load voltage. In an auto-transformer a portion of the same winding is part of both the primary and secondary winding.

When the primary side of a transformer with a winding of N_1 turns is supplied with V_1 voltage (satisfying $V=Nd\varphi/dt$), then on the secondary N_2 turns, has induced V_2 voltage, in accordance with the *transformer turns ratio* η_T:

$$\eta_T = \frac{V_2}{V_1} = \frac{N_2}{N_1} = \frac{I_1}{I_2} \tag{26.104}$$

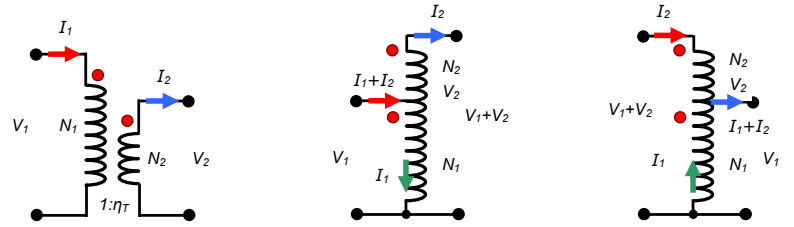


Figure 26.25. *Transformer and autotransformer connection diagram: (a) two winding transformer, (b) step-up voltage autotransformer, and (c) step-don voltage autotransformer.*

In the transformer, the power is transmitted from the primary to the secondary circuit via the magnetic field, such that the input VA is equal to the output VA, assuming losses windings and neglecting any magnetising current. That is

$$S = VA_{xfm} = V_1I_1 = V_2I_2 \tag{26.105}$$

The energy flow through the autotransformer is the summation of magnetic transformation (induction) phenomena and input to output current conduction. The conduction results from the series connection of the autotransformer's primary and secondary circuits. The accumulative (additive) series connection is assumed, as indicated by the dots in figure 26.25, since no advantages are gain when the windings are differentially connected.
The VA capability of an auto-transformer does not depend on whether the common connection point forms an input terminal or output terminal. It is assumed that each winding exploits its full VA rating, as define by equation (26.105). When the common connection forms the auto transformer input, giving a step-up voltage, as in figure 26.25b, the voltage transfer ratio is $V_{out} / V_{in} = (V_1 + V_2)/ V_1 = 1 + \eta_T$ and the input (and output) VA is:

$$\begin{aligned} VA_{autoX-i/p} &= V_1I_{in} = V_1(I_1 + I_2) \\ &= V_1\left(I_1 + \frac{I_1}{\eta_T}\right) = V_1I_1\left(1 + \frac{1}{\eta_T}\right) \\ &= VA_{xfm}\left(1 + \frac{1}{\eta_T}\right) \end{aligned}$$

When the common connection forms the auto transformer output, giving a step-down output voltage, as in figure 26.25c, the voltage transfer ratio is $V_{out} / V_{in} = V_1 / (V_1 + V_2) = 1 / (1 + \eta_T)$ and the output (and input) VA is:

$$\begin{aligned} VA_{autoX-o/p} &= V_1I_{out} = V_1(I_1 + I_2) \\ &= V_1\left(I_1 + \frac{I_1}{\eta_T}\right) = V_1I_1\left(1 + \frac{1}{\eta_T}\right) \\ &= VA_{xfm}\left(1 + \frac{1}{\eta_T}\right) \end{aligned}$$

In each case, the first term is associated with transformer action between the two windings, while the second term is that current component that conducts from the input to the output.
Generally

$$VA_{autoX} = VA_{xfm}\left(1 \pm \frac{1}{\eta_T}\right) \tag{26.106}$$

where the plus sign is applicable to when the two windings are additively connected (as in figure 26.25b and 26.25c), while the negative sign implies the two windings are connected to oppose. Opposing windings offers poor copper utilisation.

The area, hence volume, thence weight, of copper required in a winding is proportional to the number of turns and to the cross sectional area of the wire. In turn the area is proportional to the current to be carried, that is, volume of copper is proportional to NI.

Volume of copper $\propto$ length of the wire × cross sectional area of copper wire
$\propto N \times I$

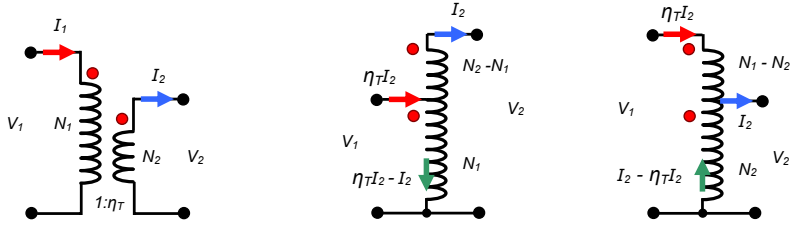


Figure 26.26. *Transformer and autotransformer diagram for V_1 and V_2 input and output voltages: (a) two winding transformer, (b) step-up voltage autotransformer, and (c) step-don voltage autotransformer.*

The magnetic circuit is assumed to be identical, satisfying $V=Nd\varphi/dt$. To quantify the copper saving, the total quantity of copper used in an auto-transformer is expressed as a fraction of that used in a two winding transformer, both with the same output VA, V_2I_2. The copper area in the two winding transformer in figure 26.26a is

$$\text{copper in two winding transformer} = k\left(N_1I_1 + N_2I_2\right) = 2kN_1I_1$$

For the step-up autotransformer, shown in figure 26.26b, with the same output VA rating (same input voltage V_1 and same output voltage and current V_2, I_2) as the two winding transformer in figure 2a:

$$\frac{\text{copper in auto-transformer}}{\text{copper in two winding transformer}} = \frac{N_1\left(\eta_T I_2 - I_2\right) + \left(N_2 - N_1\right)I_2}{N_1I_1 + N_2I_2} = \frac{2N_2I_2\left(1 - \frac{1}{\eta_T}\right)}{2N_2I_2} = 1 - \frac{1}{\eta_T}$$

The pu copper saving for the step up autotransformer is $1/\eta_T$, where $\eta_T \geq 1$.

For the step-down autotransformer in figure 26.26c:

$$\frac{\text{copper in auto-transformer}}{\text{copper in two winding transformer}} = \frac{N_2\left(I_2 - \eta_T I_2\right) + \left(N_1 - N_2\right)\eta_T I_2}{N_1I_1 + N_2I_2} = \frac{2N_2I_2\left(1 - \eta_T\right)}{2N_2I_2} = 1 - \eta_T$$

The pu copper saving for the step down autotransformer is η_T, where $\eta_T \leq 1$.

Generally, for cumulatively connected autotransformer windings:

$$\frac{\text{copper in auto-transformer}}{\text{copper in two winding transformer}} = 1 - \frac{1}{\eta_T^{\pm 1}} \qquad (26.107)$$

Generally the copper saving is $1/\eta_T^{\pm 1}$ where the positive sign is applicable to a voltage step-up connection, while the negative sign implies voltage step-down (with cumulative windings, as opposed to subtractive connection, in each case).

The current in the common part of the autotransformer winding is small compared with input and output currents, being the difference between the two currents. Thus, the cross-section of this part of the winding may be decreased, resulting considerable savings. Although the iron area is unchanged, since the voltages are unchanged ($V=Nd\varphi/dt=kNBAf$), the core window area, hence core length, can be decreased because of the copper area saving. Using smaller quantities of iron and copper results in lower losses and increased efficiency.

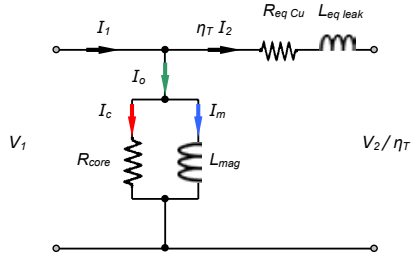


Figure 26.27. *Autotransformer equivalent circuit.*

Equivalent circuit

In figure 26.27, auto-transformer output side resistance R_2 and reactance X_2 transfer to the input according to:

$$R_{eq\ autoX} = R_1 + \left(\frac{1}{\eta_T} - 1\right)^2 R_2$$
$$X_{eq\ autoX} = X_1 + \left(\frac{1}{\eta_T} - 1\right)^2 X_2 \qquad (26.108)$$

The two winding transformer equivalent equations are

$$R_{eq\ Xfm} = R_1 + \frac{1}{\eta_T^2} R_2$$
$$X_{eq\ Xfm} = X_1 + \frac{1}{\eta_T^2} X_2 \qquad (26.109)$$

These equations show that the impedance transferred for the autotransformer is less than for the conventional two winding transformer. Thus in the case of an autotransformer, the short circuit impedance is lower. Having a smaller value of short circuit impedance is considered a disadvantage, since the short circuit currents are larger. But the full load regulation is lower (better).
The autotransformer short circuit low voltage compared to a transformer is:

$$\frac{V_{Z\ autoX}}{V_{Z\ Xfm}} = \left(1 - \frac{1}{\eta_T}\right)$$

where: $V_{Z\ autoX}$, $V_{Z\ Xfm}$ are auto-transformer and transformer short circuits voltage respectively.

A second autotransformers disadvantage concerns the galvanic (electrical non-isolated) connection of the primary and secondary circuits, due to which, all disturbances, over-voltages, etc. are transmitted directly through conduction between the input and output sides.

Advantages of the auto-transformer

- A saving in winding material (less copper or aluminium), since the secondary winding is part of the primary. Smaller volume, hence lower weight.
- Lower copper loss, lower I^2R losses, therefore efficiency is higher than in the two winding transformer.
- Lower leakage reactances, lower magnetising current.
- Variable output voltage can be obtained.
- Lower %voltage regulation.

Disadvantages of the auto-transformer

- There is a direct electrical connection between the primary and secondary sides. No electrical isolation.
- Should an open-circuit develop across the common winding portion, the full supply voltage is applied to the secondary.
- The short-circuit current is much larger than for the normal two-winding transformer.

Autotransformers are used in electromagnetic systems for connecting networks with different voltage levels, in start-up systems for large squirrel-cage induction motors, and where primary and secondary circuits with galvanic non-isolation is permissible, and where lower weight and losses out weigh the expenditure associated with limiting the short circuit current.

The Korndorfer system presented in chapter 13.4.6ii, is a frequently applied solution when starting up asynchronous motors. The start-up takes place in two stages without voltage-free interruptions.

A variable autotransformer is known as a variac. A variac is a single coil with a sweeping arm for the tap-off, which allows the ratio of primary turns:secondary turns to be readily altered.

26.8 Appendix: Soft ferrite general technical data

Tensile strength	20	N/mm^2
Flexural strength	40	N/mm^2
Resistance to compression	100	N/mm^2
Vickers hardness HV_{15}	8000	N/mm^2
Modulus of elasticity	150,000	N/mm^2
Breakage modulus	80-120	N/mm^2
Thermal conductivity	4 - 7× 10^{-3}	J/mm s K (W/mm/K)
Linear expansion coefficient	7-10×10^{-6}	/K
Specific heat	0.75	J/g K
Density	4 - 5	g/cm^3 (4 per cent Si, 7.63 g/cm^3)
Porosity	5	%
Poisson ratio	0.28	pu

Resistivity ρ	$\lambda_s = \frac{\Delta\ell^*}{\ell}$	$\varepsilon_r^{\P} / \rho^{\S}(pu)$				
Ω cm	×10^{-16}	10 kHz	100 kHz	1 MHz	100 MHz	300 MHz
10^5	-18	30/1	15/1	12/1	11/0.97	11/0.95
1	-1.5	140×10^3/1	50×10^3/0.95	30×10^3/0.65	-	-

* Magnetostriction, at saturation, contraction.
¶ Dielectric constant, $\varepsilon_r \rightarrow$ 10-20 at high frequency.
§ Resistivity normalised at low frequency.

26.9 Appendix: Technical data for a ferrite applicable to power applications

Symbol	Unit	Test condition	#1	#2
μ_i		25°C	2500±20%	2200±25%
$\hat{B}$	T T	25°C 100°C	0.48 0.37	0.54 0.45
H_s	A/m	B_s, 25°C	B_s = 0.4T : 200	B_s = 0.5T : 200
H	A/m		1600	
H_c/B_r	A/m /T	25°C 100°C	12/0.18 9.6/0.11	13/0.17 6.5/0.06
α/β	%/°C	Rev temp coeff	-0.11/-0.2	-0.11/-0.2
T_c	°C		> 200	> 250
ρ	Ω cm		100	700
η_B^*	mT^{-1} × 10^{-6}	10 kHz	0.9	
Density	g/cm^3		4.8	4.9
f_c	MHz	25°C	1.8	1.6

* - Maximum hysteresis coefficient
10 G (gauss) = 1 mT (milliTesla)
10 *e* = 80A/m

26.10 Appendix: Technical data for Iron, nickel, and cobalt applicable to power applications

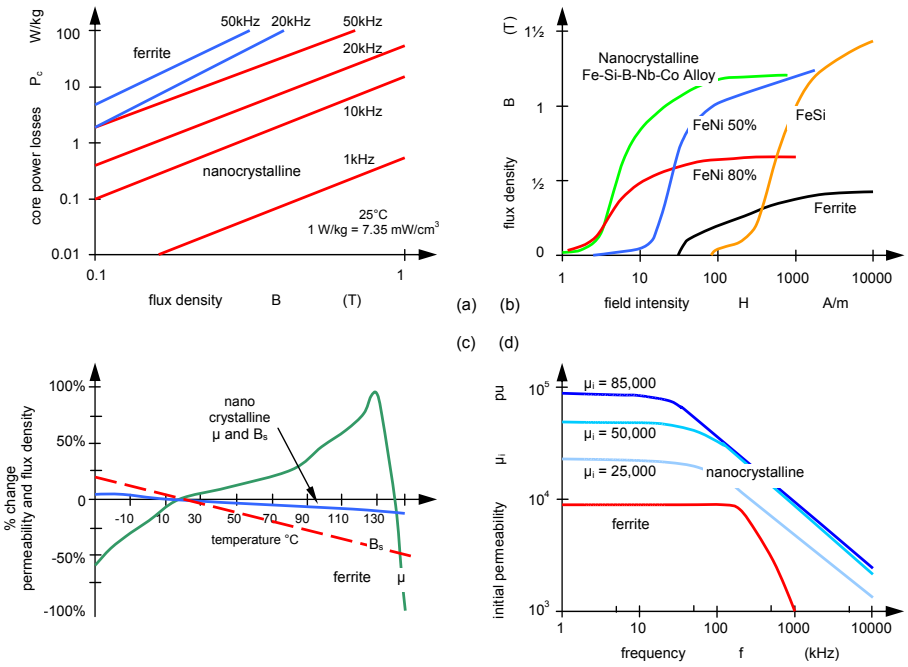


Figure 26.28. *Typical Fe-Si nanocrystalline material characteristics: (a) core losses; (b) B-H curves; (c) temperature dependence; and (d) permeability dependence on frequency.*

alloy	grain size	saturation flux density	saturation magnetostriction	coercivity	initial permeability	electrical resistivity	core losses	ribbon thickness
	d	B_s	λ_s	H_c	μ_i	ρ	P_c	t
	nm	T	10^{-6}	A/m		μΩcm	W/kg	μm
					@ 1 kHz		0.2T 100kHz	
$Fe_{73.5}Si_{15.5}$	14	1.23	0	0.4	100,000	115	35	21
$Fe_{84}Nb_7B_9$	9	1.49	0.1	8	22,000	58	76	22
$Fe_{86}Cu_1Zr_7B_6$	10	1.52	0	3.2	48,000	56	116	20
$Fe_{91}Zr_7B_3$	17	1.63	-1.1	5.6	22,000	44	80	18
$Co_{68}Fe_4(MoSiB)_{28}$	amorphous	0.55	0	0.3	150,000	135	35	23
$Fe_{76}(SiB)_{24}$	amorphous	1.45	32	3	8,000	135	50	23
80% Ni-Fe	100,000	0.75	1	0.5	100,000	55	90	50
50% Ni-Fe	100,000	1.55	25	5	40,000	45	200	70

26.11 Appendix: Cylindrical inductor design

Figures 26.29a and b show cross-sectional views of single-layer and multi-layer cylindrical inductors. The inductance of a single-layer cylindrical inductor (all dimensions are in mm) is given by

$$L = \frac{\mu_{eff}\, r^2 N^2}{228.6r + 254\ell} \quad (\mu H) \qquad (26.110)$$

while if the insulation spacing between the single layer turns are accounted for, inductance is given by

$$L = \frac{\frac{2}{3}\mu_{eff} N^{1.3} \left(d + d_w\right)^{1.7}}{\left(d_w + S\right)^{0.7}} \quad (nH) \qquad (26.111)$$

where the bare wire diameter is d_w and S is the spacing between turns, all in mm.

For the multi-layer cylindrical inductor shown in figure 26.29b, inductance is given by

$$L = \frac{\mu_{eff}\, r^2 N^2}{152.4r + 228.6\ell + 254b} \quad (\mu H) \qquad (26.112)$$

Figure 26.29d shows a family of curves used to give the effective permeability from the former l/d ratio and the core material permeability. These curves are applicable to the single-layer inductor but are a fair approximation of the multi-layer inductor. The winding is assumed to be closely wound over 95 per cent of the core length.

The inductance of a flat spiral air-core coil as shown in figure 26.29c is virtually independent of any axial core and is given by

$$L = \frac{r^2 N^2}{203r + 279b} \quad (mH) \qquad (26.113)$$

For inductance levels below 100 μH, an air core strip wound inductor as shown in figure 26.29e, has an inductance approximated by

$$L = \frac{r^2 N^2}{225r + 250\ell + 250b + 82.5\frac{lb}{r}\left(\frac{\ell + 2r}{\ell + 4r}\right)} \quad (\mu H) \qquad (26.114)$$

A toroidal core with a circular cross section has inductance given by

$$L = \mu_o \mu_{eff} \frac{r^2 N^2}{D} \quad (H) \qquad (26.115)$$

where r is the radius of the coil winding and D is the overall diameter of the toroid, all in metres.

Inductor design using these equations may require an iterative solution. Always attempt to maximise the winding surface area ($S_A \approx \pi\left(d + 2b\right)\ell$) for better cooling.

Example 26.8: *Wound strip air core inductor*

An air core inductance of 50 μH is made as a wound strip of copper, 40 mm wide and 1.5 mm thick. For cooling purposes, ½ mm spacing is used between each turn with an inner diameter of 60 mm and an outer diameter of 160 mm as physical constraints: Can the required inductance be achieved?

Solution

First calculate the parameters shown in figure 26.29e.

$$r = \tfrac{1}{4}\left(d_o + d_i\right) = \tfrac{1}{4} \times \left(160 + 60\right) = 55 \text{ mm}$$

$$b = \tfrac{1}{4}\left(d_o - d_i\right) = \tfrac{1}{4} \times \left(160 - 60\right) = 25 \text{ mm}$$

$$\ell = 40 \text{ mm}$$

$$N = \frac{b}{t_{Cu} + t_{air}} = \frac{50}{1.5 + 0.5} = 25 \text{ turns}$$

Substitution of the appropriate parameters values into equation (26.114) yields L = 51.6 μH.

♣

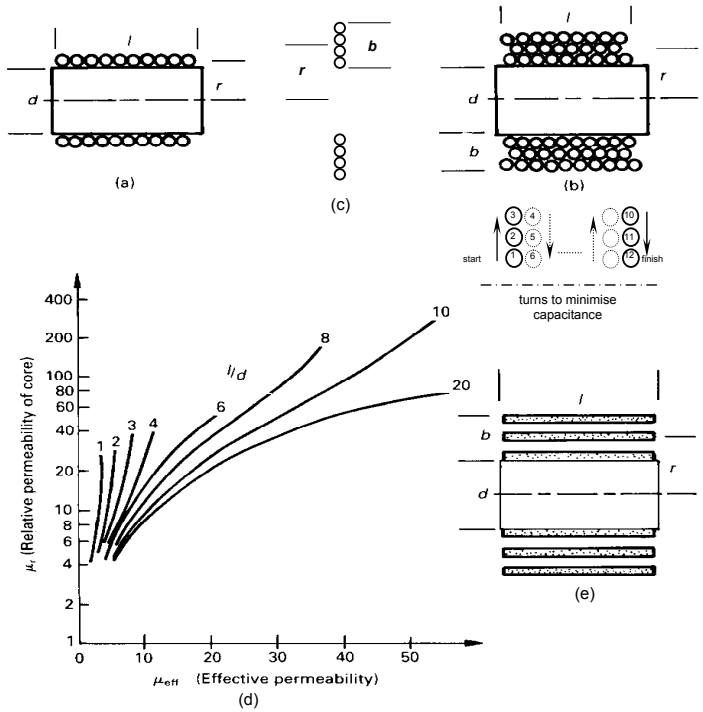


Figure 26.29. *Cylindrical inductors: (a) single-layer coil; (b) multi-layer coil; (c) coil layer; (d) effective permeability for different aspect ratios, l/d; and (e) core strip wound air core inductor.*

Example 26.9: *Multi-layer air core inductor*

An air core inductor is to have the same dimensions as the inductor in example 26.8. The same conductor area (40 mm × 1.5 mm) but circular in cross-section and number of turns is to be used. Calculate the inductance. If a ferrite solid cylindrical core 42 mm long and 60 mm in diameter with a relative permeability of 25 is inserted, what will the inductance increase to?

Solution

From example 26.8

r = 55 mm $\quad$ ℓ = 40 mm

b = 50mm $\quad$ N = 25

Substitution of these parameter values into equation (26.112) yields 62.5 μH.

From figure 26.29c, ℓ / d = 40/60 = 0.66, whence $\mu_{eff} \approx 3$. That is, with a cylindrical core inserted, a three fold increase in inductance would be expected (188 μH).

The use of end-caps and an outer magnetic sleeve would increase inductance, but importantly also help to contain the external magnetic field.

♣

26.12 Appendix: Copper wire design data

Nominal wire diameter d	Outer diameter enamelled grade 2	Approximate dc resistance at 20°C	Bare copper weight	Fusing current
mm	mm	Ω/m	gm/m	A
0.1	0.129	2.195	0.070	2.5
0.2	0.245	0.5488	0.279	7
0.376	0.462	0.136	1.117	18
0.5	0.569	8.781×10^{-2}	1.746	27.5
0.6	0.674	6.098×10^{-2}	2.50	36
0.8	0.885	3.430×10^{-2}	4.469	57
0.95	1.041	2.432×10^{-2}	6.301	79
1	1.093	2.195×10^{-2}	6.982	82
1.5	1.608	9.67×10^{-3}	15.71	145
2	2.120	5.44×10^{-3}	27.93	225
2.5	2.631	3.48×10^{-3}	43.64	310
3	3.142	2.42×10^{-3}	62.84	>
4	4.160	1.36×10^{-3}	111.7	>
4.5	4.668	1.08×10^{-3}	141.4	>
5.0	5.177	8.70×10^{-4}	174.6	>

26.13 Appendix: Minimisation of stray inductance

In many circuit layouts, it is essential to minimise stray and residual inductance. With high *di/dt* currents during switching, large voltages occur ($v = L\ di/dt$) which may impress excessive stresses on devices and components. Stray inductance within a package reduces its usable voltage rating. Stray inductance in the drain circuit of the MOSFET, within the package as shown in figure 4.11, reduces the usable voltage rail while source inductance increases the transient gate voltage. In the case of capacitors, residual inductance reduces the effectiveness of turn-off snubbers and can result in an unintentional resonant circuit.

Inductance of a straight wire of length ℓ and radius r is

$$L = \frac{\mu_o \ell}{2\pi}\left(\ell n\frac{2\ell}{r} - \tfrac{3}{4}\right) \quad \text{(H)} \quad \text{or} \quad L = 0.2\ell\left(\ell n\frac{2\ell}{r} - \tfrac{3}{4}\right) \quad (\mu\text{H}) \qquad (26.116)$$

which as a rule of thumb is about 1μH/m.

26.13.1 Reduction in wiring residual inductance

Wiring inductance can be decreased by cancelling magnetic fields in a number of ways

- coaxial cable
- parallel plates
- parallel wiring conductors.

In each case, the go and return paths are made parallel and physically close. Figure 26.30 shows the per unit length inductance for each wiring method.

coaxial cable

Minimum inductance results with coaxial cable, which is available for power application. The per unit inductance and capacitance are given by

$$L = \frac{\mu_o \mu_r}{2\pi}\ell n\frac{r_o}{r_i} \quad \text{(H/m)}$$
$$C = 2\pi\varepsilon_o\varepsilon_r / \ell n\frac{r_o}{r_i} \quad \text{(F/m)} \qquad (26.117)$$

where r_i is the inner radius and r_o the outer radius ($r_i < r_o$).

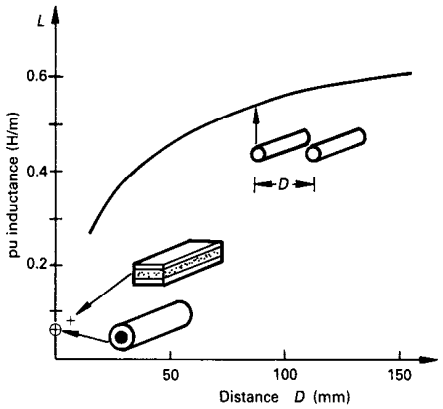


Figure 26.30. *Relative inductance of go and return wiring conductors.*

parallel plates

Very low inductance can be achieved by using parallel conducting copper plates separated by a thin insulation layer ($\mu_r \approx 1$). The inductance per unit length, neglecting skin effects, is approximated by

$$L = \mu_o \frac{d}{w} \quad \text{(H/m)} \quad (w >> d) \qquad (26.118)$$

where d is the separation of the plates and
w is the plate width.
The parallel plate capacitance is

$$C = \varepsilon_o\varepsilon_r\ w/d \quad \text{(F/m)} \qquad (26.119)$$

A complete analysis of the laminated parallel bus bar configuration is presented in appendix 26.14

parallel wiring conductors

For parallel wiring cylindrical conductors of radius r and separation D, in air,

$$L_{lo-freq} = \frac{\mu_o}{\pi}\ell n\left(\frac{D}{r}\right) \qquad L_{hi-freq} = \frac{\mu_o}{2\pi}\cosh^{-1}\left(\frac{D^2}{2r^2} - 1\right) \quad \text{(H/m)} \quad (D > 2r)$$
$$C = \frac{2\pi\varepsilon_o}{\ell n\left(\frac{D}{r}\right)} \quad \text{(F/m)} \qquad (26.120)$$

When the separation D is small over a long distance, ℓ, that is $D/\ell << 1$, the inductance, mutual coupling inductance, and capacitance, are

$$L = \frac{\mu_o}{\pi}\left\{\ell n\left(\frac{D}{r}\right) - \frac{D}{\ell} + \tfrac{1}{4}\right\} \quad \text{(H/m)}$$
$$M = \frac{\mu_o}{2\pi}\left\{\ell n\left(\frac{D}{r}\right) - \frac{D}{\ell} - 1\right\} \quad \text{(H/m)} \qquad (26.121)$$
$$C = \frac{\pi\,\varepsilon_o}{\cosh^{-1}\left(\frac{D}{2r}\right)} \quad \text{(F/m)}$$

At high frequencies, the inductance per unit length of a pair of parallel wires is better estimated by

$$L = \frac{\mu_o}{2\pi}\cosh^{-1}\left(\frac{D^2}{2r^2}\right) \quad \text{(H/m)} \qquad (26.122)$$

parallel wiring conductors over a conducting ground plane

The self inductance and mutual inductance between two conductors height h over a ground plane carrying the return current are given by

$$L_{lo-freq} = \frac{\mu_o}{2\pi}\ell n\frac{2h}{r} \qquad L_{hi-freq} = \frac{\mu_o}{4\pi}\cosh^{-1}\left(\frac{2h^2}{r^2}-1\right) \qquad \text{(H/m)}$$

$$M = \frac{\mu_o}{4\pi}\ell n\left(1+\left(\frac{h}{r}\right)^2\right) \qquad \text{(H/m)} \qquad (26.123)$$

$$C = \tfrac{1}{2}\frac{2\pi\varepsilon_o}{\ell n\left(\frac{D}{r}\frac{2}{\sqrt{D^2+4^2}}\right)} \qquad \text{(F/m)}$$

At high frequencies, the inductance per unit length of a wire parallel to a conducting wall is better estimated by

$$L = \frac{\mu_o}{4\pi}\cosh^{-1}\left(\frac{2D^2}{r^2}\right) \qquad \text{(H/m)} \qquad D > r \qquad (26.124)$$

Figure 26.30 shows that go and return power cable residual inductance decreases as separation decreases. Physical and mechanical constraints may dictate which wiring technique is most viable. All other wiring should cross perpendicularly, in order to minimise coupling effects.

Inductance of other conductor profiles

The self-inductance of a rectangular conductor, not associated with a return path in close proximity is

$$L = \frac{\mu_o}{2\pi}\left[\ell n\left(\frac{2\ell}{w+t}\right)+\tfrac{1}{2}+\frac{2}{9}\frac{w+t}{\ell}\right] \qquad \text{(H/m)} \qquad (26.125)$$

When the bus bar and its return path are side-by-side in the same plane

$$L = \frac{\mu_o}{2\pi}\left[\ell n\left(\frac{D}{w+t}\right)+\frac{3}{2}\right] \qquad \text{(H/m)}$$

$$M = \frac{\mu_o}{2\pi}\left[\ell n\left(\frac{2\ell}{D}\right)-1+\frac{D}{\ell}\right] \qquad \text{(H)} \qquad (26.126)$$

or long cylindrical wire and its return path are side-by-side in the same plane

$$L = \frac{\mu_{wire}}{8\pi}+\frac{\mu}{2\pi}\ell n\frac{D}{r} \qquad \text{(H/m)} \qquad (26.127)$$

where w is the width of the conductors
t is the thickness of the conductors
D is the distance between the midpoints of the conductors
ℓ is the conductor length.

The first component in equation (26.127) is the internal self-inductance component, which for copper and aluminium, $\mu_{wire} = \mu_o$, gives 50nH per metre.

26.13.2 Reduction in component residual inductance

26.13.2i - Capacitors

The inductance of a cylindrical capacitor winding, employing extended foils and scooping connections is given by

$$L = \frac{\mu_o}{2\pi}\left[\ell n\frac{2b}{r}-\tfrac{3}{4}\right] \qquad \text{(H)} \qquad (26.128)$$

where b is the length of the cylinder winding and $2r$ is its diameter. This equation shows that (undesirable) inductance is decreased by decreasing the length and by increasing the diameter.

26.13.2ii - *Capacitors - parallel connected*

Capacitors are extensively parallel connected, by manufacturers before potting, or by the user after potting, in order to increase capacitance. The low inductance feature of an extended foil, scoop connected capacitor can be obliterated by poor lead connection. Consider the parallel-connected capacitors shown in figure 26.31a, which shows the relative residual wiring inductance for three connections. Minimum inductance results when using a thin, double-sided copper printed circuit board arrangement, such that connections alternated between the top and bottom copper layers (go and return conductors). Cut-outs in the pcb, for the capacitors to fit into, only marginally decrease the inductance.

26.13.2iii - *Transformers*

A current balancing transformer may be used to equalise the principal currents of two parallel-connected power devices, as shown in figure 10.8. Conventionally each coil is wound on separate legs of the core, resulting in a large leakage inductance. This large leakage inductance can result in high voltage transients, which are to be avoided.
Leakage can be significantly decreased if two coils are bifilar wound on each limb and connected as shown in figure 26.31b. The same leakage flux cancelling technique can be used on the centre-tapped, push-pull transformer for the switch mode power supply shown in figure 15.16a. Because of the close proximity of bifilar wound conductors, high inter-winding capacitance and high dielectric fields may be experienced.

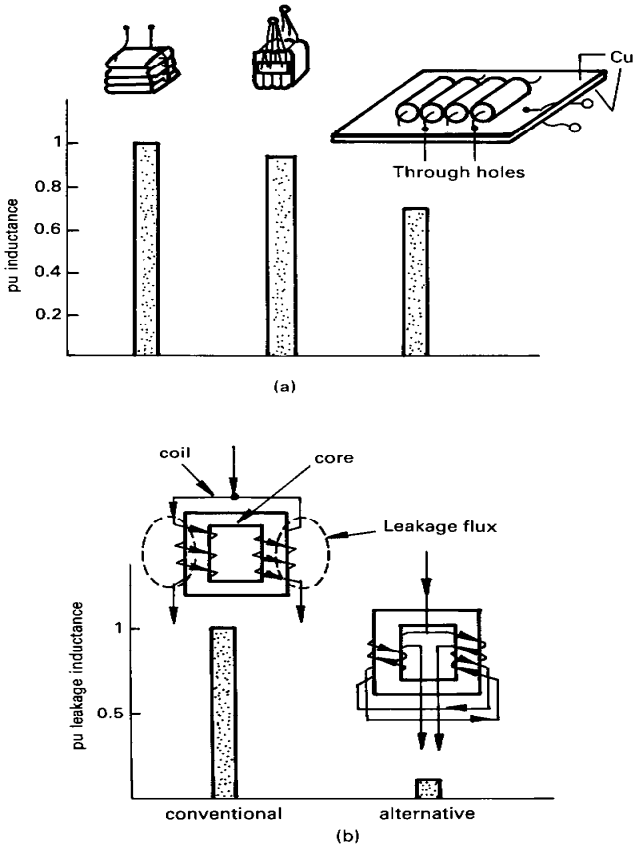


Figure 26.31. *(a) Parallel connected capacitors, inductance and (b) leakage inductance of a current balancing transformer.*

26.14 Appendix: Laminated bus bar design

As shown in figure 26.30, the use of a parallel laminated bus bar arrangement shown in figure 26.32a for go and return paths, results in a low inductance loop. If the gap between the bus bars is laminated with a dielectric material (e.g., polyester, ε_r = 3.5, $10^{18}\Omega$ insulation resistance, and a dielectric strength of 300kV/mm), distributed capacitance properties are gained. Up to five layers are available.
A laminated bus bar arrangement offers the following electrical mechanical and economic features in mitigation to the increased component costs:

- high packing density with good shielding
- better conductor cooling and thermal distribution because of flat surface area
- low voltage drop - low impedance
- high voltage and current capability
- modular, reliable, and eliminates wiring errors
- space saving, high packing density, low profile, low weight, and high mechanical strength
- increased capacitance for better noise suppression
- low inductance because thin parallel conductors allow flux cancellation
- low system costs, easy but low service costs, low installation costs
- applicable to voltages up to and in excess of 1kV

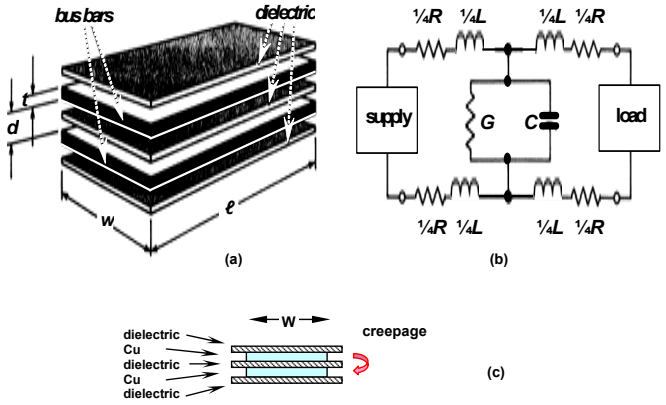


Figure 26.32. *Laminated bus bar:*
(a) parallel planar construction; (b) equivalent circuit distributed components; and (c) cross section showing creepage paths between copper conductors.

The physical bus bar dimensions determine the electrical parameters and characteristics. The two level bus bar comprises two parallel conducting plates of aluminium, brass or copper with resistivity σ, separated by a dielectric, with dielectric constant ε_r, and permeability μ_o, giving a conductance *G*, capacitance *C*, and resistance *R* that are uniformly distributed along the bus, as shown in the model in figure 26.32b.

Capacitance, C

The capacitance *C* is given by

$$C = \varepsilon_o \varepsilon_r \frac{w}{d} \quad \text{(F/m)} \tag{26.129}$$

where *w* is the width of the conductors
d is the distance between the bars, which is the dielectric thickness.
An increase in capacitance decreases the characteristic impedance, $Z_o = \sqrt{L/C}$. Lower impedance gives greater effective signal suppression and noise elimination. This is achieved with

- a smaller bar separation, *d*,
- a higher permittivity dielectric material, ε_r,
- wider conductors, *w*.

Shunt conductance, G

The shunt conductance *G* depends on the quality of the dielectric, specifically its conductivity at the operating frequency and temperature.

$$G = \frac{1}{\sigma}\frac{w}{d} \quad \text{(℧/m)} \tag{26.130}$$

Skin effect

Both the resistance and inductance are affected by the ac skin effect, which is frequency dependant. This was briefly treated in section 26.3.4ii, and specifically equation (26.50). The skin effect is when at high frequencies the current tends to flow on the surface of the conductor. The skin depth *δ*, from equation (26.50) is

$$\delta(f) = \sqrt{\rho/\mu_o \pi f} \tag{26.131}$$

where *ρ* is the resistivity of the conductor at frequency *f* and a given temperature.
The skin depth for copper and brass are

$$\delta_{Cu} = \frac{0.066}{\sqrt{f}} \quad \text{(m)} \qquad \delta_{Brass} = \frac{0.126}{\sqrt{f}} \quad \text{(m)} \tag{26.132}$$

As the frequency increases *L* decreases and *R* increases.

Inductance, L

There two inductive components,

- L_{int} - inside the conductor due to internal flux linkages,
- L_{ext} - external inductance between the two conductors due to the orientation of the two conductors carrying current.

In power applications and at the associated frequencies, the external inductance is more dominant.

$$L_{ext} = \mu_o \frac{d}{w} \quad \text{(H/m)} \tag{26.133}$$

At high frequency (taking the skin effect into account) the effective inductance is

$$L_e = \mu_o \frac{d+\delta}{w} \quad \text{(H/m)} \tag{26.134}$$

Thus to decreased inductance

- decrease the dielectric thickness, *d*
- increase the conductor width, *w*
- decrease the skin depth *δ* by using a conductor of lower resistivity.

Resistance, R

The dc resistance R_{dc} of the two conductors is

$$R_{dc}(20°\text{C}) = 2\rho\frac{1}{wt} \quad \text{(Ω/m)} \tag{26.135}$$

where *t* is the thickness of the conductors
The resistivity of copper and brass at 20°C are 1.7×10^{-8} and 7.0×10^{-8} Ωm, respectively.
The temperature effects on resistance for copper are accounted for by

$$R_{T_o} = R_{20°C}\left(1 + 0.0043\times(T_o - 20°C)\right) \quad \text{(Ω)} \tag{26.136}$$

At high frequency, taking the skin effect into account, assuming that the conductor thickness is at least twice the skin depth,

$$R_{ac} = 2\rho\frac{2}{\delta w} \quad \text{(Ω/m)} \tag{26.137}$$

Characteristic impedance, Z

The characteristic impedance *Z* of the go-and-return bus bar arrangement is given by

$$Z = \sqrt{\frac{R + j\omega L}{G + j\omega C}} \quad \text{(Ω)} \tag{26.138}$$

When the conductor resistance *R* and insulation conductance *G* are negligible

$$Z = \sqrt{\frac{L}{C}} \quad \text{(Ω)} \tag{26.139}$$

This equation illustrates that increasing the capacitance and decreasing the inductance reduces bus bar noise problems. Common to decreased inductance and increased capacitance are

- decrease the dielectric thickness *d* and
- increase the bus bar width *w*
- increase permittivity and decrease permeability

That is, characteristic impedance *Z* is proportional to *d/w*.

Figure 26.33 compares the electrical parameters obtained for one metre of twist pair plastic coated 1mm diameter solid copper wire and one metre of laminated bus bar. The copper cross section area is the same in each case, giving a dc resistance of 44mΩ per metre. The most significant electrical factor is the reduction in inductance when a bus bar arrangement is used. Better electrical parameters are gained for a significant cost increase.

The propagation delay time is approximately

$$t_{pd} = 0.04 \times C \times Z \quad \text{(ps/mm)} \qquad (26.140)$$

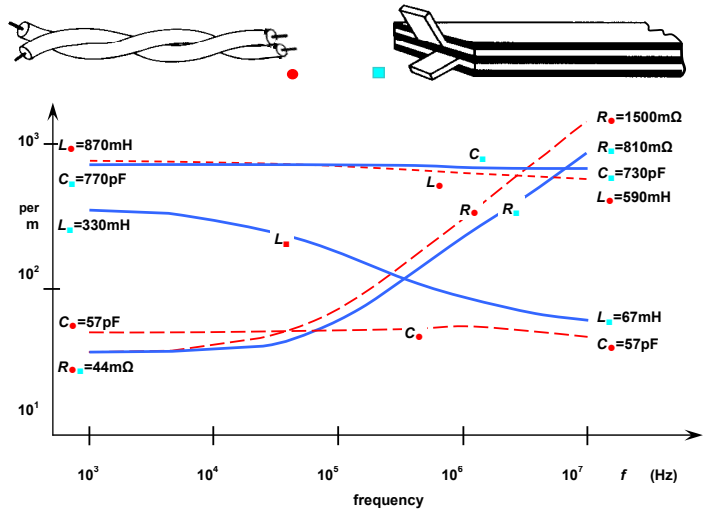


Figure 26.33. *Comparison between electrical parameters for a twisted pair and a laminated bus bar, each with the same copper cross sectional area, that is, the same dc resistance.*

26.15 Appendix: Insulating material for between bus bar conductors

If the busbar is to be edge filled, epoxy glass is recommended for the top and bottom insulators. If it is to be sealed by pinching off the insulation, Nomex type materials are recommended.

Material	Minimum thickness	K-factor	Dielectric strength
	mil	pu	kV/mil
Mylar	2	3.3	7.5
Epoxy glass	2.5	4.3	0.5
Kapton	1	3.8	4.6
Nomex	3	2.6	0.5

26.16 Appendix: Materials by types of magnetization

Diamagnetic: Ordering magnetic particles are electron pairs or magnetic nuclei. They are ordered in one direction under external magnetic field. Materials exhibit magnetization opposing to the applied magnetic field. Magnetic field is weakened in a material. Weak repelling effect to the magnetic field. The value of susceptibility is independent of temperature. Magnetization disappears when the field is removed. *Example* - most objects around - wood.

Paramagnetic: Ordering magnetic particles are unpaired electrons (spins). The magnetic moments tend to be randomly orientated due to thermal fluctuations when there is no magnetic field. They are ordered in one direction under external magnetic field such that magnetisation of the material is proportional to the applied field. Materials exhibit magnetization reinforcing (parallel, therefore - para) to the applied magnetic field. Magnetic field is strengthened in a material. Weak attracting effect to the magnetic field. Magnetization disappears when the field is removed. *Example* - aluminium metal (Al).

Ferromagnetic: Ordering magnetic particles are unpaired electrons (spins). They are ordered in one direction under external magnetic field. Below Curie point, ferromagnets exhibit magnetization without of any external magnetic field (spontaneous magnetization) due to self-ordering of spins. Exhibit reinforcing magnetization to the applied magnetic field that usually remains when the field is removed (field-induced magnetization). Magnetic field is strengthened in a material. Strong attracting effect to the magnetic field. In the periodic table of elements only Fe, Co and Ni are ferromagnetic at and above room temperature. As ferromagnetic materials are heated then the thermal agitation of the atoms means that the degree of alignment of the atomic magnetic moments decreases and hence the saturation magnetisation also decreases. Eventually the thermal agitation becomes so great that the material becomes paramagnetic; the temperature of this transition is the Curie temperature, T_C (Fe: T_C =770°C, Co: T_C =1131°C and Ni: T_C =358°C). *Example* - iron metal (Fe).

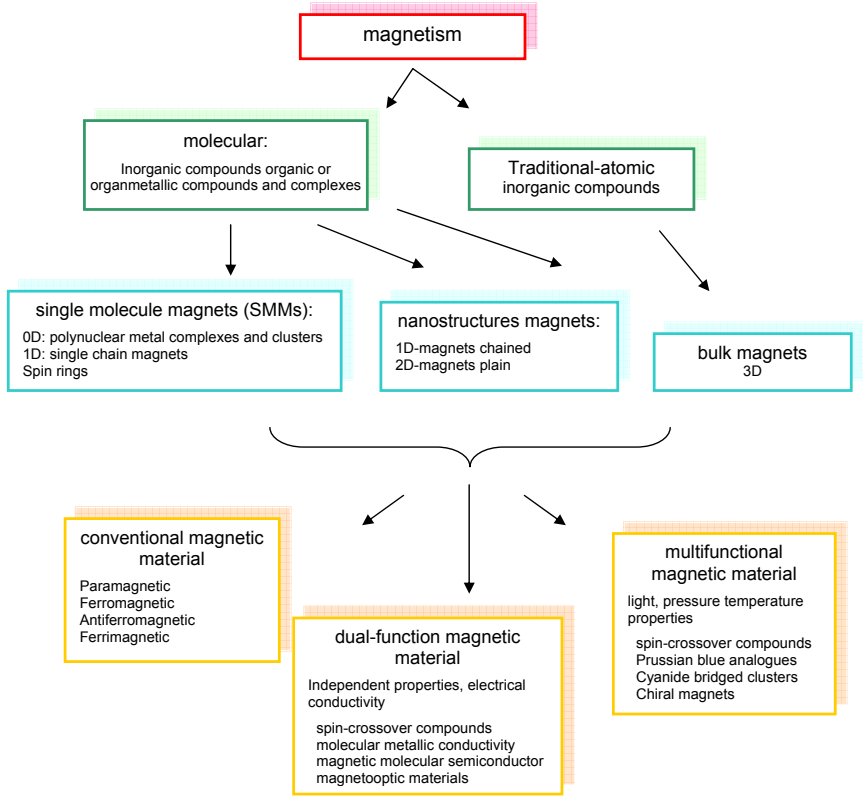


Figure 26.34. *Classification of magnetic materials.*

Antiferromagnetic: Ordering magnetic particles are unpaired electrons (spins). In contrast to paramagnetic and ferromagnetic, they are ordered in opposite directions in equal quantities under the action of magnetic field and since the field cancels out, the material appears to behave in the same way as a paramagnetic material. The ordering usually remains when the field is removed. Like ferromagnetic materials these materials become paramagnetic above a transition temperature, known as the Néel

temperature, T_C. (Cr: T_N = 37ºC). Below Neel point, self-ordering can be observed. Magnetic field is weakened in a material. Weak repelling effect to the magnetic field. In the periodic table the only element exhibiting antiferromagnetism at room temperature is chromium. *Example* - nickel oxide (Ni0).

Ferrimagnetic: Ferrimagnetism is only observed in compounds, which have more complex crystal structures than pure elements. Ordering magnetic particles are unpaired electrons (spins). They are ordered in opposite directions in unequal quantities/sizes under the action of magnetic field. Behave as weak ferromagnets. Magnetic field is strengthened in a material. Weak to strong attracting effect to the magnetic field. *Example* - magnetite (Fe_30_4).

Metamagnetic: Materials that can change their magnetic properties depending on the strength of applied magnetic field. For example, paramagnets transform to ferromagnets or vice versa when the field is increased or reduced.

Magnetic materials - structural types

Figure 26.34 classifies development in both inorganic and organic magnetic materials
Relatively recently, the science of so-called molecular magnetism branched out from traditional magnetism that explored mainly magnetic behaviour of compounds at the level of atoms (in metals). Molecular magnetism researches magnetic behaviour of molecules rather than atoms. Magnetic behaviour of organic compounds belongs to the field of molecular magnetism.
Molecules as well as atoms may form bulk magnets (traditional type of magnets), that are also known as 3D magnets. At the same time, two new types of molecular magnets recently evolved: single molecule magnets (SMMs) and nanostructured magnets.
SMMs are fully functional magnets on a molecular level. They possess some unique properties, such as quantum tunnelling of magnetization. SMMs include: polynuclear metal complexes and clusters, single-chain magnets, spin rings and some others.
Nanostructured magnets built of self-organized or self-assembled crystallites. They may possess many unique properties in addition to magnetic ones, (such as anisotropy, transparency, electrical conductance, porosity etc.). Low dimensionality (0D, 1D, 2D) is characteristic of self-aggregated molecular magnets.
Bulk molecular (and atomic) magnets usually contain materials that may be either paramagnetic, ferromagnetic, anti-ferromagnetic, or ferrimagnetic. Recently, a new type of molecular magnetic materials evolved, that may change magnetic properties under the action of different external factors, such as light, temperature, pressure, etc. They are known as multifunctional magnetic materials, which include so-called spin-crossover compounds, Prussian Blue analogues, cyanide-bridged clusters, chiral magnets.
The other type of new functional molecular magnetic materials are dual-function materials. They possess one or more independent physical properties in addition to magnetic properties, such as electrical conductivity, porosity, or optical properties. Magnetic molecular conductors are the most widely studied organic materials of this type.

Reading list

cda.org.uk (copper development agency)

EPCOS Ferrite design tool
www.epcos.com

McLyman, W. T., *Transformer and Inductor Design Handbook*,
Marcel Dekker Inc., 1978.

Snelling, E. C., *Soft Ferrites*,
CRC Press, Cleveland, Ohio, 1969.

Manufacturers Data Handbooks, Catalogues, and www

Siemens	Thomson CSF	Philips
SEI	Telmag	Neosid
Magnetics Inc (mag-inc.com)	Krystinel Corp	Stackpole
Ferroxcube Inc	Arnold Eng. Co.	Micrometals Co.
Pyroferric Inc	Fuji	Fair-rite

Hitachi Metals http://www.hitachi-metals.co.jp/e/

Vacuumschmelze http://www.vacuumschmelze.de/dynamic/en/

Mecagis http://www.mecagis.com/index.php

Magnetec http://www.magnetec.de/magnetec.htm

Problems

26.1. Rework example 26.7, taking B_r = 0.18 T into account.

26.2. Rework example 26.7, when the core temperature is 100°C.

26.3. Rework example 26.3 when
$V_e = 7.3 \times 10^{-6}$ m^3 A_e = 66 mm^2
ℓ_e = 110 mm c = 0.75 nH
What are the effects of decreasing the core volume for a given L and I?
[10A: 274 J/m^3, 0.53 mm, 1200 A/m, 13 turns; 20A: 2360 A/m, μ_e = 185]

26.4. Show that the maximum flux density for a square-wave-excited transformer is given by

$$\hat{B} = \frac{V}{4NAf}$$

26.5. A 2:1 step-down transformer with an effective area of 10 cm^2 is driven from a 240V, 1kHz square-wave source. The transformer has 240 primary turns and magnetising inductance of 10mH.
i. Calculate the maximum flux density.
ii. Calculate the peak primary current when the secondary is loaded with a 5 Ω resistor. Sketch the primary and secondary current waveforms.
[0.25 T, 30 A]

27

Hard Magnetic Materials - Permanent Magnets

All materials can be classified in terms of their magnetic behaviour, falling into one of five categories depending on their bulk magnetic susceptibility (permeability). The two most common types of magnetism are diamagnetism and paramagnetism, which account for the magnetic properties of most of the periodic table of elements at room temperature, as indicated in figure 27.1. The diamagnetic elements are usually referred to as non-magnetic, while those that are referred to as magnetic are classified as ferromagnetic. The only other type of magnetism observed in pure elements at room temperature is antiferromagnetism, Cr. Magnetic materials (as opposed to elements) can also be classified as ferrimagnetic although this is not observed in any pure element but is found in compounds, such as the mixed oxides, known as ferrites, from which ferrimagnetism derives its name. Magnetic susceptible falls into a particular range of values for each of the five material types as shown in Table 27.1, with some examples, with the five categories shown in figure 27.2.

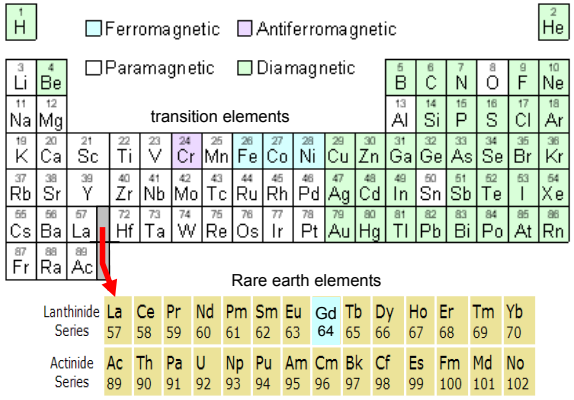
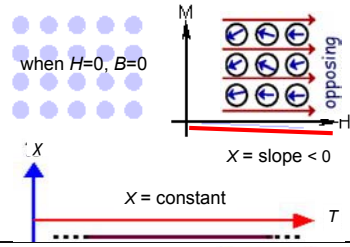
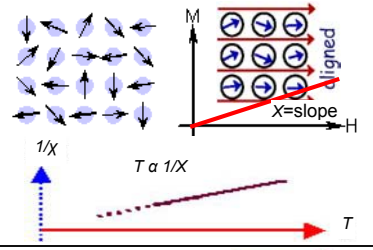
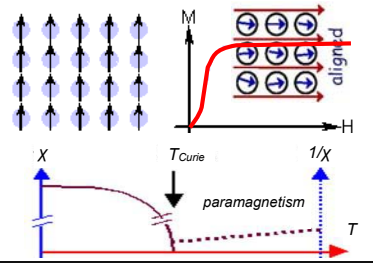
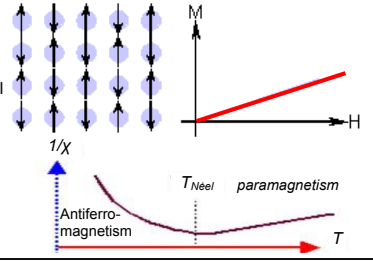
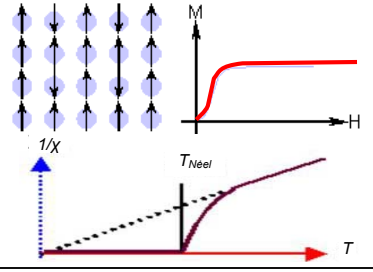


Figure 27.1. *Periodic table, classifying the type of magnetic behaviour of each element at room temperature.*

Table 27.1. Magnetic behaviour of five different types of materials ($\mu_r = \chi+1$, $B=\mu_o(H+M)=(1+\chi)\mu_o H=\mu_r\mu_o H$)

Type of Magnetism	Susceptibility	Atomic (H=0) / Magnetic Behaviour $M(H)$	Example/ Susceptibility χ	
Diamagnetism	Small and negative $\chi < 0$ M= χ H No temperature dependence	Atoms have no magnetic moment 	Bi Au Cu H_2O SiO_2 $CaCO_3$ Super Con-ductor	-166×10^{-6} -34.7×10^{-6} -9.63×10^{-6} -9.2×10^{-6} -6.2×10^{-6} -4.8×10^{-6} $\approx 10^5$
Paramagnetism	Small and positive $1 \gg \chi > 0$ M= χ H $1/\chi$(T) linear (weak)	Atoms have randomly oriented magnetic moments 	Aℓ β-Sn Pt Mn Air O_2	21.1×10^{-6} 22.7×10^{-6} 257×10^{-6} 904×10^{-6} 0.36×10^{-6} 2.1×10^{-6}
Ferromagnetism	Large and positive, function of applied field, microstructure dependent $\chi > 0$, $\lvert\chi\rvert \gg 1$	Atoms have parallel aligned magnetic moments 	Fe Ni Co SiFe	~80,000 7×10^4
Antiferromagnetism	Small and positive $\chi > 0$	Atoms have mixed parallel and anti-parallel aligned magnetic moments 	Cr FeMn NiO MnO	318×10^{-6}
Ferrimagnetism	Large and positive, function of applied field, microstructure dependent $\chi \gg 1$	Atoms have anti-parallel aligned magnetic moments 	Ba-ferrite Fe_3O_4	~3000 100

i. Diamagnetism

Diamagnetic metals have a weak, negative susceptibility to magnetic fields. In a diamagnetic material, there are no unpaired electrons and all the orbital shells are filled, hence no net magnetic moment. The intrinsic electron magnetic moments cannot produce any bulk effect, thus magnetization arises from electron orbital motion. Under the influence of an applied field, H, the spinning electrons precess and this motion, which is a type of electric current, produces a weak internal magnetisation, M $(=\chi H = (\mu_r - 1) \times H = J / \mu_o)$, that opposes the externally applied magnetic field, thus causing a repulsive effect. The cause of this interaction is in accordance with Lenz's Law by which small, localized currents generated in the material created magnetic fields in opposition to the applied changing field. All materials have a diamagnetic effect, which is often masked by a larger paramagnetic or ferromagnetic component. Although diamagnetism is generally a weak effect in most materials, superconductors exhibit a strong effect, thus repel magnetic fields from their bulk. The dimensionless volumetric susceptibility value χ ($\mu_r = \chi + 1$) is independent of temperature.

Most elements in the periodic table, including copper, silver, gold, and helium are diamagnetic. The strongest diamagnetic elements are bismuth and carbon graphite, as highlighted in Table 27.1.

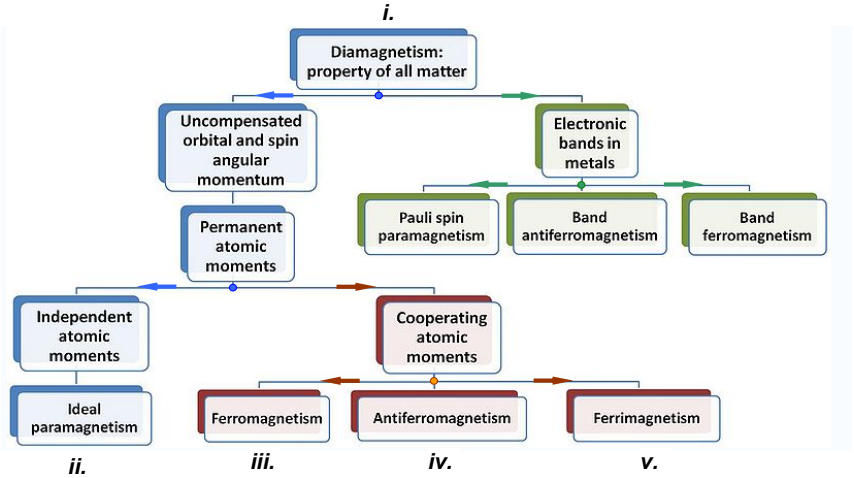


Figure 27.2. *Types of natural magnetism.*

ii. Paramagnetism

Paramagnetic metals have a small, positive susceptibility to magnetic fields. They become magnetized in the same direction as the applied magnetic field and the magnetization magnitude is proportional to the applied magnetic field. In a paramagnetic material there are unpaired electrons, specifically, atomic or molecular orbitals with exactly one electron in them. While paired electrons are required by the Pauli exclusion principle to have their intrinsic (spin) magnetic moments pointing in opposite directions, causing their magnetic fields to cancel out, an unpaired electron is free to align its magnetic moment in any direction. When an external magnetic field, H, is applied, these magnetic moments tend to align in the same direction as the applied field, thus reinforcing it.

Paramagnetic materials are attracted to magnetic fields, hence have a relative magnetic permeability slightly greater than one (that is, a positive magnetic susceptibility, χ). The force of attraction generated by the applied field is linear and weak. Unlike ferromagnets, paramagnets do not retain any magnetization in the absence of an externally applied magnetic field, because thermal motion causes the spins to return to a random orientation. Thus the total magnetization drops to zero when the applied field is removed. Even in the presence of a H field, the induced magnetization is small because only a small fraction of the spins are oriented by the field. This small fraction is linearly proportional to the field strength, H.

The randomly oriented moments result from thermal agitation. A magnetic field slightly aligns these moments whence low magnetisation results, aligned with the applied field. As the temperature increases, the thermal agitation increases and it becomes harder to align the atomic magnetic moments, hence the magnetic susceptibility, χ, decreases. This behaviour is known as the *Curie law*, as is shown in equation (27.1), where C is a material-specific constant called the Curie constant.

$$M = H \times \chi = H \times \frac{C}{T} \qquad \left(= H \times (\mu_r - 1)\right) \tag{27.1}$$

M is the resultant magnetism (with reference to the permeability of a vacuum), T is absolute temperature, K, and H is the externally produced magnetic field, A/m.

This law indicates that the susceptibility χ of paramagnetic materials is inversely proportional to temperature. Curie's law is only valid under conditions of low magnetisation, since it does not consider magnetisation saturation that occurs when the atomic dipoles are all aligned in parallel. After complete alignment, increasing the external field will not increase the total magnetisation since there can be no further alignment. However, such saturation typically requires extremely strong magnetic fields.

The Curie law is a special case of the more general Curie-Weiss law, equation (27.2), which incorporates a temperature constant, T_C, and derives from Weiss theory, proposed for ferromagnetic materials, that incorporates the interaction between magnetic moments.

$$\chi = \frac{C}{T - T_c} \tag{27.2}$$

In equation (27.2), T_C , in degrees Kelvin, K, can be positive, negative or zero. When T_C = 0, the Curie-Weiss law equates to the Curie law. A non-zero T_C indicates that there is interaction between neighbouring magnetic moments and the material is only paramagnetic above a certain transition temperature. If T_C is positive then the material is ferromagnetic below the transition temperature and the value of T_C corresponds to the transition temperature, the Curie temperature, T_C. If T_C is negative then the material is antiferromagnetic below the transition temperature, the Néel temperature, T_N, however the value of T_C does not relate to T_N. Equation (27.2) is only valid for a paramagnetic material state and is not valid for many metals, as the electrons contributing to the magnetic moment are not localised. However, the law does apply to some metals, for example, the rare-earth elements shown in figure 27.1, where the 4f electrons, that create the magnetic moment, are closely bound.

Paramagnetic materials include magnesium, molybdenum, lithium, sodium, and tantalum. Paramagnetic materials like aluminium, uranium and platinum become more magnetic when they are very cold.

iii. Ferromagnetism

Ferromagnetic materials have a large positive susceptibility to an external magnetic field. Ferro-magnetism is the term for the basic mechanism by which specific materials (such as iron) form permanent magnets and/or exhibit strong magnetic interaction with magnets.

A paramagnetic substance has unpaired electrons, but additionally ferromagnetism is only possible when atoms are arranged in a lattice and the atomic magnetic moments can interact to align parallel to each other. The effect is due to the presence of a molecular field within the ferromagnetic material, which is sufficient to magnetise the material to saturation. This creates a net internal magnetic field much greater than the applied field. Even when the applied external field is removed, the electrons in the material maintain each other orientated in the same direction.

Magnetic domains within the material are the regions of atomic magnetic moments that are aligned. The movement of these domains determines how the material responds to a magnetic field and consequently the susceptible is a function of the applied magnetic field. Therefore, ferromagnetic materials are usually compared in terms of saturation magnetisation (magnetisation when all domains are aligned) rather than susceptibility.

In the elemental periodic table, only the three consecutive elements Fe, Co and Ni are ferromagnetic at and above room temperature. As ferromagnetic materials are heated, thermal agitation of the atoms decreases the degree of alignment of the atomic magnetic moments, hence saturation magnetisation decreases. Eventually the thermally agitated disorder overwhelms the energy lowering due to ferromagnetic order and the material becomes paramagnetic; the temperature of this critical transition is the Curie temperature, T_C (Fe: T_C = 770°C, Co: T_C = 1131°C and Ni: T_C = 358°C). Above T_C, susceptibility varies according to equation (27.2).

iv. Antiferromagnetism

Antiferromagnetic materials are similar to ferromagnetic materials but the exchange interaction between valance electrons of neighbouring atoms leads to the anti-parallel alignment (as opposed to parallel alignment) of the atomic magnetic moments. Therefore the magnetic field cancels out and the material appears to behave like a paramagnetic material. It is difficult to magnetize such materials in the direction of the applied field but they still demonstrate a relative permeability slightly greater than 1. The magnetic susceptibility of an antiferromagnetic material is typically a maximum at the Néel temperature. Like ferromagnetic materials, these materials become paramagnetic above a transition temperature, the Néel temperature, T_N, where equation (27.2) is applicable. The Néel temperature is similar to the Curie temperature in ferromagnetism. In the periodic table, the only element exhibiting antiferromagnetism at near room temperature is chromium (Cr: T_N = 37°C).

Alloys such as iron manganese (FeMn), and oxides such as nickel oxide (NiO), manganese oxide, and iron oxide (FeO), exhibit antiferromagnetism.

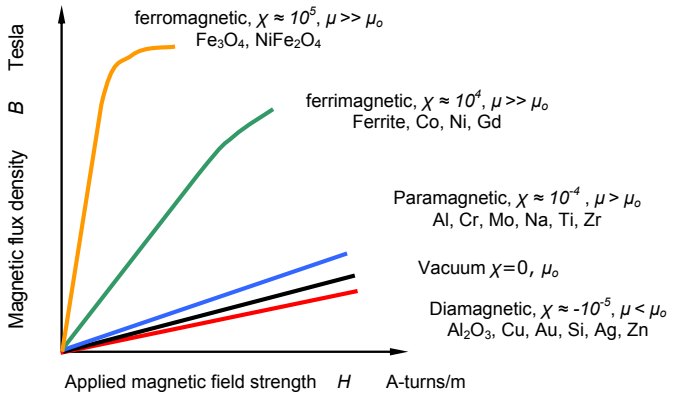


Figure 27.3. *Relative susceptibility χ (= μ_r - 1) of different materials.*

v. Ferrimagnetism

Ferrimagnetism is only observed in compounds that have more complex crystal structures than pure elements. Within such materials, the exchange interactions lead to parallel alignment of atoms in some of the crystal sites and anti-parallel alignment of others. The material devolves into magnetic domains, just like a ferromagnetic material and the magnetic behaviour is also similar, although ferrimagnetic materials usually have lower saturation magnetisations. In an optimal geometrical arrangement, there are more magnetic moments from the sublattices of electrons which point in one direction, than from the sublattices which points in the opposite direction. Both ferrimagnetic and ferromagnetic materials retains spontaneous magnetization below the Curie temperature, and show no magnetic order (are paramagnetic) above this temperature. However, there may be a temperature below the Curie temperature at which the two sub-lattices have equal (but opposite polarity) moments, resulting in a net magnetic moment of zero; this is called the *magnetization compensation point*.
Examples of ferrimagnetic materials are iron oxide (Fe_3O_4) and ferrite.

The different magnetic types are summarised in figure 27.3 in terms of relative susceptibility. There are other types of magnetism, such as spin glass, superparamagnetism, superdiamagnetism, and metamagnetism.

27.1 Magnetic properties

The intrinsic properties of a magnetic material are those properties that are characteristic of the material and are unaffected by the microstructure (for example, grain size or crystal orientation of grains). These properties include Curie temperature, saturation magnetisation, and magneto-crystalline anisotropy.

i. Saturation Magnetisation

Saturation magnetisation, M_S, is a measure of the maximum field that can be generated by a material. It depends on the strength of the dipole moments on the atoms that make up the material and how densely they are packed. The atomic dipole moment is affected by the nature of the atom and the overall electronic structure within the compound. The packing density of the atomic moments are determined by the crystal structure (that is, the spacing of the moments) and the presence of any non-magnetic elements within the structure.
For ferromagnetic materials, M_S also depends on how well the moments are aligned, as thermal vibration of the atoms causes moment misalignment and a reduction in M_S. For ferrimagnetic materials not all of the moments align parallel, even at zero degree Kelvin, hence M_S depends on the relative alignment of the moments as well as the temperature. In the case of a single magnetic domain, its saturation magnetisation is referred to as the spontaneous magnetisation.
Table 27.2 gives examples of the saturation polarisation, J_S, ($=\mu_o M_S$) and Curie temperature, T_C, of commonly used magnetic materials.

ii. Magnetic Anisotropy

In a crystalline magnetic material, magnetic properties vary depending on the crystallographic direction in which the magnetic dipoles are aligned.

A measure of *magnetocrystalline anisotropy* in the easy direction of magnetisation, is the anisotropy field, H_a, which is the field required to rotate all the moments by 90° in a saturated single crystal.

Temperature dependant anisotropy is caused by a coupling of the electron orbitals to the lattice, and in the easy direction of magnetisation, this coupling creates orbitals in the lowest energy state.
The easy direction of magnetisation for a permanent magnet, based on ferrite or the rare earth alloys, is uniaxial. However, it is also possible to have materials with multiple easy axes or where the easy direction can lie anywhere on a certain plane or cone surface. The fact that a permanent magnet has uniaxial anisotropy means that it is difficult to demagnetise since it is resistant to rotation from the magnetisation direction. When no preferred crystallographic direction exists within a material, shape anisotropy may arise if there are non-spherical particles present within the material. The long axis of such particles is the preferred axis of magnetization, as with Alnico magnets.

In addition to magnetocrystalline anisotropy, there is another effect related to spin-orbit coupling called stress anisotropy or *magnetostriction*. Magnetostriction (which is temperature dependant) arises from the strain dependence of the anisotropy constants. Upon magnetization, a previously demagnetized crystal experiences a strain that can be measured as a function of the applied field along the principal crystallographic axes. A magnetic material will therefore change its dimension when magnetized. The inverse affect or magnetization change with stress, also occurs. A uniaxial stress can produce a unique easy axis of magnetization if the stress is sufficient to overcome all other anisotropies.

The third type of anisotropy is due to the shape of a grain. A magnetized body will produce magnetic charges or poles at its surface. This surface charge distribution, acting in isolation, is itself a magnetic field source, and is a demagnetizing field, opposing the magnetization that produced it. Anisotropy is temperature dependent, decreasing and vanishing at the Currie temperature.

Table 27.2. The saturation polarisation, B_S, and Curie temperature, T_C, of a range of magnetic materials.

Material	Magnetic Structure	J_S at 298K $J_S = \mu_o M_S$	T_C Curie temperature	*reversible coefficient*	
				α (B_r)	β (H_{ci})
		Tesla	ºC	%/ºC	%/ºC
Fe	Ferro	2.15	770		
Co	Ferro	1.80	1121		
Ni	Ferro	0.62	368		
Alnico 5/8	Ferro	1.08/0.80	900/860	-0.02	-0.03
$Nd_2Fe_{14}B$	Ferro	1.59	312	-0.11	-0.060
$SmCo_5$	Ferro	1.14	720	-0.04	-0.3
Sm_2Co_{17}	Ferro	1.25	820	-0.03	-0.3
FeCrCo	Ferro	1.35	630	-0.03	-0.04
Fe, 3wt% Si	Ferro	2.00	740		
Fe, 35wt% Co	Ferro	2.43	940		
Fe, 78wt% Ni	Ferro	0.70	580		
Fe 50wt% Ni	Ferro	1.55	500		
$Ba0.6Fe_2O_3$	Ferri	0.48	450	-0.19	0.40
$Sr0.6Fe_2O_3$	Ferri	0.48	450	-0.19	0.40
$MnO\text{-}Fe_2O_3$	Ferri	0.51	300	-0.19	0.40

27.2 Classification of magnetic materials

Magnetic materials can be classified as either *hard* or *soft*.

Magnetically soft materials can be readily magnetized, but the magnetism induced is only temporary. An alloy of iron with 4% silicon is used to make the cores of electromagnets and transformers. Stroking a magnet along a steel pin from one end to the other will weakly magnetize the steel pin. This is because large numbers of iron atoms (domains) of the steel become aligned in the same direction. Such materials typically have an intrinsic coercivity of less than $1000Am^{-1}$ and are primarily used to enhance and/or channel the flux produced by an electric current. A parameter often used as a figure of merit for soft magnetic materials is the relative permeability, μ_r, where $\mu_r = B/\mu_o H$, which is a measure of how readily the material responds to an applied magnetic field, H. Other important parameters are coercivity, saturation magnetisation, and electrical conductivity, plus as small hysteresis loop area.

Magnetically hard materials can be permanently magnetized by a strong magnetic field. Steel and special alloys, which contain various amounts of aluminium, nickel, cobalt, and copper, are used to make permanent magnets. Its magnetic properties are due to its atomic structure. The electrons in the outer orbit of an iron atom behave as an electric charge and produce a strong magnetic field. In magnetized iron, millions of individual iron atoms, called a domain, are aligned in the same direction. Each domain has a north and a south (seeking) pole. Iron is an example of a natural hard magnetic material. The strongest energy permanent magnets are NdFeB, made in a vacuum and at high temperature from powders of various alloy grains.

Hard magnets, also referred to as permanent magnets, are of magnetic materials that retain their magnetism after being magnetised, when the inducing field energy source has been removed. Practically, this occurs with materials that have an intrinsic coercivity of greater than ~10kAm^{-1}. The hysteresis loop area should be large, corresponding to a high stored energy. Saturation, coercive force and remanence should also be large. Figure 27.4 shows this intrinsic coercivity, H_c, variation difference between hard and soft magnetic materials.
The development progress time-lines of hard magnetic materials is shown in figure 27.5.

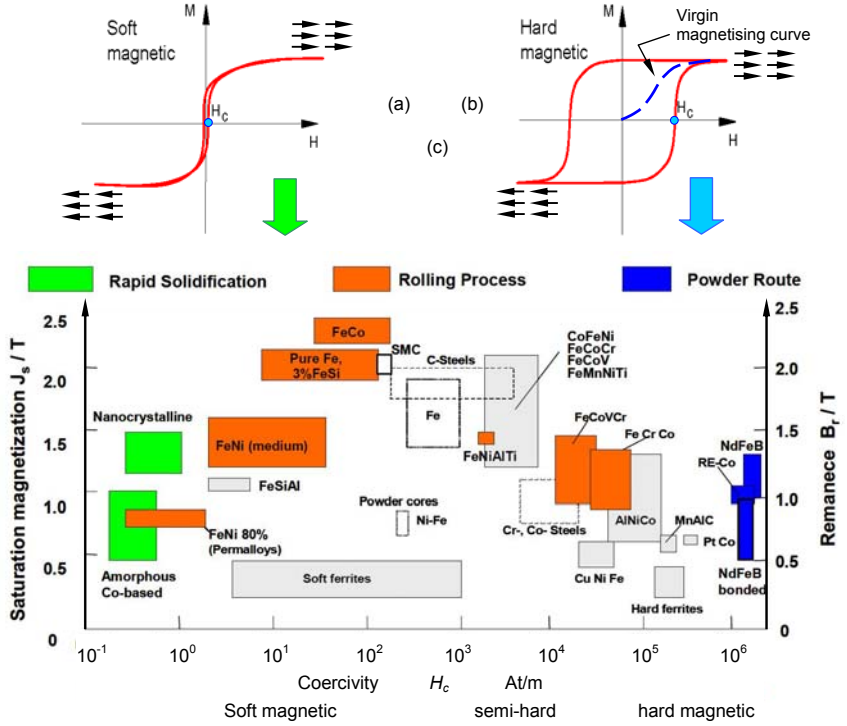


Figure 27.4. *Typical hysteresis loop and intrinsic coercivity H_c for:*
(a) soft, (b) hard magnetic materials, and (c) an overview of magnetic metals, alloys, and materials.

Hard magnetic materials can be divided into four categories: hard magnetic alloy material, hard magnetic ceramic material, bonded hard magnetic material, and flexible (or rubber) hard magnetic material, which in turn have the following subcategories:

1. Hard magnetic alloy material:
   - AlNiCoTi
   - FeCrCo
   - FeCoVCr (RE - rare earth)
   - RE Cobalt
   - RE FeB, REFeN
   - PtCo
   - CuNiFe

2. Hard magnetic ceramic material:
   Hard magnetic ceramic material ($M.nFe_2O_3$; M refers to Barium, Strontium and Plumbum; while *n* is within the range 4.5 to 6.5)

3. Bonded hard magnetic material:
   - bonded NiCoFeTi
   - bonded RE Cobalt
   - bonded NdFeB
   - bonded ferrite

4. Flexible hard magnetic material:
   Individually or mixed ferrite and neodymium magnetic powders

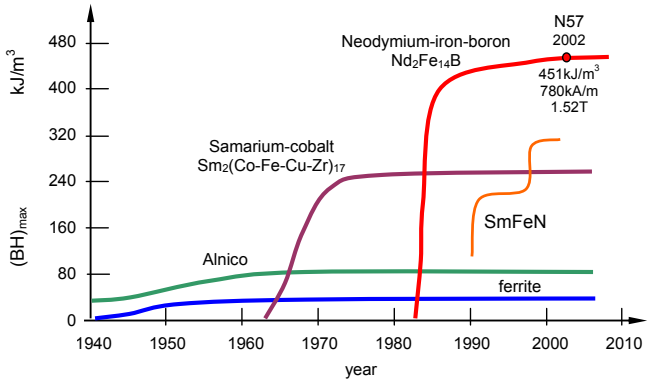


Figure 27.5. *The progressive development of permanent magnets.*

Magnet alloys that utilize elements from the Lanthanide series of the periodic table shown in figure 27.1 are commonly referred to as Rare Earth (RE) magnets. Specific magnetic materials containing rare earth elements, plus transitional elements (TE), are Neodymium Iron Boron (NdFeB) and Samarium Cobalt (SmCo) magnets, as shown in figure 27.6.

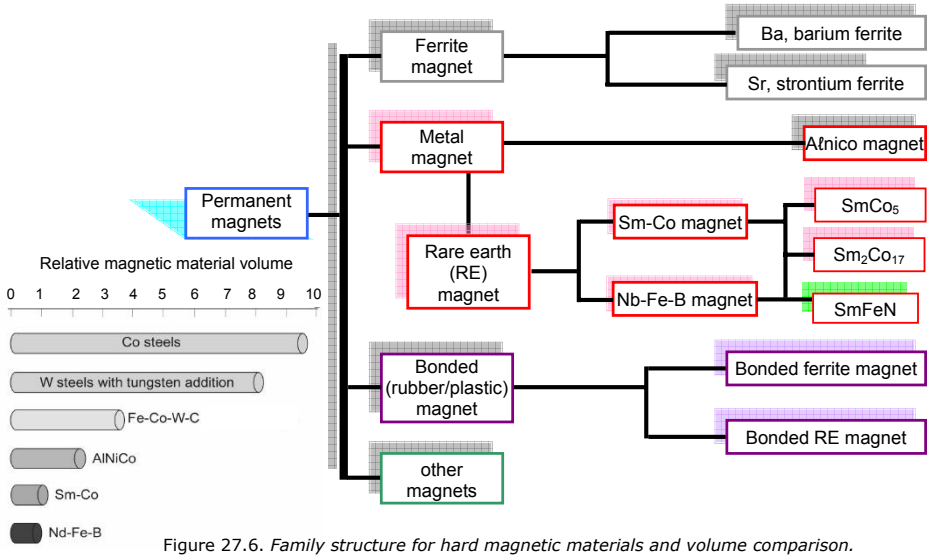


Figure 27.6. *Family structure for hard magnetic materials and volume comparison.*

27.2.1 Alloys

i. Nickel-iron alloys

Nickel-iron alloys, known as permalloy, are versatile and are composed of a wide range of magnietc material, from 30 to 80wt%Ni. Widely varying properties result and the optimum composition is selected for a particular application. High Ni content alloys have high permeability; around 50wt%Ni gives high saturation magnetisation while a low Ni content produces a high electrical resistance (low ac losses).
Some Ni-Fe alloys have zero magnetostriction and zero magnetic anisotropy, such as mumetal which is produced by a specific heat treatment and incorporates minimal additions of Cu and Cr. Such alloys have extremely high relative permeability, up to 300,000, and an intrinsic coercivity as low as 0.4Am^{-1}.

ii. Cobalt alloys

(a) Alnico hard magnet material: (alloys based on Al, Ni, and Co)

Chemical composition:

Wt %	Al	Ni	Co	Cu	Ti	Nb	Si	Fe
AlNiCo	6 - 13.5	12 - 28	0 - 42	2 - 6	0 - 9	0 - 3	0 - 0.8	balance

A hard magnet that incorporates aluminium, nickel, cobalt, ferrum, and titanium, is called an AlNiCo magnet. Typical weight% is: Fe-35, Co-35, Ni-15, Al-7, Cu-4, Ti-4.
This group of magnets offer far more magnetic hardness than magnetic steels. Their properties rely on the shape anisotropy improvement associated with the two-phase nanostructure comprised of ferromagnetic Fe-Co needles in a matrix of non-magnetic Al-Ni. Due to their high Curie temperature, ~850°C, they have more stable properties around room temperature than some other alloys. This material has the lowest temperature coefficient (-0.02% per degree centigrade) of all permanent magnets, thus producing a constant field over a wide temperature range (-270°C to +500°C).

Alnico is classified as either an isotropic or anisotropic hard magnetic material. Alnico 2, 3 and 4 are unoriented – magnetic properties are isotropic and equal in all directions. Grades 5 to 9 are anisotropic.

There are two manufacturing techniques for AlNiCo hard magnet, namely casting and powder metallurgy (sintering for smaller magnets). If the cobalt content is more than 15%, the introduction of an external magnetic field during heat treatment at 1260°C in a hydrogen (or inert gas or a vacuum) atmosphere creates anisotropy for the sintered magnet, allowing the magnetic property to be increased in the preferred direction. For column crystal or monocrystalline shaped material, a strong external magnetic field parallel to the column crystal axis during heat treatment (annealing), produces optimal properties for the cast (and sintered) hard magnetic material.
Disadvantages of the Alnico materials are relatively low coercivity and mechanical brittleness.
Anisotropic columnar Alnico 9, has an energy product of ~80kJm^{-3}. However, although having a relative high B_r, its main disadvantage is low intrinsic coercivity (H_c ~50kAm^{-1}) thus it must be made in the form of horseshoes or long thin cylinders/rods, which cannot be exposed to significant demagnetising fields.

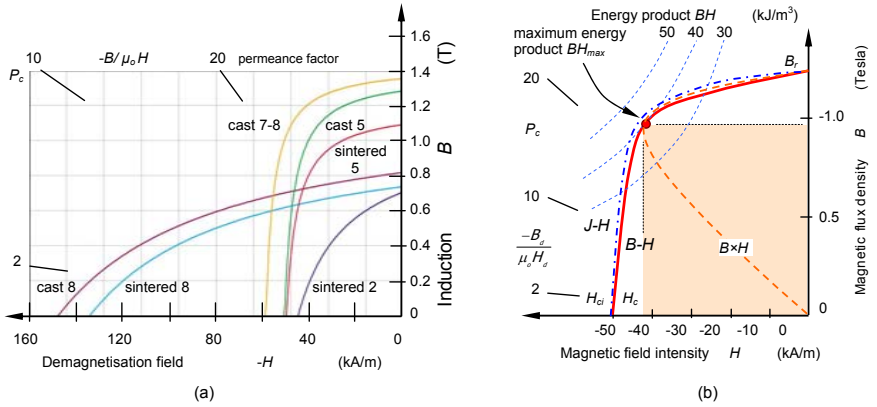


Figure 27.7. *Second quadrant hysteresis loop for: (a) Alnico family and (b) Alnico 5.*

(b) CrFeCo hard magnet material

Chemical composition:

Wt %	Cr	Co	Si, Ti, Mo, Al, V	Fe
CrFeCo	25 - 35	7 - 25	0.1 - 3	balance

CrFeCo can be classified into isotropic and anisotropic hard magnetic materials.
There are two production techniques for CrFeCo hard magnet alloy: casting and powder metallurgy. The material is ductile (good mechinability) and can be easily hot rolled or cold rolled into strip or drawn into bar shapes, or, after punching, turning or drilling, can be made into the required shape and still maintain its magnetic properties. It has a low coercivity, and is relatively high in cost. Although mostly iron, which is inexpensive, there is significant cobalt and chrome content which is considerably more expensive. Typically strong, hard, tough, but brittle. It is hydrogen resistant.
[B_r=1.35T, BH_{max}=52kJ/m^3, H_c= 49kA/m, α=-0.03%/°C, β=-0.04%/°C, T_C=640°C, T_{op}=500°C and μ_{rc} = 2.5]

(c) FeCoVCr hard magnetic material

Chemical composition:

Wt %	Co	V+Cr	Fe
FeCoVCr	49 - 54	4 - 13	balance

FeCoVCr hard magnetic anisotropic alloy is produced using a casting process, and can be hot rolled or cool rolled into strip or drawn into wire shapes. Cool deformation (80% to 95%) and subsequent annealing heat treatment at 500 to 600ºC is a necessary processing step to obtain the required anisotropic magnetic properties.
[B_r = 0.85T, BH_{max} = 15kJ/m^3, H_{ci} H_c = 28kA/m, α=-0.01%/°C, β=0%/°C, T_C=720°C, T_{op}=500°C, and μ_{rc}= 5]

iii. Samarium Cobalt alloys

Chemical composition, typically:

Wt %	Sm	Co	Fe and Cu
SmCo	35	60	balance

This alloy group combines cobalt, iron (transition metals - TM) and a light rare earth (RE) element to exhibit permanent magnetic properties. They exhibit high-energy hard-magnetic behaviour, but are costly. These magnets have good thermal stability, thus are used where magnets are exposed to high temperatures (>150°C) or very low temperatures (<100K). They are brittle and hard to machine.
The combination of a rare earth and a transition metal is ideal as the rare earth provides the anisotropy to the phase and the transition metal provides a high magnetisation and Curie temperature. By varying the percentages of the composition and changing the sintering and heat treatment cycles, two grades are produced, namely $SmCo_5$ termed 1:5 and Sm_2Co_{17} termed 2:17. They differ in energy product, temperature coefficient, and magnetisation force required to saturate. In the case of $SmCo_5$

$$Sm_2O_3 + 10\,Co + 3\,Ca \rightarrow 3\,CaO + 2\,SmCo_5$$

The reaction takes place at 1100°C for 1 to 4 hours in a hydrogen (or inert gas) atmosphere or a vacuum. Usually excess Ca is added, typically 20 to 40%, to assure that the reaction goes to completion. $SmCo_5$ sintered magnets have energy products of ~160kJm^{-3}. These magnets have excess Sm that forms a smoothing grain boundary phase and coercivity is achieved by preventing the nucleation of reverse domains.
The maximum energy product is increased to 240kJm^{-3}, with a Sm_2Co_{17} based alloy. The theoretical limits are 230kJ/m^3 and 270kJ/m^3 for $SmCo_5$ and Sm_2Co_{17}, respective These materials are based on the composition $Sm_2(Co,Fe,Cu,Zr)_{17}$ and achieve permanent magnetic properties by control of the microstructure. The magnets are produced by powder metallurgy and are solution treated at ~1100ºC, where they are single phase, then heat treated at 500ºC to 800ºC. This homogenising stage is followed by aging treatments at lower temperature where a cellular microstructure is formed. The cells are based on the Sm_2Co_{17} type phase, which is enriched in Fe and the cell boundaries comprise a layer of $SmCo_5$ type phase, which is enriched in Cu. The intrinsic magnetic properties of the cells and their boundaries vary such that the magnetic domain wall energy is greatly reduced within the cell boundary and hence pin the domain walls, leading to permanent magnetic properties. They can operate at high temperature, ~500ºC, when small quantities of zirconium, hafnium, etc. are added to give better heat treatment response.

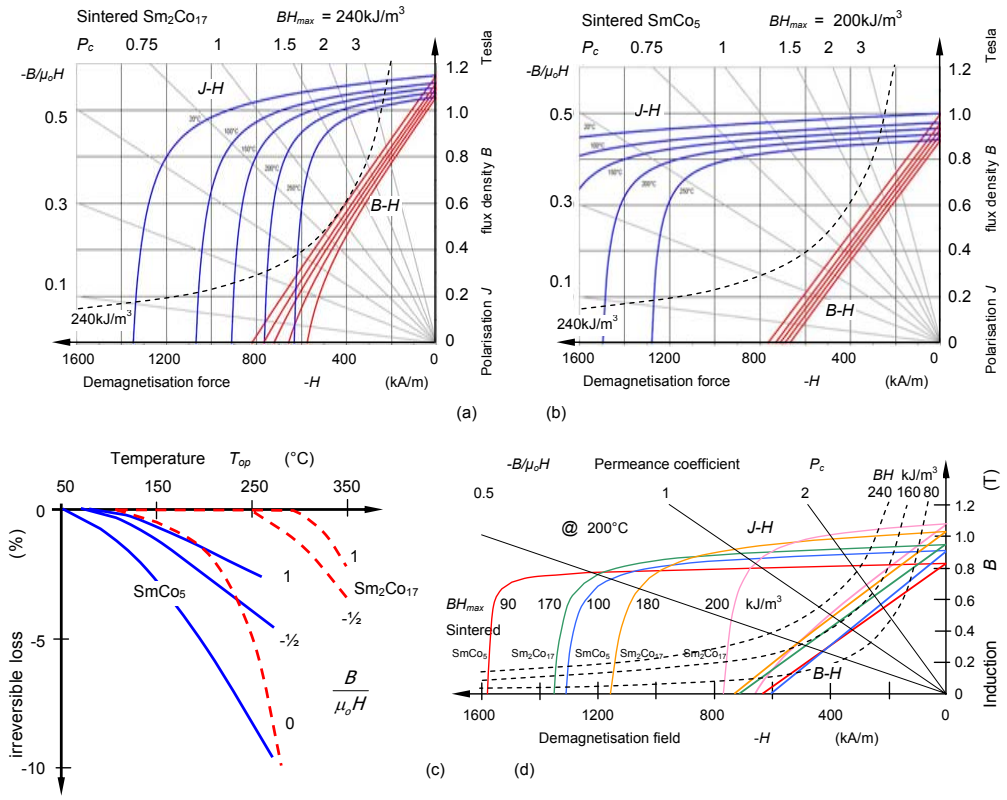


Figure 27.8. *Second quadrant hysteresis loop for samarium cobalt alloys: (a) Sm_2Co_{17}, (b) $SmCo_5$, (c) characteristics at 200°C (with constant energy contours shown), and (d) temperature dependant irreversible losses as a function load line slope.*

Both grades exhibit excellent energy, temperature, and corrosion resistant properties, applicable when a compact high energy field is required at temperatures of over 100°C or in hostile environments. These alloys have the best reversible temperature coefficient of all rare earth alloys, typically -0.03%/°C, as shown in Table 27.2. SmCo requires large magnetizing fields (in excess of 28MAt/m) to saturate and under operational conditions, are difficult to de-magnetise.

A problem with Sm/Co based magnets is the expense of the raw materials. Samarium is much less abundant than other light rare-earth elements, such as La, Ce, Pr, and Nd, which account for over 90% of rare-earth metals in typical rare-earth ores.

Samarium-cobalt magnets are manufactured by similar routes as ferrites, being formed into either a fully dense sintered magnet, or a compression - or injection-moulded bonded magnet. Because both samarium and cobalt are relatively expensive elements, anisotropic magnets are usually produced with optimized properties along a pre-determined axis.

Rare-earth magnets have a rather more complicated domain wall mechanism than that for pure magnetocrystalline anisotropy, such that the best magnetic properties are only achieved with the powder milled to a grain size which is an order of magnitude larger than the single domain size. This means that, not only can domain walls exist, but also they can move relatively freely within a grain. While this allows saturation magnetization to be achieved with only a modest applied field, a high intrinsic coercivity will depend upon the grains' ability to resist the formation of a reverse domain when a demagnetizing field is applied. This vital property is therefore controlled by the grain boundaries, which are composed of deviations from the primary composition that provides a strong pinning of the domain walls at these sites. This mechanism, known as *nucleation*, where a grain undergoing nucleation, occurs in $SmCo_5$ magnets.

Sm_2Co_{17} differs from $SmCo_5$ in that its grains contain a fine cell structure. Heat treatment of this compound promotes the formation of Sm_2Co_{17} cells, separated by thin walls of $SmCo_5$ which can provide pinning of the domain walls (rather than the grain boundaries). Pinning, rather than nucleation, is therefore the controlling mechanism in Sm_2Co_{17} magnets, and this requires that a much greater field be applied to initially magnetize it into saturation.

Whether it is a nucleation-type $SmCo_5$ or a pinning-type Sm_2Co_{17} magnet, that is, whether the domain walls are pinned at the grain or cell boundaries, they will move quite freely once these pinning forces are overcome, and M_{sat} will abruptly flip over into the opposite direction when an applied field of $-H_{ci}$ is reached. This effect is seen in the shape of the demagnetization curves for samarium-cobalt type magnets, figure 27.8. The change in H_{ci} with temperature is seen in figure 27.8a for Sm_2Co_{17} and largely for $SmCo_5$ in figure 27.8b. The theoretical Sm_2Co_{17} limit is 270kJ/m^3, N34, with the single phase $SmCo_5$ limit is 230kJ/m^3, N28.8,

iv. Samarium Iron Nitride alloys

The rare earth alloy $Sm_2Fe_{17}N_3$ incorporates Fe rather than the more costly transition metal Co. While its magnetic properties can theoretically exceed those of $Nd_2Fe_{14}B$, its processing is more complex. Only bonded magnets (extrusion-flexible and injection-rigid) are made from $Sm_2Fe_{17}N_3$, (1-2μm powder) which are more easily made in isotropic form than anisotropic. Such permanent magnet alloys offer high resistance to demagnetisation, high magnetisation, lightweight, and increased resistance to corrosion and temperature, compared with neodymium iron boron. They oxidize readily, have high irreversible loss at high temperature and are expensive.

[Extrusion moulded, flexible: B_r = 0.78T, BH_{max} = 111kJ/m^3, H_{ci} = 812kA/m, H_c = 520kA/m, α=-0.05%/°C, β=-0.45%/°C, T_C = 498°C, T_{op} = 80°C, and μ_{rc} = 1.15:
Injected moulded, rigid: B_r = 0.81T, BH_{max} = 115kJ/m^3, H_{ci} = 756kA/m H_c = 533kA/m, α = -0.05%/°C, β=-0.45%/°C, T_C = 498°C, T_{op} = 110°C, and μ_{rc} = 1.15]

Explosion sintered $Sm_2Fe_{17}N_y$ yields a squarer $B{\times}H$ area, which is further improved, Curie temperature wise, with the addition of Ta. [B_r = 0.83T]

v. Neodymium Iron Boron alloys

NdFeB combines a high saturation magnetisation with good resistance to demagnetisation. The high cost of samarium and the price volatility of cobalt have lead to NdFeB magnets becoming the preferred material for applications requiring high-energy magnets. Despite the high energy-product, these magnets have a relatively low Curie temperature, 312°C, which prohibits their use in high temperature applications. Additions of Co/Tb and Dy improve the temperature and coercivity characteristics but decrease the saturation polarisation and increases production costs. Co improves corrosion resistance. NdFeB is easier to machine than Alnico and SmCo magnets. NdFeB are materials based on the magnetic phase $Nd_2Fe_{14}B$, with two different powder metallurgy processing routes being employed.

Process route #1 *Sintering (orient-press-sinter)*

Powder NdFeB based sintered permanent magnets produced a maximum energy product of ~451kJm^{-3}, by accurate heat treatment, controlled processing, and with the use of iron rich compositions.

Sintered NdFeB based magnets achieve their coercivity by virtue of an Nd-rich phase at the grain boundaries, which acts to produce liquid phase sintering, smooth the boundaries, hence prevent nucleation of reverse magnetic domains.

The processing sequence for sintered NdFeB based magnets is shown in figure 27.9. The *as-cast* (untreated) ingot must first be broken into a powder. This is achieved most conveniently by exposing the ingot to hydrogen, which is absorbed at the surface. The hydrogen enters the material in the spaces between the atoms and causes the material to expand. The differential expansion generates stress in the ingot and the alloy breaks down into a fine powder, with an average grain size of 100μm. This process is known as *Hydrogen Decrepitation*, HD. The HD powder is then broken up further by a jet milling stage, which reduces the grain size to less than 5μm – the size of a few domains and therefore inherently anisotropic. When the alloy is in a powdered form, it is flammable, thus is handled under an inert gas.

Each particle of the broken down powder is a single crystal, 3 to 5μm, which can be aligned in a magnetic field; a few 10ms alternating pulses at 6.4MA/m. This alignment is held in place by pressing the powder into a green (unsintered) compact, which is about 60% dense. The compact is then heated in a vacuum (or inert gas) to ~1060ºC for 4 hours. During the heating stage, the hydrogen is driven out of the material and is pumped away. Sintering occurs and the compact densifies, with the assistance of a liquid formed by the melting of the Nd-rich phase. After sintering, then quenched the magnets are heat-treated, annealed, at ~550ºC to 800ºC for 1 hour, thereby achieving optimum magnetic properties.

The magnet is then machined to the dimensions for the intended application. Due to the large degree of shrinkage that occurs during sintering, which is greater in the direction of alignment, it is not possible to press compacts that will shrink to the exact size required. Machining is an expensive operation and, particularly for small magnets, a large proportion of the material may need to be machined away.

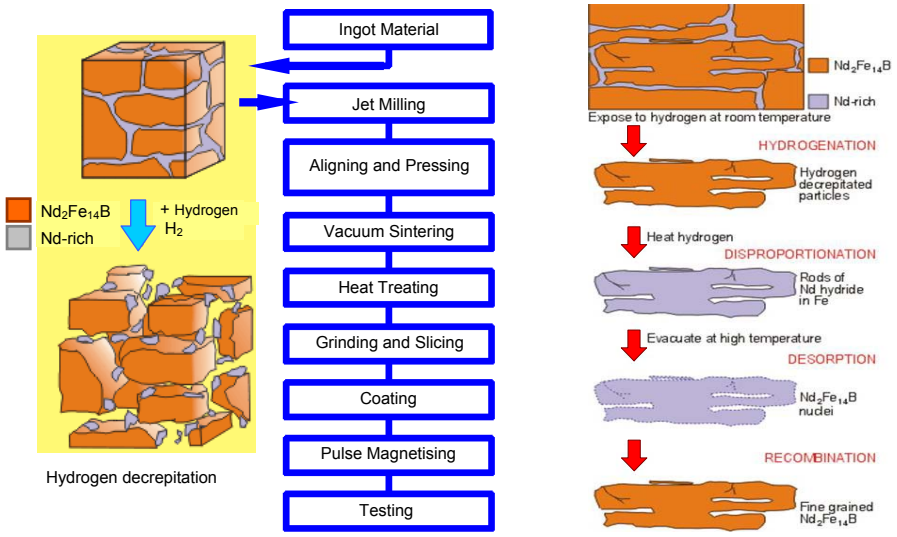


Figure 27.9. *The processing route for sintered NdFeB permanent magnets, left HD and right HDDR.*

Due to the highly reactive nature of the Nd-rich phase, the magnets tend to corrode rapidly, particularly in moist environments. Therefore, the next processing stage is to provide a protective barrier on the magnet surface, usually a nickel (plus copper) coating, although aluminium, zinc, and epoxy resins are also used. Finally, the magnet material is magnetised and tested.

An extension to the HD process is the *hydrogenation disproportionation desorption and recombination* (HDDR) process. Unlike the more straightforward HD-process, the HDDR route involves heating the bulk alloy in 1 bar of hydrogen to ~800°C. The first stage is the decrepitation of the alloy due to the initial absorption of the hydrogen, as described previously. At ~700°C, the $Nd_2Fe_{14}B$ matrix phase disproportionates during an exothermic and reversible reaction involving hydrogen absorption

$$Nd_2Fe_{14}B + (2 \pm x)H_2 \Leftrightarrow 2NdH_{2\pm x} + 12Fe + Fe_2B + \Delta H$$

where *x* depends on temperature and pressure. Then desorption results in isolated 0.3um grains, which exhibit appreciable coercivity. Fully dense isotropic magnets are produced by hot pressing the powder mixed with thermosetting resins in a compression mould, at 750°C in an inert atmosphere.

In summary, neodymium-iron-boron magnets may be made from alloy powder, which is:

- sintered, nucleation-type;
- rapidly quenched, magnetocrystalline; and
- HDDR, magnetocrystalline.

Processing route #2 *Melt Spinning*

Melt-spinning is used to produce a ribbon like powdered material. Molten alloy is ejected onto the surface of a rotating water-cooled wheel, and cooling rates (rapid quenching) of the order of one million °C/s are achieved. The microstructure and magnetic properties of the NdFeB ribbons are sensitive to the quench rate, that is, the speed of the rotating wheel. High quench rates produce essentially amorphous ribbons (that is, no crystal grains) having negligible intrinsic coercivity. Optimum quench rates yield ribbons with high coercivities; they are comprised of roughly spherical $Nd_2Fe_{14}B$ grains (20 to 100 nm in diameter), which are single domain particles thus have a high coercivity, ~1MAm^{-1}. At wheel velocities below optimum, the slow cooling rate produces ribbons with larger grains and are characterised by low coercivities. This powder cannot be sintered to produce fully dense magnets without destroying the magnetic properties, but is utilised in one of three ways:

MQ-I

The melt spun ribbon (typically milled into thin platelets 200μm wide and 35μm thick) is blended with a resin to produce a bonded permanent magnet. The crystals of *MQ-I* material are randomly oriented so that the magnets are isotropic and can be magnetised along any axis to a BH_{max} of ~80 kJm^{-3}.

MQ- II

Improved densification of melt-spun ribbons can be achieved by hot pressing at ~750ºC, without adversely effecting the coercivity of the powder. These *MQ-II* type magnets exhibit a slight degree of magnetic alignment, ~10%, and are 100% dense, that is, the magnetic properties are not diluted by a non-magnetic material, such as a resin. This gives *MQ-II* a higher BH_{max} than *MQ-I*, 100 - 120 kJm^{-3}.

MQ-III /sintered

Substantially greater alignment (>75%), and hence greater maximum energy products, can be obtained by heating the *MQ-II* material to ~750ºC in a die cavity having a larger diameter and then it is slowly deformed. This second hot press, termed die upset forging, produces plastic flow and a reorientation of the crystals. Such magnets, known as MQ-*III*, are 100% dense, and because of the alignment of crystals, have maximum energy products of ~400 kJm^{-3}.

NdFeB magnets made by rapid quenching, as opposed to sintering, require a much greater applied field to initially align the grains' magnetizations and magnetize this material to its saturation level. Neodymium-iron-boron magnets therefore exhibit similarly shaped demagnetization characteristics to the other classes described, with a well-defined *knee* (Cunic point) at which M_{sat} reverses as an applied field of $-H_{ci}$ is approached. As an example, figure 27.9b shows the demagnetization characteristics measured at different temperatures for one grade of fully dense anisotropic $Nd_2Fe_{14}B$. In all cases the higher the energy density, the lower the Curie temperature. $Nd_2Fe_{14}B$ with an energy density of 451kJ/m^3 must be operated at a temperature below 50°C. The theoretical maximum is 512kJ/m^3, N64.

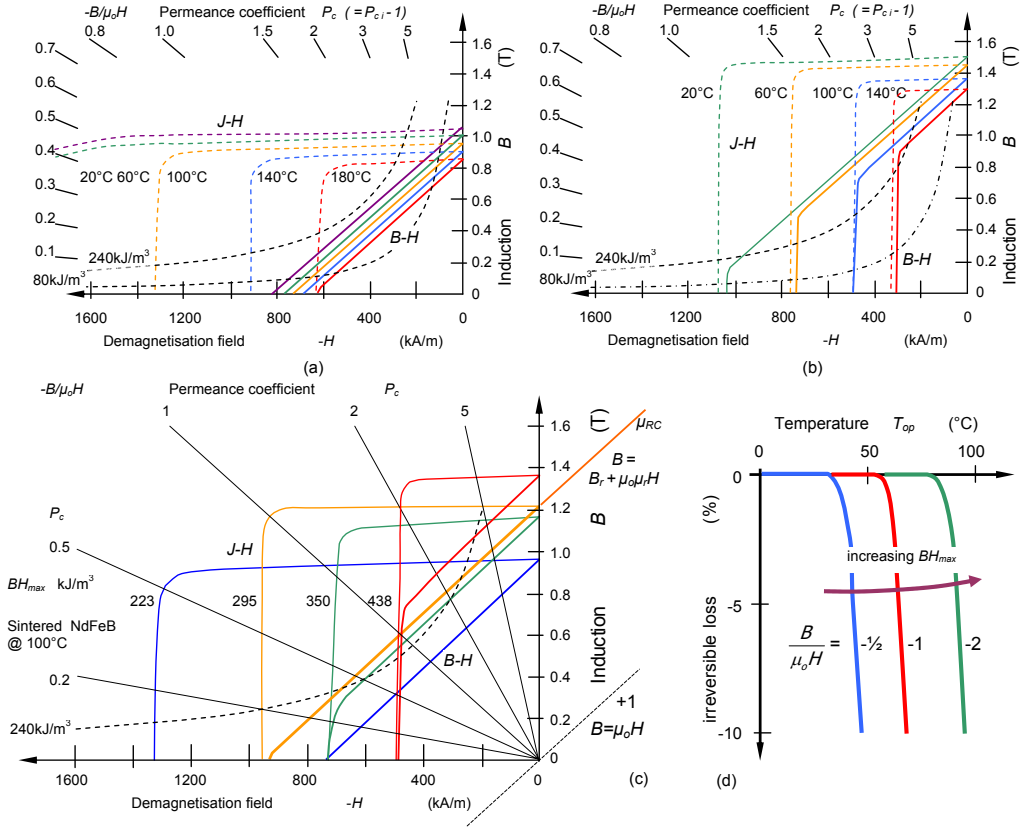


Figure 27.10. *Second quadrant hysteresis loop for sintered neodymium-iron-boron magnets: (a) N28, (b) N55, (c) characteristics at 100°C (with constant energy contour shown), and (d) temperature dependant irreversible losses as a function load line, N55.*

Table 27.3 NdFeB magnetic characteristics - sintering results in higher B_r than *MQ* processes

NdFeB Grade		Melt spin property (Magnequench)			sintered		
	symbol	*MQI*	*MQII*	*MQIII*	N28	N55	units
Maximum energy product	BH_{max}	68 – 88/ 9 - 11	111 – 119/ 15 - 15	255 – 335/ 32 - 42	223 / 28	438 / 55	kJ/m^3 / MGOe
Residual induction	B_r	0.61 – 0.71	0.8 – 0.83	1.16 – 1.31	1.06	1.50	T
Coercive force	H_c	412 - 446	557 - 573	875 - 979	835	1082	kA/m
Intrinsic coercive force	H_{ci}	716 - 1353	1393 - >1433	1173 - >1593	2705	1074	kA/m
Rev. Temp. Coefficient	α - B_r	-0.105	-0.10	-0.09	-0.11	-0.11	%/°C
Rev. Temp. Coefficient	β - H_{ci}				-0.55	-0.65	%/°C
Magnetising filed	H_s	1990 - 2785	3581	2785 - 3581			kA/m
Recoil relative permeability	μ_{rc}	1.15 – 2.31	1.14	1.05 – 1.09	1.05	1.05	pu
Maximum operating temperature	$T_{op\text{-}max}$	80 - 180	160 - 200	150 - 200	200	80	°C
Curie temperature	T_C	305 -470	325 - 370	325 - 370	365	310	°C

vi. Amorphous and Nano-Crystalline Alloys

These materials can be produced in the form of a tape by melt-spinning. The alloys consist of iron, nickel and/or cobalt with one or more of the following elements: boron, carbon, phosphorous and silicon. They have extremely low coercivity, an order of magnitude less than standard Fe-Si, and consequently have lower hysteresis losses. However, they have a relatively low magnetisation and are not suitable for high current applications.
Instead of casting the alloy onto a rotating wheel to produce tapes, it is also possible to direct a stream of molten alloy into a bath of water or oil to produce amorphous wires of typically 50mm thick. These wires exhibit a square hysteresis loop with large changes in magnetisation at low fields, making them ideal for sensing and switching.
Nano-crystalline material is produced by annealing the amorphous material. These alloys can be single phase but are usually comprised of nano-sized grains, in the range 10 to 50nm, in an amorphous matrix. They have relatively high resistivity, low anisotropy, good mechanical strength, and are better suited for soft-magnetic applications.

27.2.2 Ceramics

Hard Ceramic hexaferrites: ($BaFe_{12}O_{19}$ or $SrFe_{12}O_{19}$)

Hard hexagonal ferrite materials, often referred to as ceramic magnets, are ferrimagnetic and considering the proportion of iron within the material, have a low remanence, ~400mT. The coercivity of these magnets is typically ~250kAm^{-1}. The low remanence means that the maximum energy product is only ~40kJm^{-3}, which is lower than the alnicos, but due to the high coercivity, these magnets can be made into thinner sections. The magnets can be exposed to moderate demagnetising fields, and coupled with good mechanical characteristics and low cost, are suitable for applications such as permanent magnet motors.
Ceramic ferrites are made using an iron oxide powder, to which either barium or strontium (carbonate) is added to improve alignment of the crystal lattice structure. The chemical formula is $M{\cdot}n(Fe_2O_3)$, where M is Ba or Sr and $5.8<n<6.0$. After milling, which produces small grains, <1mm, which is essential for generating coercivity in ferrites, the powder is pressed in a die, with an orienting field applied through the cavity if desired, enhancing the remanence and the maximum energy product. If no field is applied, an isotropic magnet results with no preferred magnetic properties in any particular direction. If an orienting field is applied, an anisotropic magnet results having preferred magnetic properties along a specific axis. The compacted powder is then sintered at a temperature of 1100°C to 1300°C (hence the name *ceramic*) to achieve full densification, cooled at a rate of less than 90°C per hour to avoid thermal shock, and then diamond ground to the required final dimensions. There are no problems with oxidation of the powder during processing, as the material is a stable iron oxide. Alternatively, the powder may be blended with a polymer binder, and then either extruded or formed in a die by compression, or injection-moulding, thus producing a bonded ferrite magnet of near net shape; again, anisotropic properties may be achieved by applying an orienting field through the die cavity. Due to their anisotropic structure ferrites exhibit relatively high coercivity, however, the energy product is low.

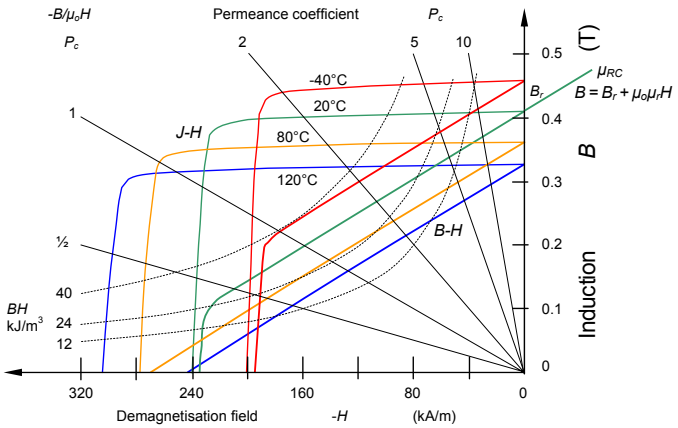


Figure 27.11. *Second quadrant hysteresis loop demagnetization characteristics measured at different temperatures for a sintered ceramic ferrite magnet (with constant energy contours shown).*

A hexagonal ferrite structure is found in both $BaO{\cdot}6(Fe_2O_3)$ and $SrO{\cdot}6(Fe_2O_3)$, with Sr ferrite having slightly superior magnetic properties.
Generally, magnet materials respond negatively to heat because thermal energy reduces the flux density and the ability of domains to remain aligned. Increased thermal energy increases the disorder present and at a particular temperature, the Curie point, the material loses its ferromagnetism. Excessive heating of a magnet may cause metallurgical changes to occur; in many magnets the safe operating temperature is well below the Curie point. The one exception is ferrites (ceramic magnets), which, due to their chemical nature, can withstand temperatures significantly past their Curie points. Ferrites also exhibit the unusual characteristic of increased coercivity with temperature, as shown in figure 27.11. Heat treatments during manufacturing are precisely controlled to precipitate desired phases and to control metallurgical changes.

The main advantage of ferrites is low cost, due to the ease of processing and the abundance, non-strategic, and low cost of raw materials, and complex shape possibilities. Ceramic magnet material (ferrite) has modest resistance to corrosion and can operate in moderate heat but demagnetise at low temperatures, typically, 20°C, as the knee moves up the characteristic line. Ferrites are brittle.

27.2.3 Bonded

Rigid bonded magnet materials can be made from Alnico, Ceramic, isotropic NdFeB (*MQ*), anisotropic NdFeB (HDDR), SmCo, SmFeN, or nano composite powders which are combined (forming a thermoset or thermoplastic) with a variety of plastic binders, a matrix with polyphenylene sulphide (PPS), chlorinated polyethylene (CPE) polyester, PVC, nylon (polyamide), duroplast or nitrile rubber. Rigid bonded magnets are processed by compression, injection, extrusion or calendaring shaping methods.
They can be *injection moulded* thermoplastic bonded into mechanically-strong complex magnet shapes with accurate finished dimensions. Bonded magnet materials have a moderate resistance to corrosion and a low tolerance to heat, -40°C to 180°C, because of the binder material. The same bonder properties tends to make them chip and break resistant.
Bonded thermoset magnets can also be *compression bonding*, which offers a higher magnetic output (than injected moulded bonded magnets), due to a higher magnetic particle density than either NdFeB or SmCo powders, but are restricted to simpler geometries than injection moulded materials. The use of an epoxy binder and epoxy surface coating prevents oxidation and is resistant to normal industrial solvents and automotive fluids. Because of the compression bonding process, the tooling tolerances and mechanical strength properties are slight less than those materials that have been injection moulded.
In the case of NdFeB, the bonded material is isotropic, offering approximately 80kJ/m^3 or 30% of the energy produced by the sintered fully dense material, but can be magnetised in any direction. Typical of bonded magnets, the operating temperature range (specifically the upper limit) is restricted to -40°C to 165°C, with a poor temperature coefficient.

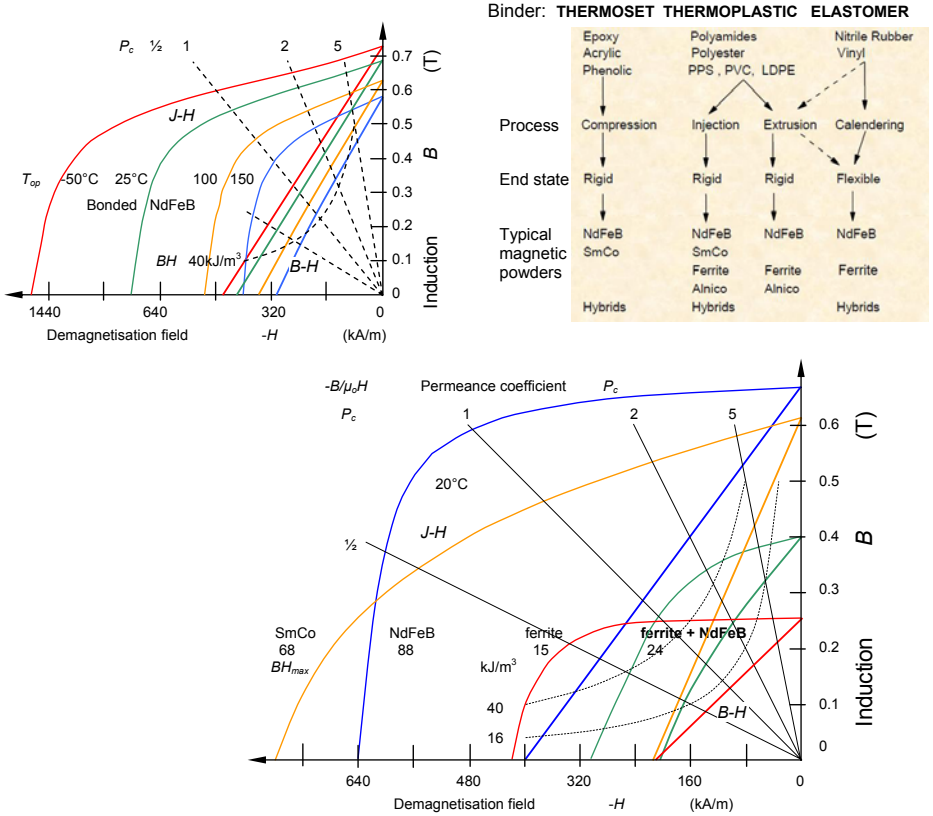


Figure 27.12. *Second quadrant hysteresis loop demagnetization characteristics measured at different temperatures for bonded magnets (with constant energy contours shown).*

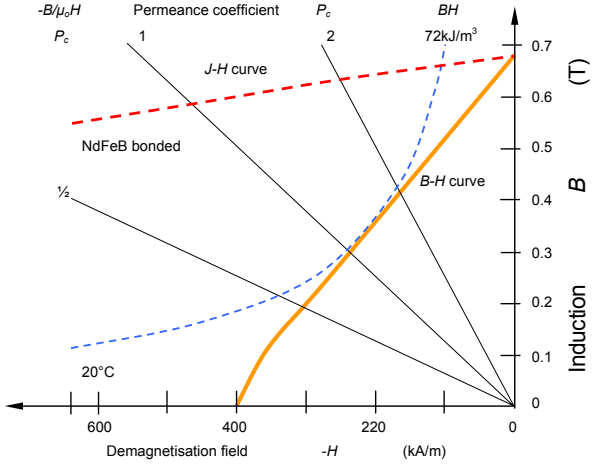


Figure 27.13. *Second quadrant hysteresis loop demagnetization characteristics for a flexible bonded magnet (with a constant energy contour shown).*

The energy product $(BH)_{max}$ of bonded magnets is also lower than for a fully dense material. This is due to the magnetic material being diluted by the binder. Bonded samarium cobalt bonded materials are typically anisotropic with energy products of 40kJ/m^3 to 120kJ/m^3.
The polymer matrix associated with bonded magnets weakens magnetic properties and limits thermal properties. Their attribute is freedom of shape and design.

27.2.4 Flexible (rubber)

Flexible (rubber) magnets are manufactured by mixing (bonding) ferrite or neodymium magnet powders and synthetic or natural rubber (nitrile, vinyl, etc.) and resin elastomer binders. Flexible (rubber) magnets are manufactured by rolling (calendaring), injection or extrusion methods. Versatility, low cost, and ease of use are among the reasons to choose ferrite based flexible (rubber) magnets. Flexible magnets are usually manufactured in strip or sheet form. There are *isotropic* and *anisotropic* forms. The isotropic rubber magnet is normally magnetized with multipoles magnetization on one side only while the anisotropic form can have magnetics on both sides. The magnet material is low energy, as shown in figure 27.13, and does not usually replace fully dense magnet materials. Flexible neodymium material is higher in strength, but is made in limited quantities because the cost is high.

Table 27.4 Permanent magnet material features

Permanent magnet material	advantages	disadvantages
Alnico	Thermal stability	Brittle
	High service temperature	Low coercive force
	High flux density	Required thickness
	Complex shapes	Cost variability
	Easy to magnetise	
	Low tooling cost	
Hard ferrite	Low cost	Brittle
	High coercive force	Limited shape and tolerances
	Easy to magnetise/demagnetise	Low energy product
	Stability to oxidation	Performance varies – hi temperature dependency
		High tooling costs
Bonded ferrite	Flexible	Low energy product
	Easily shaped	Low service temperature
	Can build subassemblies	Performance varies – hi temperature dependency
	Requires little machining	
Samarium Cobalt	High energy product	Brittle
	High coercive force	High cost
	Compact	Difficult to magnetise/demagnetise
	High service temperature	Easy to oxidize
	Corrosion resistant	Powders pyrophoric – spontaneous ignition
	Thermal stability, low TC	
Neodymium-iron-boron	Highest energy product	Performance varies – medium temperature dependency
	High coercive force	Susceptible to corrosion
	Compact	Low service temperature
	Low energy cost	Difficult to magnetise
		May require coating
		Powders pyrophoric
Bonded Neodymium-iron-boron	High energy product	Performance varies – hi temperature dependency
	High coercive force	Susceptible to corrosion
	Low energy product cost	Low service temperature
	Easily shaped	
	Can build subassemblies	

27.3 Properties of hard magnetic materials

A permanent magnet is a material that when inserted, in a virgin form, into a strong magnetic field will not only begin to exhibit a magnetic field of its own, but also continues to exhibit a magnetic field once removed from the original magnetising field. Permanent-magnet materials are characterized by high values of remanent magnetization and coercivity. Such materials produce significant magnetic flux, even in magnetic circuits with air gaps.
A permanent magnet's magnetization *M* [the sum of the individual (unpaired electron spins) magnetic moments per unit volume] provides a magnetizing force *H* which establishes a flux density *B;* these being related by $B = \mu_o(H + M) = \mu_o H + J$. This relationship can be used to convert the intrinsic *M versus H* magnetization characteristic into the usual *B versus H* magnetization characteristic shown in figure 27.12. The *B versus H* characteristic is far more useful for practical magnet design since *M* exists only within the magnet while *B* 'flows' through the magnet and in the surrounding media,.

The *B* within a magnet is indicative of the flux density it will deliver into the adjacent air gap, and the point at which a magnet operates on its demagnetization curve relates *B* to the *d*emagnetizing force *-H* it experiences. The demagnetization curve shows that as the magnitude of *-H* increases, the flux density delivered by the magnet falls, ultimately at $H = -H_{ci}$ to $B = 0$. However, an unconstructive phenomenon begins within the magnet before $-H_{ci}$ is reached, because the *knee* (Cunic point) in the demagnetization curve represents the onset of a reversal of the material's *M* (that is, the threshold for irreversible loss, where the characteristic rapidly leaves the second quadrant and moves into the third quadrant). Thus it is desirable that the operating point of a permanent magnet always remain above the temperature dependant *knee* in the demagnetization curve.

The second quadrant of a hysteresis loop (the *demagnetization curve*) is usually employed for analyzing a permanent-magnet material. Residual flux density or remanent magnetization is B_r, while coercivity, H_c, is a measure of the magnitude of the mmf required to demagnetize the material, and the capability of the material to produce flux in a magnetic circuit that includes an air gap.

Magnets are characterized by the three main characteristics.

- Residual Induction or Remanent flux density (B_r, measured in Tesla).
  - An indication of how strong the magnet is capable of being.
- Coercive Force or Coercivity (H_c, measured in kA/m).
  - An indication of inherent stability and how difficult it is to demagnetize the magnet.
- Maximum Energy Product per unit volume (BH_{max}, measured in kJ/m^3).
  - An indication of the magnet material volume required to produce a given level of magnetic flux. It is where the magnetic field energy density into the air gap surrounding the magnet, is at a maximum, i.e., the operating point that minimises magnet volume.

The temperature dependence of residual induction and coercive force are modelled by

$$B_r(T) = B_{r\,20°C}\left(1 + \alpha(T - 20°C)\right)$$
$$H_{ci}(T) = H_{ci\,20°C}\left(1 + \beta(T - 20°C)\right) \qquad (27.3)$$

where reversible temperature coefficient of B_r and H_{ci} (and H_c), specifically α and β respectively, are defined as

$$\alpha = \frac{\Delta B_r}{B_r} \times \frac{1}{\Delta T} \times 100\% \qquad \beta = \frac{\Delta H_{ci}}{H_{ci}} \times \frac{1}{\Delta T} \times 100\% \qquad (27.4)$$

Behaviour of a permanent magnet is referenced to the second quadrant of its *B versus H* curve, where it is termed the *demagnetization curve*. The second quadrant of the curve (b-o-c), shown in figure 27.14, can be used to demonstrate the demagnetisation characteristics of a permanent magnet material.

Most magnetic materials are classed as *anisotropic*, that is, a preferred magnetic axis is determined during manufacturer so that the magnets can only be magnetised in that predetermined direction. *Isotropic* materials may be magnetized in any direction, but generally have inferior performance than anisotropic grades.

- Remanence is the intercept of the *B versus H* curve on the positive *B* axis. For an *ideal* material, its value $B_r = \mu_o M_{sat}$, but B_r is always the value of flux density for the condition when a magnet develops no magnetizing force ($H = 0$). Magnetization *M*, a bulk material property, is the sum of the magnetic dipole moments μ_m per unit volume [M_{sat} is saturation magnetisation, A/m]. B_r is coincident to both the intrinsic *J-H* curve and the normal *B-H* curve.
- Coercivity is the intercept of the *B versus H* curve on the negative *H* axis. Its value $-H_c$ is the magnetizing force required to reduce the magnet's flux density *B* to zero, which on the *ideal* curve in figure 27.14, is $-H_c = M_{sat}$. There is no observable (external) field because the applied field *H* (namely, H_c) is balanced out by the flux *M* of the magnet material. Because they are

opposing, the net observable induction *B,* is zero. Notice that by comparing the *normal* and *intrinsic* characteristics, the values of $-H_c$ and $-H_{ci}$ are not the same, that is, the magnetizing force H_c required to make $B = 0$ may be less than that required to reduce the net observable flux density B to zero, namely H_{ci}, after which the material's magnetization direction is reversed.
- Maximum energy product BH_{max} is the point on the second quadrant *B versus H* curve at which the product of *B* and *H* for the magnet is a maximum. On the *ideal* curve in figure 27.14, it is located exactly halfway down the second quadrant line, with a value of $-BH_{max} = \mu_o(½M_{sat})^2$.

Generally $\mu_o M_s = J_s > B_r$, $B_r \geq \mu_o H_c$, $H_{ci} \geq H_c$, and $¼B_r^2 \geq \mu_o BH_{max}$. The slope of the normal curve at the *B*-axis intercept is called the recoil permeability, $\mu_o\mu_{rc}$. The third and fourth quadrants are the same as the first and second quadrants, except have the 'opposite' polarities. The whole curve forms the hysteresis loop

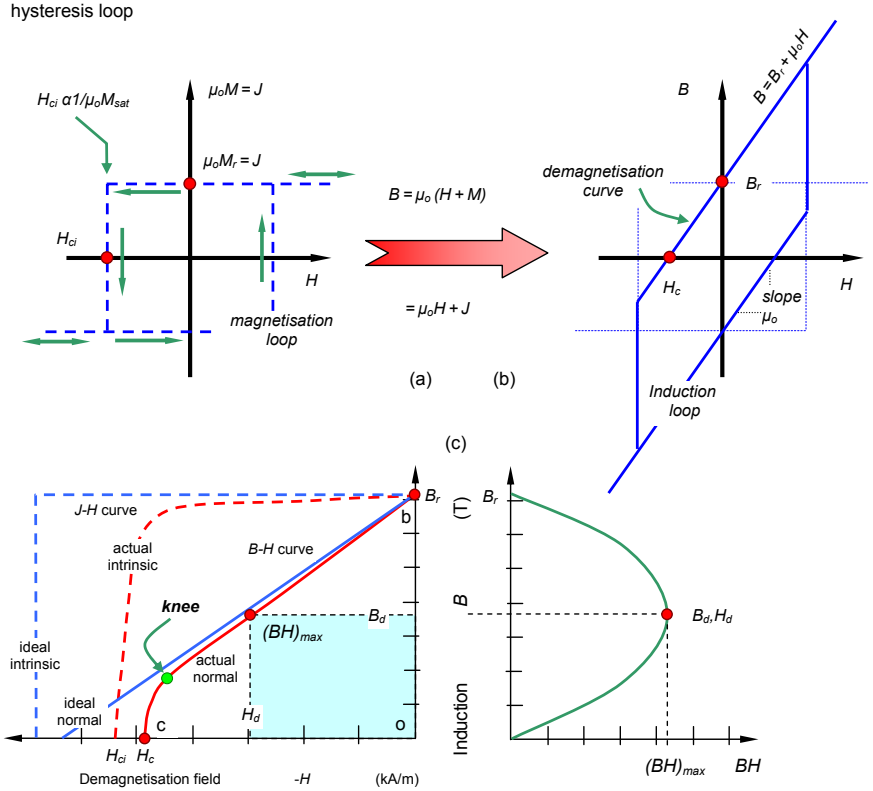


Figure 27.14. *Magnetic material curves: (a) ideal magnetisation loop, (b) ideal induction loop, and (c) second quadrant hysteresis loop for a sintered magnet.*

Several characteristics have to be taken into consideration when deciding which permanent magnetic material to use, namely:

- Flux related requirements of
  - Energy stored in the material;
  - Flux Density on the surface of the magnet; and
  - Pull required to remove the magnet from a flat piece of steel.
- Maximum operating temperature.
- Cost.
- Availability.
- Degree of corrosion likely to be encountered.
- Magnetic stability required.
- Size and/or weight limitations.

General relative characteristics are shown in figure 27.15 and the following design points are related sequentially to the data columns in Table 27.5.

i. A method of comparing the magnetic performance or the capability of different types and grades of permanent magnet is to consider their maximum energy product BH_{max}. This is the operating point where the magnet will provide most energy for a minimum volume. When the permanent magnet behaves linearly with a constant recoil permeability μ_{rc} the maximum energy product can be expressed simply by:

$$BH_{\max} = \frac{B_r^2}{4\mu_{rc}} \qquad \text{J/m}^3 \tag{27.5}$$

A material with the largest available maximum energy product results in the smallest required magnet volume. Maximum energy products for various hard magnetic materials are shown in Table 27.5.
Ignoring leakage flux, for a permanent magnet with an air gap, the air gap flux density and mmf's are:

$$B_g = \frac{A_m}{A_g} B_m \tag{27.6}$$

$$\frac{H_m \ell_m}{H_g \ell_g} = -1 \tag{27.7}$$

Combining gives $$B_g^2 = \mu_o \frac{\ell_m A_m}{\ell_g A_g}\left(-H_m B_m\right) = \mu_o \frac{Vol_{mag}}{Vol_{air-gap}}\left(-H_m B_m\right) \tag{27.8}$$

where $B_g = \mu_o H_g$. That is

$$Volume_{mag} = Volume_{air-gap} \frac{B_g^2}{\mu_o} \frac{1}{\left(-H_m B_m\right)} \tag{27.9}$$

Equation (27.9) indicates that, to achieve a desired flux density in the air gap, the required magnet volume can be minimized by operating the magnet at the point of maximum energy product.

ii. Table 27.5 shows typical pole face flux densities for the various hard magnetic materials when operating near their BH_{max} point. This point is not at the remanence B_r, which is the induction under closed magnetic circuit conditions.

iii. The attraction force F between a permanent magnet and a soft magnetic body depends on three factors:

B = flux density on the pole face
A = area of the pole face
μ = permeability of the material being attracted

and the force is given by:

$$F = B^2 A \mu \tag{27.10}$$

iv. There are two distinct temperature related loss categories, namely reversible and irreversible losses.
The reversible changes with temperature are dependant on material composition and are unaffected by shape, size or working point on the demagnetisation curve. This loss vanishes without need for remagnetisation when the magnet returns to its initial temperature.
Irreversible losses do not occur until a certain temperature, the Curie temperature T_C, has been exceeded. Unrecoverable losses occur when excessive temperatures are reached and metallurgical changes occur within magnet. This loss can also be limited by operating at a high as possible working point (above the *knee*, near B_r).

v. The maximum working temperature, before irreversible losses commence, is dependant on the magnet working point in the circuit. The higher the working point up the B-H demagnetising characteristic curve, the higher the temperature at which the magnet can operate. Irreversible losses can be recovered by remagnetising the magnet.

vi. Each magnetic material has a maximum temperature, above the Curie temperature, where metallurgical changes occur within the magnet structure and where the individual magnetic domains randomise due to the high level of thermal agitation. Once these losses occur they cannot be reversed by remagnetising.

vii. The effects of low temperatures are different for each material group and are also related to the magnet shape and therefore its working point on its demagnetization curve.

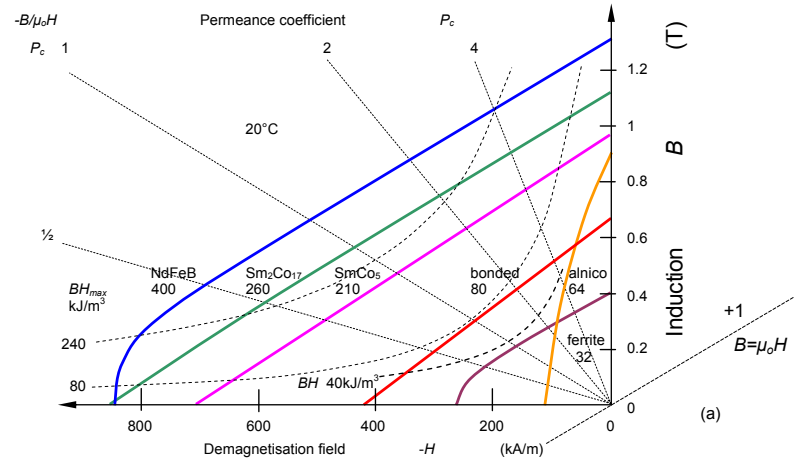
viii. The temperature has a significant effect on magnet stability, but exposure to high external fields, H, influences magnets as follows:

- It is possible to ensure magnetic stability, by pre-exposing the magnet to possible detrimental influences by thermal cycling and controlled demagnetizing fields (ageing) techniques.
- Another cause of performance loss is total composition structural breakdown due to corrosion or in the case of NdFeB, exposure to hydrogen.

ix. There are many protective coatings used to help prevent corrosion in magnets.
Samarium cobalt, alnico, and ceramic materials are corrosion resistant, and typically do not require to be coated against corrosion.
Alnico is easily coating for cosmetic purposes with a powder coating or electroplating when required.
Ceramics may be coated to seal the surface, which will otherwise be covered by a thin film of ferrite powder, which does not create a problem in most applications.
Magnets using rare earth (non-precious) elements are highly reactive at elevated operating temperatures in the presence of moisture, due to their strong negative electrochemical standard potential, between -2.2V and -2.5V. Rare-earth-hydroxide is formed, releasing hydrogen, which then reacts with free rare earth metal, forming rare earth metal hydrides. With $SmCo_5$ or Sm_2Co_{17}, by adding cobalt, the reaction with water is suppressed.
In the case of NdFeB magnets, which are particularly susceptible to corrosion, the individual magnet grains are held together mechanically and fixed to each other by the so-called neodymium-rich phase, so the neodymium reacts with water to form neodymium hydroxide. Addition of cobalt to the neodymium-rich phase improves the corrosion behaviour by systematically stopped intergranular corrosion in a warm, humid atmosphere. The corrosion behaviour of such alloys is similar to that of pure iron materials (steel). In cases where the humidity turns to condensation, the materials gradually rust forming non-magnetic metal oxides or hydroxide, similarly to red rust in the case of iron. In applications with possible condensation, hard magnetic parts should be coated. Spray coatings are either metallic (galvanic nickel 10µm thick and/or tin 15 to 30µm thick, or 5µm of electrically isolating aluminium –aluminium chromate vacuum deposited) or organic (5µm and 40µm of epoxy resin, lacquer, parylene or polyxylylene polymers.

x. Shape, tolerances, and quantity influence the prices of specific magnets but the most significant factor is the cost of the basic raw material.

Other design factors include time (magnetic creepage, energy loss of typically less than 1×10^{-5} pu per annum at 20°C), radiation, shape, weight (for example, NdFeB magnets are about 13% lighter than SmCo magnets), and shock and vibration. All these are factors affecting magnet stability. Stability can be described as the repeated magnetic performance of a material under specific conditions over the life of the magnet.



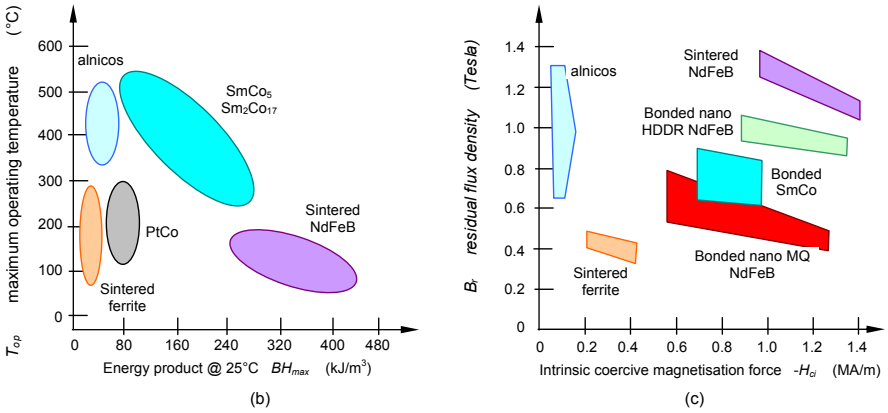


Figure 27.15. *Comparison of different magnetic materials: (a) second quadrant hysteresis loop demagnetization characteristics (with constant energy contours shown), (b) energy product versus maximum operating temperature, and (c) remanence versus intrinsic coercive force.*

Summary of 2nd quadrant demagnetisation characteristics of different hard magnetic materials

Table 27.5 Hard magnetic material characteristics and properties

Material	structure	Energy BH_{max}/ *recoil* μ_{rc}	fux density/ coercivity	pull	α_r	T_{op}	T_C	Low T_{op}	magnetic stability	corrosion resistant	cost relative
		kJm^{-3}	T / kA/m	kg/cm^2	% / °C 20°C - 150°C	°C	°C		Thermal α/B_r %/K		
		i	ii.	iii.	iv	v	vi	vii	viii	ix	x
Ferrite	anisotropic	24/1.15	0.5/180 0.1T typ	0.26	- 0.19	250	460	Large irreversible losses below - 60 °C	poor -0.2	excellent	low x1
Ferrite	bonded	3-8/1.1	0.2/130	0.06			450				
Alnico	anisotropic	40/3	1.0/120 0.13T typ	0.42	- 0.02	540	860	Permanent losses of no more than 10% are to be expected down to 4K	excellent -0.02	fair	medium x5
Alnico	casted	38/	0.7/100	0.42		540	850				
SmCo	Sintered 1:5	144/1.05	1.0/ 0.28T typ	1.50	- 0.04	250			excellent -0.035	excellent	
SmCo	Sintered 2:17	200/1.1	1.2/800 0.32mT typ	2.08	- 0.03	300	750	Minimal losses down to 4K	excellent -0.03	excellent	very high x20
SmCo	Bonded 1:5	55/1.1	0.7/750	0.60			870				
NdFeB	N38H	256/1.05	1.3/800 0.4T typ	3.06	- 0.12	50-200	320	No irreversible losses down to 77K	moderate -0.13	poor	high x10
NdFeB	bonded	110/1.1	1/800	1.20			310				
ferrite	flexible	5/	0.16/110			100					low

27.4 Permanent Magnet Magnetization Curve (hysteresis loop) and recoil

The magnetic circuit in figure 27.16a includes a section of un-magnetised hard magnetic material in a core of highly permeable soft magnetic material and an *N*-turns excitation winding. In figure 27.16c, it is assumed that the hard magnetic material is initially un-magnetized (corresponding to point *a*, the origin, in the figures 27.16a and b) and current is applied to the excitation winding. Because the core is assumed to be of infinite permeability, the horizontal axis of figure 27.16c can be considered to be both a measure of the applied current $i = H\ell_m/N$ as well as the *H* field intensity in the magnetic material.

The magnetic circuit of figure 27.16a can be used to magnetize hard magnetic materials. The process requires a large excitation current be provided through the winding, thereby attaining point *b*, which is then reduced to zero, leaving the material with a remanent magnetization (or residual induction) B_r (point *c* in figure 27.16c).

With reference to figure 27.16c, the magnetization process progresses through four stages:

Stage 1. The permanent magnet starts in an initially unmagnetized state at field intensity $H = 0$ and flux density $B = 0$, point '*a*' in figure 27.16c.

Stage 2. As the current *i* in the coil is increased to a maximum, i_{max}, the *B-H* trajectory rises from point *a* in figure 27.16c to a maximum (H_{sat}, B_{sat}) at point *b*. It is assumed that driving the core well into saturation with field strength H_{sat} at point *b*, fully magnetizes the hard magnetic material.

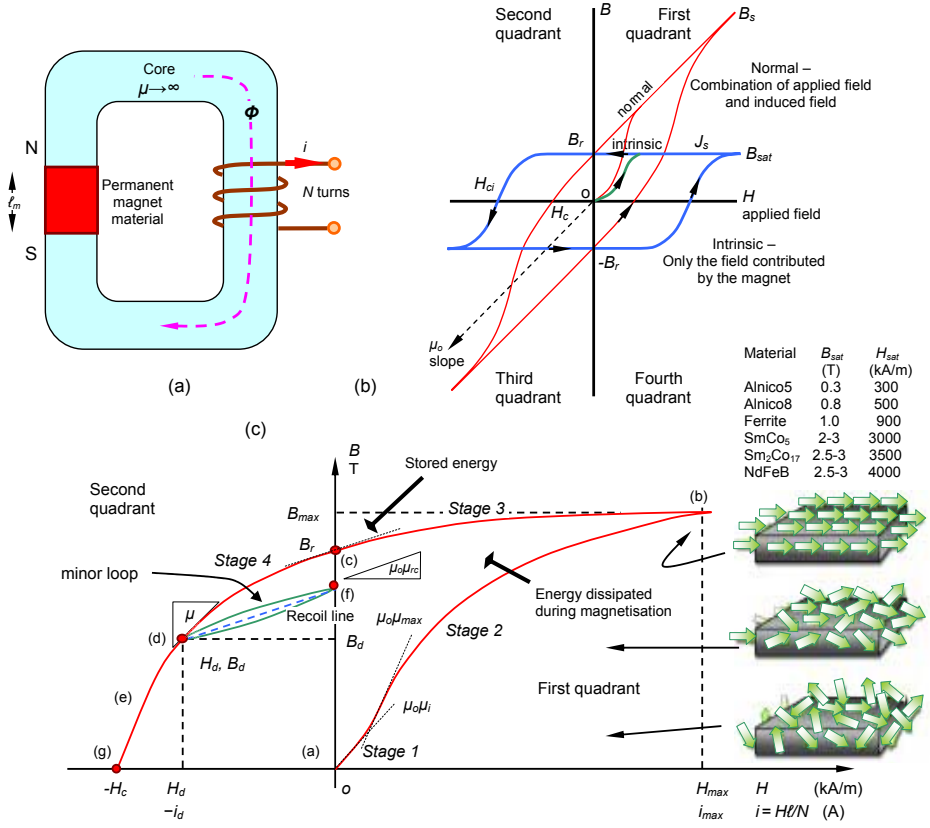


Figure 27.16. *Magnetisation of a permanent magnet:*
(a) magnetic circuit incorporating a permanent magnet and an excitation winding, (b) four quadrant B-H characteristics, and (c) portion of a B-H characteristic showing a minor loop, a recoil line and magnetic induction B_{sat} required to magnetise various permanent magnet materials.

Stage 3. Subsequently, when the coil current is decreased to zero (the external applied magnetising field intensity decreases to zero); the flux density from within the permanent magnet material also decreases. The *B-H* characteristic traverses part of a hysteresis loop, moving to point *c* at zero current. At point *c*, the applied field intensity *H* in the material is zero but the flux density *B* is at its remanent value B_r. Through hysteresis mechanisms, the magnet retains flux density B_r (retentivity if the magnetising force reached saturation - residual flux density if the magnetizing force did not reach saturation).

Stage 4. With the introduction of an air gap and/or an external demagnetizing field (current direction in the coil is reversed), the magnet operates within the second quadrant of the demagnetization curve, where *H* is negative and *B* is positive. The slope of the demagnetization curve in the second quadrant, as shown in figure 27.16c, is μ_m. Permeability μ_m is typically close to μ_o, the magnetic permeability of free space. With increased current, the intersection of the demagnetization curve with $B = 0$ occurs at $H = -H_c$, point *g*, where H_c is known as the coercivity (coercive force) of the magnetic material.

Assuming a linear demagnetising curve, the magnetic circuit second quadrant operating point should conform to the following mathematical relationship between flux density, *B*, and field intensity, *H*:

$$B = B_r + \mu_m H \qquad (27.11)$$

The created (hysteresis) loops are a graphical representation of the relationship between an applied magnetic field and the resulting induced magnetization within a material. The field that is generated by the now magnetized material (orientated domains), B_i, when added to that of the applied field, *H*, is known as the normal induction, B_n, or simply *B*. This induction has two components (*normal* and *intrinsic* curves), and they are related by:

$$B_n = B = \mu_o M + \mu_o H = J + \mu_o H \qquad (27.12)$$

The *B* versus *H* curve is known as the *normal* curve, while the *J* versus *H* curve is called the *intrinsic* curve (polarisation $J=\mu_o M$). These curves, also called hysteresis loops, are shown in figure 27.16b. The loops show the properties of the magnetic material as it is cycled between magnetization (saturated) and demagnetization (under the influence of an external magnetic field). The second (and third) quadrant of the loops display the magnetic properties as the magnet performs work.

When the current is reversed from zero to $-i_d$, the *B-H* characteristic operating point traverses along a hysteresis loop with a trajectory from point *c* to point *d*. If the current is maintained at $-i_d$, the magnet operating point is maintained at point *d*, where H_d, B_d generally denotes the operating point on the second quadrant demagnetising curve. This same operating point would be attained if the material were to start at point *c*, and with zero current excitation, from equation (27.8), an air gap of length $\ell_g = \ell_m (A_g / A_m)(-\mu_o H_d / B_d)$ were introduced into the core. If the current becomes more negative, the trajectory continues to trace out the hysteresis loop toward point *e*. If instead of increasing the current magnitude, it is reduced from $-i_d$ to zero, the trajectory does not retrace the hysteresis loop toward point *c*. Rather it inscribes a minor hysteresis loop, reaching point *f* at zero current. If the current is then varied between zero and $-i_d$, the *B-H* characteristic operating point traces out the minor loop shown in figure 27.16c. Unlike soft magnetic material, the absence of an external magnetic field does not lead to demagnetization.

The *B-H* trajectory between points *d* and *f* can be represented by a straight line, termed the *recoil line*, with a slope $\mu_o\mu_{rc}$, where μ_{rc} is defined as the relative recoil permeability. At the vertex of the minor hysteresis loop, point *d,* the material has been partially demagnetized, with the effective remanent magnetization of the magnetic material having been reduced to that represented by point *f*, which is less than the remanent magnetization B_r, point *c*. If the demagnetization is increased past point *d*, to point *e* of figure 27.16c, a new minor loop will be created, with a new recoil line but with a similar recoil permeability (slope $\mu_o\mu_{rc}$). If the demagnetizing field is increased beyond point *g*, H_c, the operating point of the magnet now moves into the third quadrant of the normal curve. Recoil still results in a positive remanence flux density. Before the intrinsic coercivity H_{ci} is reached, which is show in figure 27.16b, the magnet becomes completely demagnetized, since recoil is back to the origin, point *a*. When H_{ci} is reached and exceeded, the recoil flux density is negative, below point *a*. Intrinsic coercivity H_{ci} is therefore a measure of a magnet's ability to resist demagnetization.

The demagnetization effects of negative current excitation are equivalent to the introduction of an air gap into the magnetic circuit (up to an operating point, H_{ci}, 0 on the intrinsic curve). If the magnetized material were removed from the core, this is equivalent to creating a large air gap in the magnetic circuit, therein demagnetizing the hard magnetic material. The magnet is effectively weakened, since if it were re-inserted into the magnetic core, it would follow a recoil line and return to a remanent magnetization operating point (to point *f*) less than B_r. Hard magnetic materials such as Alnico materials, often do not operate stably with varying mmf and geometry conditions, since they can be significantly demagnetize due to improper operation. An advantage of magnetic materials such as ceramic ferrite, samarium-cobalt and neodymium-iron-boron is that, because of their 'straight-line' characteristic in the second

quadrant (with slope close to μ_o), their recoil lines closely match their magnetization characteristic. As a result, demagnetization effects are reduced. At the expense of a modest reduction of the remanent magnetization, hard magnetic materials can be stabilized to operate over a specified field region.

27.5 Permanent Magnet model

When a magnetic field is imposed on a permanent magnet, intrinsic induction *J* is created. The free poles establish a field potential $-H_d$ between free poles. The field potential is due to some of the magnetization *J* lost internally. The field potential $-H_d$ associated with a permanent magnet is a due to the magnetization *J* and is 180° opposed to *J*. The magnitude of $-H_d$ depends on the geometry of the magnet including the spacing of the poles. These relationships are shown pictorially in figure 27.17.

The behaviour of a permanent magnet is described with reference to the second quadrant of its *B-H* curve, termed the demagnetization curve. For reference, the *ideal* curves in figure 27.17 are shown in blue. A practical permanent magnet material neither achieves its theoretical intrinsic coercivity $-H_{ci}$, nor does the entire magnetization flip over exactly when a reverse field of $-H_{ci}$ is reached. The practical intrinsic demagnetization curve, shown dashed red in figure 27.17 does not have an abrupt transition at $-H_{ci}$, but rather a *knee* in the curve represents a more gradual reversal of the material's magnetization. The actual normal demagnetization curve, shown solid red in figure 27.17 mirrors this *knee*. The coercivities $-H_{ci}$ and $-H_c$ are now defined as the intercepts of the practical intrinsic and normal curves with the *H*-axis.

The *B* from within the magnet is the flux density that can be delivered into the adjacent air gap, and the point at which a magnet operates on its demagnetization curve relates *B* to the experienced demagnetizing force *-H*. The demagnetization curve shows that as the magnitude of *-H* increases, the flux density produced by the magnet falls, ultimately at $H = -H_c$ to $B = 0$. However, before $-H_{ci}$ is reached, the *knee* in the demagnetization curve signifies the onset of a reversal of the material's magnetisation *M*. It is therefore desirable that the operating point of a permanent magnet remain above any *knee* in the demagnetization curve - at all operating temperatures.

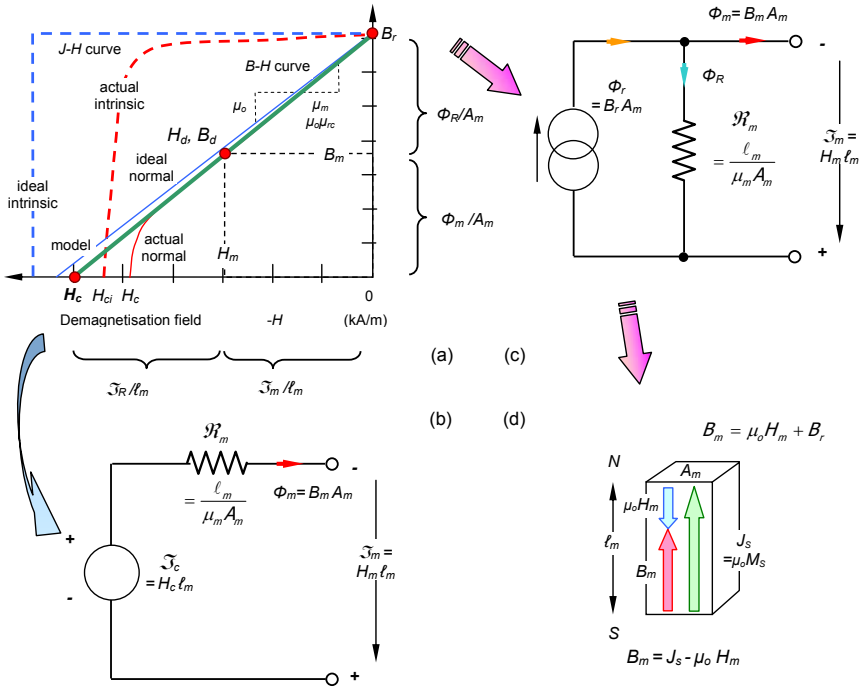


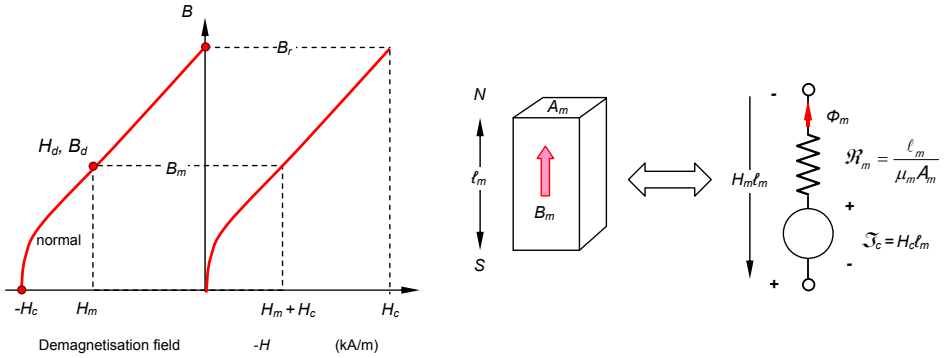
Figure 27.17. *Magnetic circuit models of a magnet based on: (a) a linear demagnetization curve, (b) mmf source equivalent series circuit, (c) flux source equivalent shunt circuit, and (d) magnet internal fluxes.*

Consider a permanent magnet of uniform cross sectional area A_m and length ℓ_m. The demagnetization curve of the magnet is a straight line with a coercive force of H_c and a remanent flux density of B_r as shown in figure 27.17. A point H_d, B_d on a linear demagnetization B_m - H_m curve can be defined by

$$\boldsymbol{B_m} = B_r + \mu_m \boldsymbol{H_m} = \frac{B_r}{H_c}\left(H_c + \boldsymbol{H_m}\right) = \mu_m \left(H_c + \boldsymbol{H_m}\right) \qquad (27.13)$$

where $\mu_m = B_r / H_c$ is the permeability of the permanent magnet, which is slightly greater than μ_o, the permeability of free space. For NdFeB and $SmCo_5$ sintered magnets, μ_m=1.05 μ_o, as listed in Table 27.5.

Equation (27.13) shows that the demagnetisation curve can be defined in terms of μ_m and either B_r or H_c. If H_c is used as a reference, the model will be based on an mmf source $H_c\ell_m$ in series with the magnet reluctance $\mathcal{R}_m$. If B_r is used as a reference, the model will be based on a flux source B_rA_m in parallel with the magnet reluctance $\mathcal{R}_m$. The two models are equivalent, and in electrical circuit terms, behave as Thevenin and Norton equivalent circuits, as illustrated in figure 27.17.



27.18. *Magnetic circuit Thevenin model of a magnet with nonlinear demagnetization curve.*

Thevenin's equivalent magnetic circuit

The traditional, less intuitive modelling approach is based on an mmf source in series with the magnet reluctance. From equation (27.13) rearranged, the magnetic mmf across the magnet, using flux $\Phi_m = B_mA_m$, can be expressed as

$$\boldsymbol{H_m}\ell_m = \left(\frac{\boldsymbol{B_m}}{\mu_m} - H_c\right)\ell_m = \frac{\ell_m}{\mu_m A_m}\boldsymbol{\phi_m} - H_c\ell_m = \mathcal{R}_m\boldsymbol{\phi_m} - \mathcal{F}_c \qquad (27.14)$$

$\mathcal{R}_m = \ell_m / \mu_m A_m$ is the magnet reluctance and $\mathcal{F}_c = H_c\ell_m$ the magnetomotive force (energy source) of the magnet. In a magnet, B_m and H_m, being in the second quadrant, oppose. The series equivalent circuit satisfying the sum of mmf's around a closed loop summing to zero, is shown in figure 27.17b.

For a magnet with a nonlinear demagnetization curve, the magnetic circuit model in figure 27.17 remains valid, except that the magnetic permeability is

$$\mu_m = \frac{\boldsymbol{B_m}}{\boldsymbol{H_m} + H_c} \qquad (27.15)$$

which is a function of the magnetic field in the magnet. The magnetic circuit model derivation of a nonlinear magnet is illustrated graphically in figure 27.18.

The operating point (H_d, B_d) does not move along the nonlinear demagnetization curve if a small cyclic external magnetic field (such that the magnet will not be demagnetized) is applied to the magnet. Instead, the operating point moves along a minor loop or for simplicity, a straight line (centre line of the minor loop) as illustrated in figure 27.19.

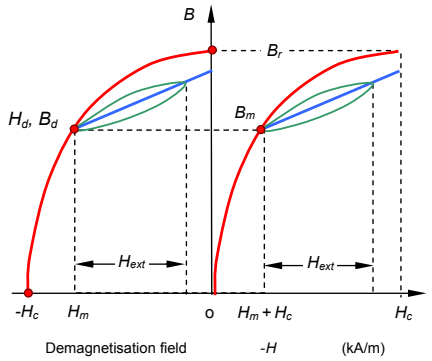


Figure 27.19. *Movement of operating point of a nonlinear magnet under an external field H_{ext}.*

Norton's equivalent magnetic circuit

From equation (27.13) rearranged, the magnetic flux from magnet, using flux $\Phi_m = B_mA_m$, can be expressed as

$$\boldsymbol{B_m}A_m = B_rA_m + \mu_m\boldsymbol{H_m}A_m = B_rA_m + \frac{\mu_m A_m}{\ell_m}\boldsymbol{H_m}\ell_m = B_rA_m + \frac{1}{\mathcal{R}_m}\boldsymbol{H_m}\ell_m$$
$$\boldsymbol{\phi_m} = \phi_r + \frac{\boldsymbol{\mathcal{F}_m}}{\mathcal{R}_m} \qquad (27.16)$$

$\mathcal{R}_m = \ell_m / \mu_m A_m$ is the magnet reluctance, $\boldsymbol{\mathcal{F}_m} = \boldsymbol{H_m}\ell_m$ the magnetomotive force across the magnet, and Φ_r is the energy source. In a magnet, B_m and H_m, being in the second quadrant, oppose. The equivalent circuit satisfying the three fluxes at a node summing to zero, is shown in figure 27.17c.

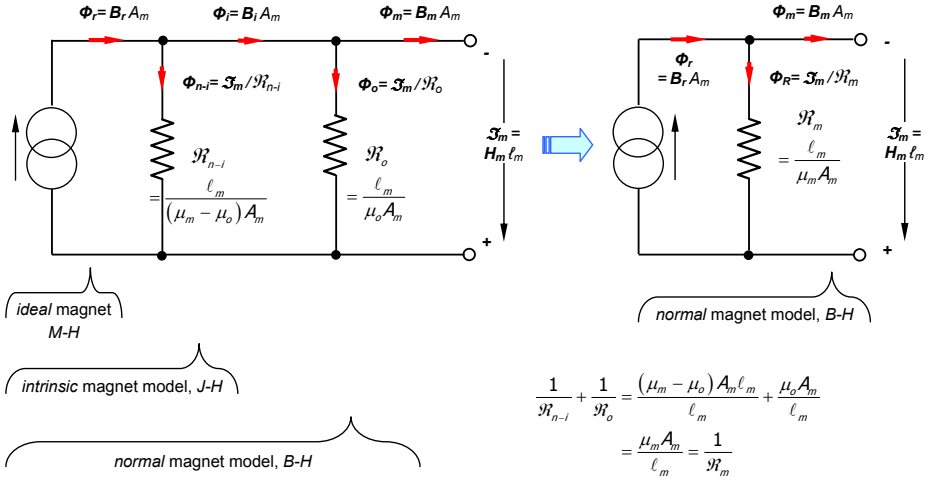


Figure 27.20. *Magnet internal model showing an intrinsic magnet model.*

In the case of NdFeB and SmCo magnets, there are several advantages in using a flux source model based on B_r rather than the series mmf model based on H_c.

- B_r is common to both the intrinsic and normal demagnetisation curves (H_c and H_{ci} differ).
- At a given temperature, the demagnetisation curve slope is near constant, $1/\mathcal{R}_m$, for an increasing reverse field, until a knee is reached. (With H_c as a reference, if a knee exists in the demagnetisation curve, the implied B_r is under estimated if a straight line demagnetising characteristic is used for modelling or analysis).

- A magnet is normally operated in the near linear region above any knee - B_r as a reference models this region independent of the existence and effects of any knee.
- The current source model in figure 27.20 clearly illustrates with two reluctances in parallel that the mmf, hence field intensity, must be the same value for the load operating point on the normal and intrinsic demagnetisation curves.

Both models yield the same mathematical solutions, since both yield the same demagnetisation load line equation, when using $B_r = \mu_m H_c$:

$$A_m \boldsymbol{B_m} = A_m B_r + \frac{1}{\mathcal{R}_m} \boldsymbol{H_m} \ell_m = \mu_m A_m H_c + \frac{1}{\mathcal{R}_m} \boldsymbol{H_m} \ell_m \qquad (27.17)$$

$$or \qquad \boldsymbol{B_m} = B_r + \mu_m \boldsymbol{H_m}$$

27.6 Load lines

One or more air gaps introduced into a magnetic circuit enable useful work to be exploited.
Magnet material can be designed to provide a specific air gap flux density, within a reasonable range. The exact flux density is determined by the magnetic circuit dimensions, particularly those of the permanent magnet and the air gap. The second quadrant of the magnet demagnetization *B-H* curve gives magnetic properties per unit magnet volume. The external magnetic circuit, termed the *magnetic load*, together with the demagnetization curve, specify the magnet's operating point on its demagnetisation curve.
Consider the magnetic circuit shown in figure 27.21a, where a permanent magnet, with magnetization *M* oriented as shown, drives flux clockwise into an air gap via two steel high permeability pole pieces. The steel high permeability, in conjunction with their shape, confines most of the magnetic flux from the magnet (area A_m, length ℓ_m) into the air gap (area A_g, length ℓ_g). The lines of magnetic flux are shown in figure 27.21b, and magnified in figure 27.21c, where leakage fluxes outside the air gap volume, termed fringing, are shown.

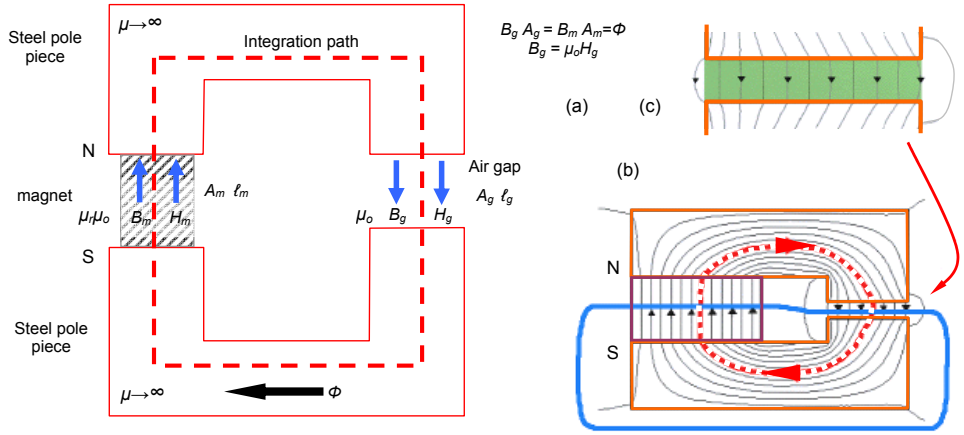


Figure 27.21. *Magnet in an iron circuit with an air gap:*
(a) magnetic circuit, (b) magnetic circuit flux lines, and (c) air gap flux and its leakage.

27.6.1 Magnetic Circuit Equations

Flux conservation gives the integral of flux density through a closed surface (equivalent to Kirchhoff's current law), while Ampère's law yields the integral of magnetizing force around a closed loop (equivalent to Kirchhoff's voltage law). These can be applied to the magnetic circuit shown in figure 27.21a, to derive its load line and hence find the flux density levels within the magnetic circuit. Any convenient surface and loop may be chosen to achieve these. The subscript *m* denotes the permanent magnet and *g* denotes the air gap, while *s* is for both steel pole pieces, and ℓ denotes the external regions into which some magnetic field leaks.

The magnetizing force $\boldsymbol{H}$ is integrated around the closed loop in the direction shown in red. No electrical currents passing through any surface spanning this loop, so $i = 0$. From figure 27.21a, as with Kirchhoff's voltage law around a closed loop, the closed integral has three series components:

$$\int_m H.d\ell + \int_s H.d\ell + \int_g H.d\ell = 0$$

$$\boldsymbol{H_m}\ell_m + \boldsymbol{H_s}\ell_s + \boldsymbol{H_g}\ell_g = 0$$

Since the steel pole pieces have a significantly higher permeability than either the magnet or the air gap, their mmf contribution is small, $\boldsymbol{H_s}\ell_s = 0$. It is therefore a simplification to either ignore this small component, $k_r = 1$ or to lump together the two external terms using a dimensionless *loss or reluctance factor* k_r, where $k_r > 1$. Outside the magnet, the magnetisation is zero, that is, $M = 0$. Each individual $\boldsymbol{H}\ell$ term is a magnetomotive force (mmf), $H_m\ell_m$ being the mmf across the permanent magnet and $\boldsymbol{H_g}\ell_g$ the mmf across the air gap.

$$\boldsymbol{H_m}\ell_m + k_r \boldsymbol{H_g}\ell_g = 0$$

Rearranging yields

$$\boldsymbol{H_m} = -k_r \frac{\boldsymbol{H_g}\ell_g}{\ell_m} \qquad (27.18)$$

Integration of the flux density $\boldsymbol{B}$ through the closed surface shown in blue (Kirchhoff's current law), yields:

$$\boldsymbol{B_m}A_m - \boldsymbol{B_\ell}A_\ell - \boldsymbol{B_g}A_g = 0$$

Each term is the flux in the respective magnetic component, $B_m A_m$ being the flux from the permanent magnet, $\boldsymbol{B_g} A_g$ the flux in the air gap, and $\boldsymbol{B_\ell} A_\ell$ the total leakage flux. The magnet and air gap volumes are well defined, while the leakage regions are not. Because of the relatively high permeance of the steel, only a small percentage of the magnet flux does not pass directly through the air gap. A simplification is to incorporate the $\boldsymbol{B_\ell} A_\ell$ term into a dimensionless *leakage coefficient* k_ℓ, where $k_\ell > 1$, such that:

$$\boldsymbol{B_m}A_m - k_\ell \boldsymbol{B_g}A_g = 0$$

After lumping together the two fluxes external to the magnet, which are dominated by the gap flux, rearranging gives

$$\boldsymbol{B_m} = k_\ell \frac{\boldsymbol{B_g}A_g}{A_m} \qquad (27.19)$$

The permanent magnet's demagnetization curve, shown in red in figure 27.22, gives its magnetic properties per unit volume as a characteristic relationship between B_m and H_m. The load line (in blue) gives the characteristics of the magnetic circuit, also in terms of B_m and H_m. The operating point (H_d, B_d) is the intersection of the two lines.

Since $\boldsymbol{B_g} = \mu_o \boldsymbol{H_g}$ in the air gap (where $M = 0$), equations (27.18) and (27.19) can be combined to give the load line (of slope $-\mu_o P_c = -S$) as:

$$\frac{\boldsymbol{B_m}}{\boldsymbol{H_m}} = -\mu_o \frac{k_\ell}{k_r}\frac{A_g\ell_m}{A_m\ell_g} = -\mu_o P_c = -S$$

where $P_c = \dfrac{A_g\ell_m}{A_m\ell_g}$ and k_ℓ / k_r is assumed to be unity.

That is

$$\boldsymbol{B_m} = -\mu_o \frac{A_g\ell_m}{A_m\ell_g}\boldsymbol{H_m} = -\mu_o P_c \boldsymbol{H_m} = -S\boldsymbol{H_m} \qquad (27.20)$$

The intercept between the demagnetization curve ($B_m = B_r + \mu_m \boldsymbol{H_m}$) and the load line, equation (27.20), gives the magnet operating point B_d and H_d. The flux density and magnetizing force elsewhere in the magnetic circuit can then be derived from the foregoing equations, assuming $k_r = k_\ell = 1$.

$$\boldsymbol{B_m} = \underbrace{B_r + \mu_m \boldsymbol{H_m} = \mu_m H_c + \mu_m \boldsymbol{H_m}}_{\text{Second quadrant demagnetising curve}} = \underbrace{-\mu_o \frac{A_g}{A_m}\frac{\ell_m}{\ell_g}\boldsymbol{H_m}}_{\text{Load line}} = -S\boldsymbol{H_m} \qquad (27.21)$$

The second and fifth equalities give

$$H_d = -\frac{B_r}{S+\mu_m} = -\frac{\mu_m H_c}{S+\mu_m} \tag{27.22}$$

From the first and last equalities in equation (27.21), on substitution of H_m from equation (27.22)

$$B_d = -SH_m = B_r \frac{S}{S+\mu_m} \tag{27.23}$$

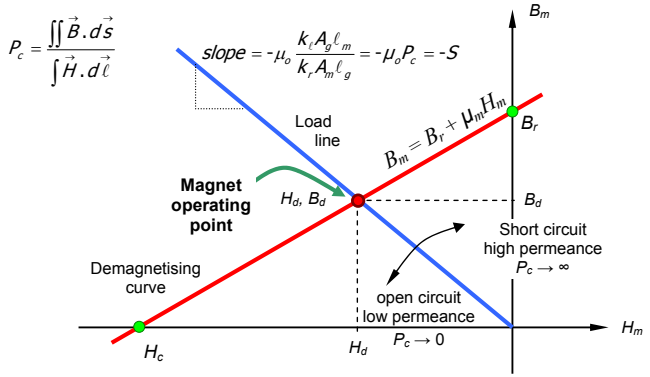


Figure 27.22. *Load line for a permanent magnet, air gap, and iron circuit.*

From equation (27.19), the gap flux density $\boldsymbol{B}_g$, which can be used to calculate the gap energy, is

$$\boldsymbol{B}_g = \frac{A_m}{A_g}\boldsymbol{B}_m = \frac{A_m}{A_g} B_r \frac{S}{S+\mu_m} \tag{27.24}$$

As the gap closes to zero (soft magnetic material forms at least one closed path), $\ell_g \to 0$, with $A_g = A_m$ and $S \to \infty$:

$$H_m \to 0 \quad \text{and} \quad B_m \to B_r$$

As the gap goes to infinity (open circuit - no soft magnetic material), $\ell_g \to \infty$ and $S \to 0$:

$$H_m \to -H_c \quad \text{and} \quad B_m \to 0$$

The open circuit magnetic flux intensity specifies the Thevenin equivalent circuit mmf source, namely

$$\mathfrak{I}_c = H_c \ell_m$$

From the first and third equalities in equation (27.22), rearranged:

$$\boldsymbol{H}_m = -H_c + \frac{\boldsymbol{B}_m}{\mu_m} \tag{27.25}$$

Equation (27.18), in conjunction with the fact that the mmf's in the closed magnetic circuit loop are defined by multiply equation (27.25) by magnet length ℓ_m, gives

$$\boldsymbol{H}_m \ell_m = -H_c \ell_m + \frac{\boldsymbol{B}_m}{\mu_m}\ell_m$$

$$-\boldsymbol{H}_g \ell_g = -H_c \ell_m + \frac{\boldsymbol{B}_m}{\mu_m}\ell_m$$

Rearranging gives

$$H_c \ell_m = \boldsymbol{H}_g \ell_g + \frac{\boldsymbol{B}_m}{\mu_m}\ell_m \tag{27.26}$$

Assuming no leakage flux, then using $\boldsymbol{\Phi} = \boldsymbol{B_g} A_g = \boldsymbol{B_m} A_m$ this equation reduces to

$$\mathfrak{I}_c = H_c \ell_m = \boldsymbol{\phi}\left(\mathfrak{R}_g + \mathfrak{R}_m\right)$$

where the gap (external magnetic load circuit) reluctance (whence gap permanence) is

$$\mathfrak{R}_g = \frac{\ell_g}{\mu_o A_g} = \frac{1}{P_g}$$

and the (Thevenin/Norton equivalent) reluctance (whence permanence) of the magnet, where $\mu_m = B_r / H_c$, is

$$\mathfrak{R}_m = \frac{\ell_m}{\mu_m A_m} = \frac{1}{P_m} \tag{27.27}$$

From equations (27.18) and (27.19), the magnet volume, V_m, is

$$V_m = \ell_m A_m = \frac{\ell_m \boldsymbol{H_g}^2 A_g}{B_m H_m}$$

Note: the magnet B_m-H_m curve intersects the general load line at the operating point H_d, B_d, and that BH_{max} will minimise the necessary magnet volume.

27.6.2 Intrinsic permeance coefficient

From equation (27.20), the dimensionless permeance coefficient P_c is:

$$P_c = \frac{B_m}{\mu_o H_m} = \frac{A_g \ell_m}{A_m \ell_g} = \frac{\mu_m}{\mu_o}\frac{\mathfrak{R}_m}{\mathfrak{R}_g} = \frac{\mu_m}{\mu_o}\frac{P_g}{P_m} = \frac{1}{N_i}$$

$$S = \mu_o P_c = \frac{B_m}{H_m} = \mu_m \frac{\mathfrak{R}_m}{\mathfrak{R}_g} = \mu_m \frac{P_g}{P_m}$$

This permeance coefficient is also known as the *normal* permeance coefficient because it provides the slope for a line that originates at the origin and intersects the *normal* demagnetisation curve at the magnet operating point (H_d, B_d). This operating condition point of the magnet also leads to the operating point on the *intrinsic* curve (provided $i = 0$).

The intrinsic operating point (H_d, B_{di}), has an intrinsic permeance coefficient P_{ci} which is obtained by adding 1 to the P_c value, that is, $P_{ci} = P_c + 1$.This is because, from equation (27.12), $J = \mu_o(\mu_m - 1)H$. The intrinsic demagnetising factor N_i is defined as $N_i = 1/P_{ci}$, thus both factors are determined by the shape of the magnetic circuit.

The P_c and P_{ci} load lines are shown as positive on demagnetisation curves but arithmetically are negative slopes. Figure 27.23 shows the case of the P_c and P_{ci} plotted for a NdFeB magnet when $P_c = 1$ and $P_{ci} = P_c + 1 = 2$.

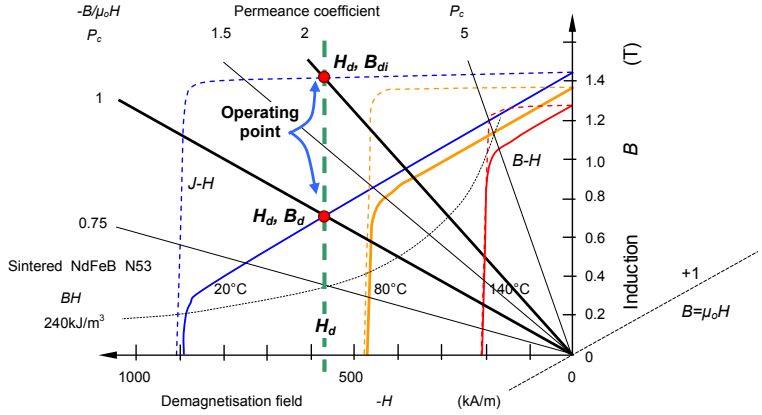


Figure 27.23. *Permanent magnet and gapped circuit load lines (constant energy contour shown).*

The P_{ci} and P_c straight load lines intersect the intrinsic and normal curves respectively with the same H_d value. This condition also occurs with an externally applied field (magnetising or demagnetizing influence) with or without air gap, which vertically shifts the common intersection origin of the P_{ci} and P_c lines by an amount related to the applied field (see section 27.6.3 and figure 27.28).

Example 27.1: ***Magnet load dependant operating point***

Using an equivalent circuit approach, figure 27.17c; determine the magnet operating point, assuming a linear demagnetization characteristic, when:

i. A coil of *N* turns carries a demagnetising current *i*, wound on a core of infinite permeance, which is also the permeance seen by the magnet, and
ii. with zero coil current, an air gap of length ℓ_g is introduced.

Solution
From equation (27.21) the equation describing the magnet linear demagnetising characteristic in terms of three flux components, shown in figure 27.24, is given by

$$\boldsymbol{\phi_m} = \phi_r + \frac{\mathfrak{I}_m}{\mathfrak{R}_m}$$

where $\boldsymbol{\Phi_m}$ is the intrinsic flux, $B_r A_m$. This three-term equation is common to any magnetic circuit loading the magnet. In order to obtain an operating point, the external magnetic loading circuit is expressed in terms of the magnet flux Φ_m and the magnet mmf $\mathfrak{I}_m$, that is, B_m and H_m respectively.

i. When the magnet is loaded by an external mmf *Ni*, which can tend to have a remagnetisation or demagnetisation effect, the load circuit in terms of magnetic variables related to H_m and B_m is

$$\mathfrak{I}_m = -Ni \qquad (= \boldsymbol{H_m} \ell_m)$$

Substituting $\mathfrak{I}_m$ into the magnet load line equation yields magnet flux Φ_m:

$$\boldsymbol{\phi_m} = \phi_r - \frac{Ni}{\mathfrak{R}_m} \qquad (= \boldsymbol{B_m} A_m)$$

These two equations in terms of H_m and B_m give the operating point H_d B_d, specifically

$$\boldsymbol{H_m} = -\frac{Ni}{\ell_m} \tag{27.28}$$

and

$$\boldsymbol{B_m} = B_r - \mu_m \frac{Ni}{\ell_m} \tag{27.29}$$

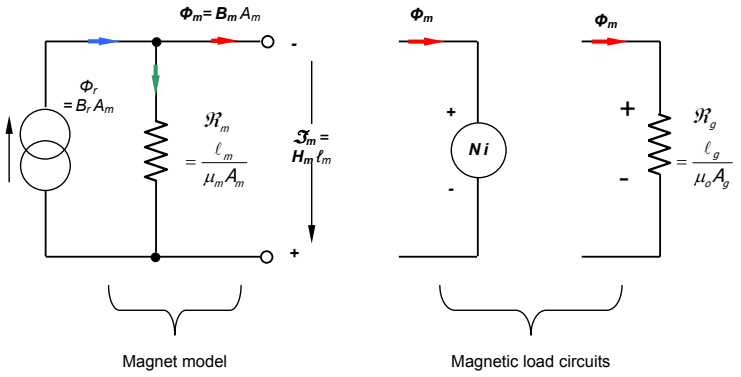


Figure 27.24. *Magnetic model for Example 29.1*

ii. When the magnet is loaded by the introduction of an air gap (progressively increased in length from initially zero to ℓ_g), the reluctance load circuit in terms of magnetic variables related to H_m and B_m is

$$\mathfrak{I}_m = -\boldsymbol{\phi_m} \mathfrak{R}_g$$

Solving the magnet load line equation and the load equation for magnet flux Φ_m and magnet mmf $\mathfrak{I}_m$ yields:

$$\phi_m = \frac{\phi_r}{1 + \frac{\mathfrak{R}_g}{\mathfrak{R}_m}} = \phi_r \frac{\mathfrak{R}_m}{\mathfrak{R}_g + \mathfrak{R}_m}$$

(This is the same concept as a current (flux) dividing between two parallel-connected resistors (reluctances)).
Next, the magnet mmf is given by

$$\mathfrak{I}_m = -\phi_r \frac{\mathfrak{R}_m \mathfrak{R}_g}{\mathfrak{R}_g + \mathfrak{R}_m} = -\phi_r \frac{1}{\frac{1}{\mathfrak{R}_g} + \frac{1}{\mathfrak{R}_m}}$$

These two operating point expressions in terms of the specific operating point H_d and B_d are

$$B_d = \frac{B_r}{1 + \frac{\mathfrak{R}_g}{\mathfrak{R}_m}}$$

and

$$H_d = \frac{-B_r A_m}{\ell_m} \frac{\mathfrak{R}_m \mathfrak{R}_g}{\mathfrak{R}_m + \mathfrak{R}_g} = -H_c \frac{\mathfrak{R}_g}{\mathfrak{R}_m + \mathfrak{R}_g} = -H_c \frac{1}{1 + \frac{\mathfrak{R}_m}{\mathfrak{R}_g}}$$

H_b and B_b can be rearranged, using $S = \mu_m \mathfrak{R}_m / \mathfrak{R}_g$, to yield equations (27.22) and (27.23).

For a given *B-H* demagnetisation characteristic, the operating point is determined solely by the ratio of the air gap reluctance and magnet reluctance.

The load line solutions are shown in figure 27.25. The operating point on the *B-H* characteristics will yield energy change per unit volume, *B*×*H*, while the operating point on the BA_m versus $H\ell_m$ axes yields energy, $B \times H \times A_m \ell_m = B \times H \times Volume_m$.

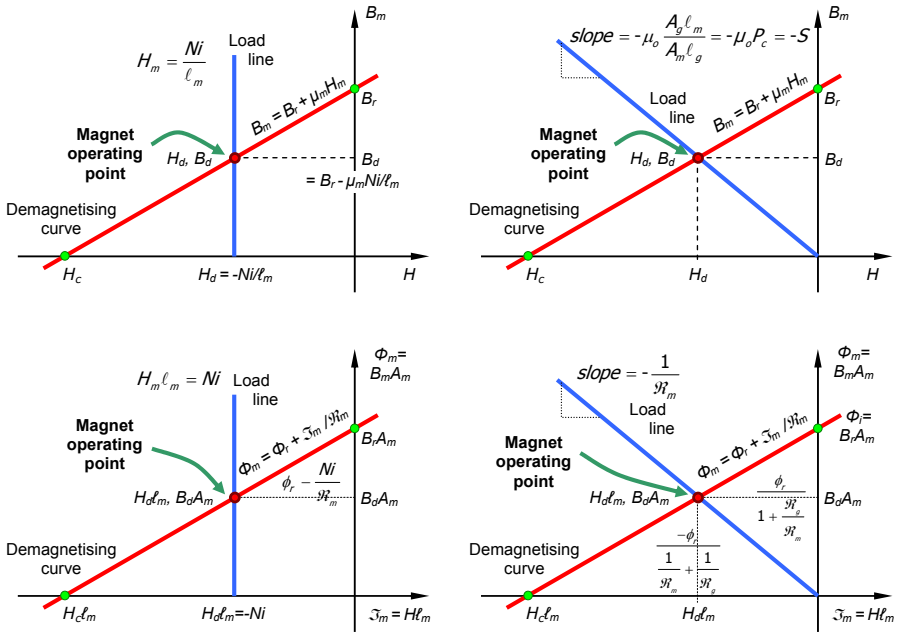


Figure 27.25. *Magnetic load line characteristics for problem 27.1.*

♣

27.6.3 Demagnetizing field

When a permanent magnet provides flux to a single air gap via two steel pole pieces, the specific dimensions of the magnet and air gap yields a unique load line, as shown in blue on figure 27.22. The permanent magnet material is characterized by its demagnetization curve (in red), the intercept between the two giving an operating point with a unique combination of B_d and H_d for the magnet in this condition. If the air gap length ℓ_g increases, the permeance, whence permeance coefficient, decrease, and the slope of the load line decreases and the magnet's operating point changes, moving down the curve.

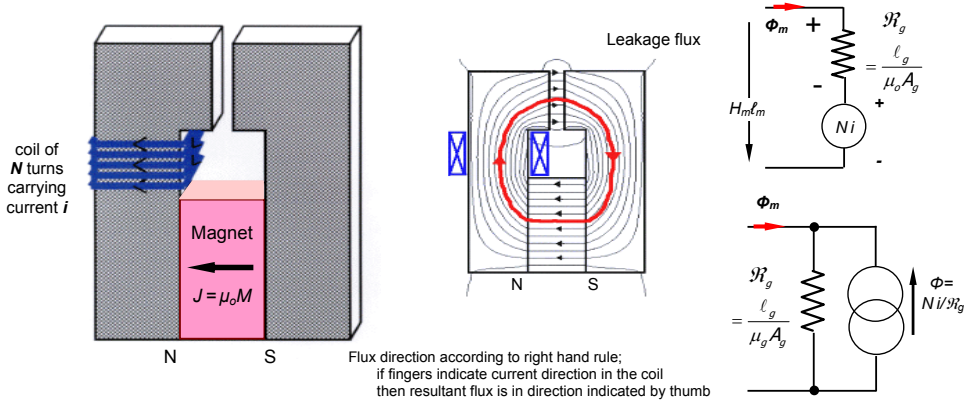


Figure 27.26. *Magnetic circuit with a permanent magnet, an air gap, and an excitation coil.*

Figure 27.26 shows the magnetic circuit in figure 27.21 with the addition of a coil of N turns carrying current i, with a direction that opposes (tending to demagnetise) the magnet (of area A_m, length ℓ_m) from delivering flux into the air gap (of area A_g, length ℓ_g). By integrating the magnetizing force H around the closed loop shown in red (Ampère's Law), where the loop is linked N times by the current i, the derived magnetic circuit equations are

$$\int H.d\ell = Ni$$

$$\boldsymbol{H_m}\ell_m + k_\ell \boldsymbol{H_g}\ell_g = \boldsymbol{Ni} \tag{27.30}$$

Since

$$\boldsymbol{B_m}A_m = k_\ell \boldsymbol{B_g}A_g = k_\ell \mu_o \boldsymbol{H_g}A_g$$

Then, assuming $k_\ell/k_m = 1$:

$$\boldsymbol{B_m} = -\mu_o \frac{A_g \ell_m}{A_m \ell_g}\left(\boldsymbol{H_m} - \frac{Ni}{\ell_m}\right)$$

$$\boldsymbol{B_m} = -S\left(\boldsymbol{H_m} - \frac{Ni}{\ell_m}\right) = -\mu_o P_c\left(\boldsymbol{H_m} - \frac{Ni}{\ell_m}\right) \tag{27.31}$$

When $i = 0$, this equation reduces to equation (27.20).
The load line expression now contains an additional term (the last term) representing the demagnetizing field produced by the coil current. The slope $B_m/\mu_o H_m$, namely P_c, of the load line is unchanged, but its parallel position is determined by the coil excitation term Ni/ℓ_m (H_m when $B_m = 0$ in equation (27.31)) as shown in figure 27.27.

The induction at the load line operating point is given

$$\boldsymbol{B_m} = B_r + \mu_m \boldsymbol{H_m} \tag{27.32}$$

Substituting the magnet demagnetising field, H_m from equation (27.31) for H_d gives

$$B_m = B_r + \mu_m\left(\frac{Ni}{\ell_m} - \frac{B_m}{S}\right) \tag{27.33}$$

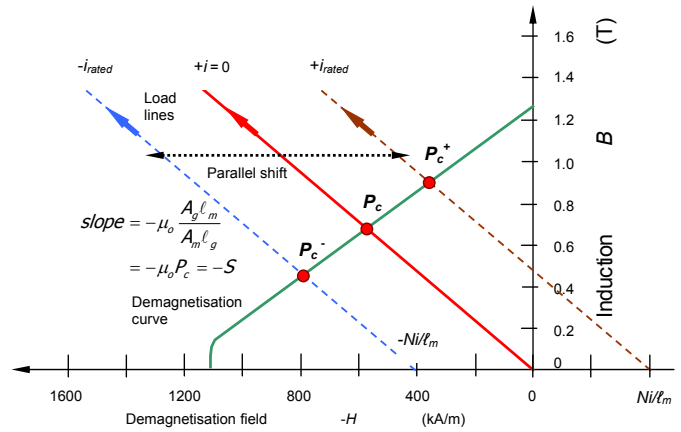


Figure 27.27. *Magnetic circuit operating point shift due to coil excitation.*

Isolating B_m and H_m gives the operating point H_d, B_d:

$$B_d = \frac{B_r + \mu_m \frac{Ni}{\ell_m}}{1 + \frac{\mu_m}{S}} = \frac{SB_r + \mu_m S\frac{Ni}{\ell_m}}{S + \mu_m} \tag{27.34}$$

$$H_d = -\frac{B_r + S\frac{Ni}{\ell_m}}{S + \mu_m} = -\frac{B_r + \mu_o P_c \frac{Ni}{\ell_m}}{\mu_o P_c + \mu_m} \tag{27.35}$$

Since linear models are used, by superposition, the operating point given by equations (27.34) and (27.35) reduce to:

- equations (27.29) and (27.28) when the air gap is reduced to zero such that $\ell_g = 0$ and $S \to \infty$.
- equations (27.23) and (27.22) when the current is reduced to zero such that $Ni = 0$.

Operating point based on the intrinsic demagnetization curve

Applying a magnetising field H (a demagnetizing or magnetising influence) creates a shift in the load line equal to the applied field, as shown in figure 27.27. A common unnecessary practice is to use of the intrinsic magnetisation curve J as the basis for operating point analysis. The intrinsic curve related load line has a slope $P_{ci} = P_c + 1$, but the H axis intercept is not coincident with normal load line, unless $i = 0$, as is highlighted in the characteristics in figure 27.28. But the B and J axis intercepts are coincident.

Applying an H field creates a horizontal shift of the normal load line equal to the applied field. The horizontal shift of the intrinsic load line is related to but not equal to the applied H field. The intrinsic operating point H_{mi}, J_m is defined in terms the normal curve operating point related equations, namely equations (27.32) and (27.31), except μ_m is replace by $\mu_m - \mu_o$ in the magnet equation and P_c is replace by $P_c + 1$ for the slope in the load equation (the load line equation constant term, that is, the B axis intercept, is unchanged). That is, from the magnet equation, equation (27.32)

$$\boldsymbol{J_m} = B_r + (\mu_m - \mu_o)\boldsymbol{H_{mi}} \tag{27.36}$$

and from equation (27.31) the load line is

$$\boldsymbol{B_m} = -\mu_o P_c\left(\boldsymbol{H_m} - \frac{Ni}{\ell_m}\right) = -\mu_o P_c \boldsymbol{H_m} + \mu_o P_c \frac{Ni}{\ell_m}$$

$$\boldsymbol{J_m} = -\mu_o (P_c + 1)\boldsymbol{H_{mi}} + \mu_o P_c \frac{Ni}{\ell_m} \tag{27.37}$$

In the load line equations (27.31) and (27.37), $B_m(H_m = 0) = J_m(H_{mi} = 0) = SNi/\ell_m$. That is, both load line equations have the same flux density axis intercept, as illustrated in figure 27.28. Also, from the load line equations, the magnetising field H axis intercepts are Ni/ℓ_m *and* $P_c Ni/\ell_m P_{c+1}$, respectively, again shown in figure 27.28.

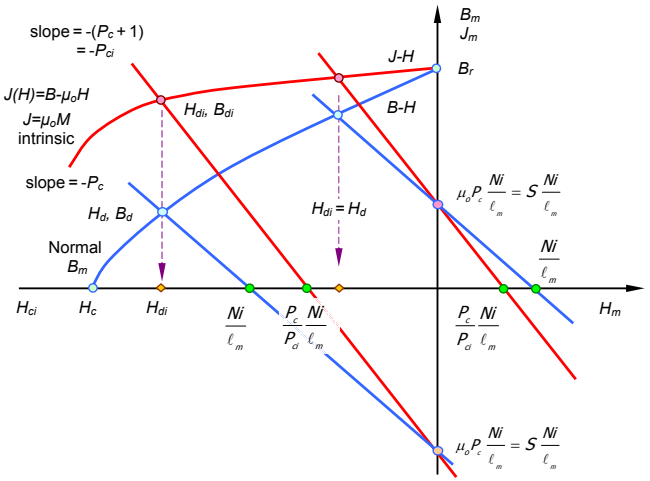


Figure 27.28. *Intrinsic demagnetisation characteristic and load lines for a magnet experiencing coil excitation Ampere turns bias.*

Solving for the operating point H_{mi}, J_m

$$J_m = B_r - \frac{\mu_m - \mu_o}{S + \mu_m}\left(S\frac{Ni}{\ell_m} + B_r\right) = \frac{(\mu_m + S)B_r - (\mu_m - \mu_o)S\frac{Ni}{\ell_m}}{S + \mu_m} \qquad (27.38)$$

$$H_{mi} = -\frac{B_r + S\frac{Ni}{\ell_m}}{S + \mu_m} = H_m \qquad (27.39)$$

where, comparing equations (27.35) and (27.39), as expected $H_m = H_{mi}$, since $\boldsymbol{B_m} = \boldsymbol{J_m} + \mu_o\boldsymbol{H_m}$. This equation can be confirmed by substitution of B_m from equation (27.34) and H_m from equation (27.35), with the resulting equation for $\boldsymbol{J_m}$ being confirmed by equation (27.38).

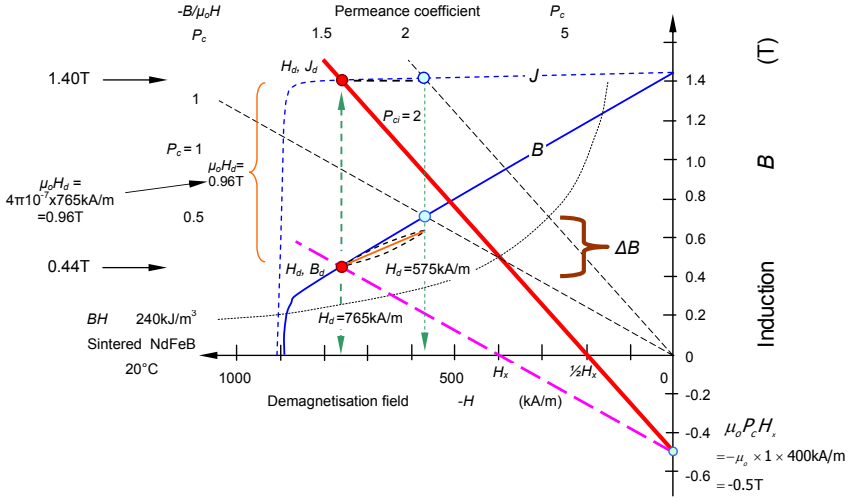


Figure 27.29. *Magnetic circuit operating point shift and flux density loss ΔB due to coil excitation (240kJ/m³ constant energy contour shown).*

Figure 27.28 shows load lines for positive and negative coil currents. The intrinsic (magnetization) curve operating point can be used to find the normal curve operating point, and vice versa, since $\boldsymbol{B_m} = \boldsymbol{J_m} + \mu_o\boldsymbol{H_m}$; where $\boldsymbol{H_m}$ is negative in the second quadrant). These load line and operating point characteristics are illustrated numerically in figure 27.29.

The commonality condition of the *J*/*B* axis intercept to both the normal and intrinsic related load lines is readily extracted (confirmed) using an equivalent circuit approach with $H_m\ell_m$=0, as shown in figure 27.29a. Analysis is based on the fact that since the mmf is zero on the Y axis, then the two model current sources must be of equal magnitudes, with one sourcing flux - the other sinking the same flux.

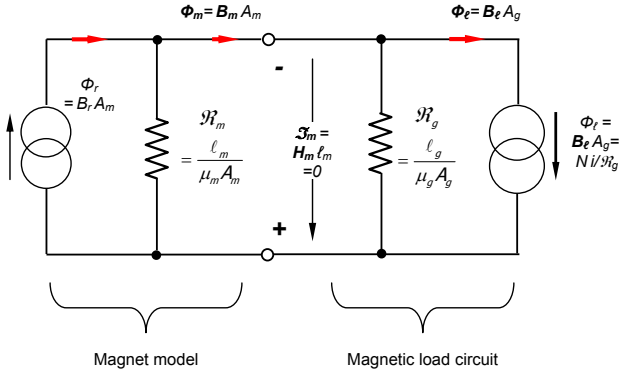


Figure 27.30. *Magnetic circuit for demagnetisation and load line characteristic of a magnet experiencing an air gap and coil excitation Ampere turns bias.*

That is, with zero net flux $\Phi_r = \Phi_m$, (the intrinsic and normal characteristics merge)

when $H = 0$ then $\phi_m = \phi_\ell$ thus $B_m A_m = B_\ell A_g$

$$B_m(H=0) = \frac{B_\ell A_g}{A_m} = \frac{Ni/\Re_g}{A_m} = S\frac{Ni}{\ell_g}$$

The operating point analysis approach adopted for externally applied fields *H*, highlights that use of the intrinsic demagnetising curve *J* is redundant. This is expected since *J* is artificially generated from the normal demagnetising curve *B* by adding $\mu_o H_m$. However, use of *J*-*H* does offer two features:

- All practical load line analysis can be performed in the second quadrant. Inspection of the *B*-*H* characteristics in figures 27.39 and 27.40 shows that the demagnetising curves project into the third quadrant, down to $B_m=\mu_o H_m$, and figure 27.38 shows operating point analysis in the third quadrant.
- The slope of the intrinsic characteristic gives a better visual indication of recoil and thermal properties – the closer to horizontal, the better the magnet. On the other hand, visualising how close the normal characteristic slope is to μ_o is more subjective.

27.7 Generalising equivalent magnetic circuits in terms of permeances

The magnetic circuits considered thus far are simple circuits comprising a permanent magnet, a single air gap, and a coil - real magnetic circuits are generally more complex. For example, the effect of leakage flux was accounted for by the correction factor k_ℓ, but this is an important effect, which normally requires calculation using estimates of leakage gap areas and lengths. Leakage flux follows paths in parallel with the main air gap in a magnetic circuit, which complicates the calculation of the load line slope thus A_g and ℓ_g, must represent the net load experience by the magnet. While it is convenient to find the operating point of a magnet by the intersection of its load line with its demagnetization characteristic, determination of the effective load line for a complex magnetic circuit may be convoluted. Rather than resorting to a finite element CAD approach, an approach is to defining a characteristic equivalent magnetic circuit involving at most two parameters for each circuit component - its permeance and possibly an mmf source, A flux source and its permeance can be converted into an equivalent mmf source and the same permeance, and vice versa.

A magnetic circuit model represents each component in terms of its dimensions and material properties. Unity k_r and k_f correction factors are assumed so the load line in equation (27.31) can be rearranged, in terms of flux $\boldsymbol{\Phi_m}$, the gap permeance $P_g = \mu_o A_g / \ell_g$, and the magnets mmf $\boldsymbol{\mathfrak{I}_m = H_m \ell_m}$, as

$$\boldsymbol{\phi_m} = \boldsymbol{B_m} A_m = -P_g \left(\boldsymbol{\mathfrak{I}_m} - \boldsymbol{Ni} \right) \tag{27.40}$$

This linear relationship is a *load line* with a slope based on the air gap load. By defining the mmf across the gap as $\boldsymbol{\mathfrak{I}_g = H_g \ell_g}$, using the magnetic equivalent to Ohms law, the gap mmf $\boldsymbol{\mathfrak{I}_g}$ is obtained from

$$\boldsymbol{\phi_m} = P_g \boldsymbol{\mathfrak{I}_g} \tag{27.41}$$

The gap permeance P_g is the slope of a flux versus mmf load line related either to the air gap or to the magnet. Hence in terms of the magnet's permeance:

$$\boldsymbol{\phi_m} = P_m (\boldsymbol{\mathfrak{I}_m} + B_r \ell_m / \mu_o) \tag{27.42}$$

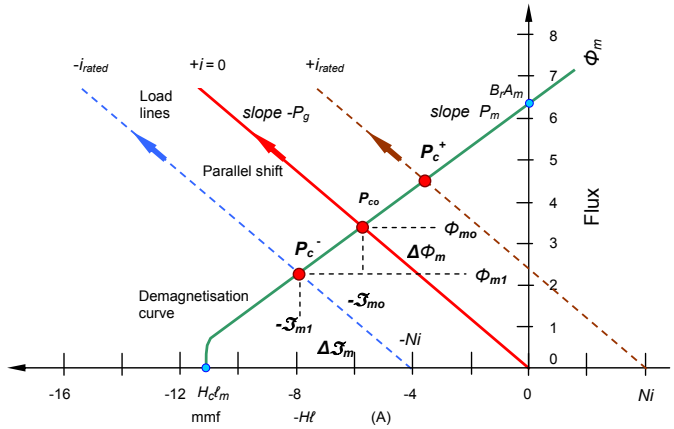


Figure 27.31. *Demagnetisation curve and load line for a magnet, expressed as permeances.*

Equations (27.40) and (27.42) are plotted in Figure 27.31, which is equivalent to Figure 27.27. The original demagnetization characteristic is scaled, the coercivity being multiplied by magnet length to give the equivalent mmf $H_c \ell_m$. The new load line is laterally displaced by the actual coil excitation Ni.
A coil current $-i$ moves the magnet's operating point from $-\mathfrak{I}_{mo}$, Φ_{mo} to $-\mathfrak{I}_{m1}$, Φ_{m1}, and the following relationships can be deduced from the geometry of Figure 27.31:

$$P_m = \frac{\phi_{m1} - \phi_{mo}}{\mathfrak{I}_{m1} - \mathfrak{I}_{mo}} = \frac{\Delta\phi_m}{\Delta\mathfrak{I}_m}$$

$$P_g = -\frac{\phi_{mo}}{\mathfrak{I}_{mo}} \tag{27.43}$$

$$Ni = \mathfrak{I}_{m1} + \phi_{m1} / P_g$$

The change in magnet flux due to coil excitation Ni can be found by rearranging the equalities in equation (27.43):

$$\Delta\phi_m = \phi_{m1} - \phi_{mo} = \frac{Ni}{P_m^{-1} + P_g^{-1}} \tag{27.44}$$

The magnetic circuit flux $\Delta\Phi_m$ can be decreased by decreasing either of the slopes P_m or P_g, which can be achieved by raising ℓ_m or ℓ_g respectively – with the advantage that additional magnet length also stabilizes temperature incurred permanent flux changes in the magnet.
In some applications, it is desired to maximise the flux change for a given excitation, which requires increasing P_m and/or P_g. This can be achieved not by reducing ℓ_m or ℓ_g, but by raising the permeability of the component materials since the magnet operates on a recoil line of slope $\mu_o\mu_{rc}$, the definition of P_m is changed to include μ_{rc}, thus equations (27.27) and (27.42) become

$$P_m = \frac{\mu_o \mu_{rc} A_m}{\ell_m} \tag{27.45}$$

$$\boldsymbol{\phi_m} = P_m \left(\boldsymbol{\mathfrak{I}_m} + \frac{B_r \ell_m}{\mu_o \mu_{rc}} \right) \tag{27.46}$$

A magnet with a *high* recoil permeability $\mu_o\mu_{rc}$, such as one of the alnico materials can be used to improve $\Delta\Phi_m$, equation (27.44). Similarly, a soft magnetic material of relative permeability μ can be considered for the *air* gap. The load line equation (27.40) is unchanged provided the gap component is defined by

$$\boldsymbol{B_g} = \mu_o \mu \boldsymbol{H_g} \tag{27.47}$$

$$P_g = \frac{\mu_o \mu A_g}{\ell_g} \tag{27.48}$$

These equations are more general versions of the equations for air, for which $\mu = 1$.

The alternative to plotting the two intersecting lines, is to solve equations (27.40) and (27.42) for Φ_m and $\mathfrak{I}_m$

$$\phi_m = \frac{Ni + B_r \ell_m / \mu_o}{\mathfrak{R}_m + \mathfrak{R}_g} \tag{27.49}$$

The load circuit may comprise a number of parallel-connected components, such as a leakage path around a main gap, and because these are in parallel with the same mmf, equation (27.41) shows that the permeance P_g used as the slope of the load line will be the sum of the parallel component permeances.

$$P_g = P_1 + P_2 + \ldots$$

That is, parallel connected reluctance components

$$\frac{1}{\mathfrak{R}_g} = \frac{1}{\mathfrak{R}_1} + \frac{1}{\mathfrak{R}_2} + \ldots$$

The magnetic load circuit may also comprise a number of series connected components, such as pole pieces and the main air gap, and because these experience the same flux, the net load line slope will be the sum of the series component reluctances

$$\mathfrak{R}_g = \mathfrak{R}_1 + \mathfrak{R}_2 + \ldots$$

or

$$\frac{1}{P_g} = \frac{1}{P_1} + \frac{1}{P_2} + \ldots \quad P_g^{-1} = P_1^{-1} + P_2^{-1} + ..$$

Note the analogy of magnetic flux, mmf, reluctance, and permeance to electrical current, emf, resistance, and conductance, respectively. Kirchhoff's voltage and current laws are analogous to mmfs around a closed loop and fluxes at a node, summing to zero.
The electrical equivalent approach involves reducing the magnetic load circuit to a Thevenin or Norton equivalent. With a graphical approach based on fluxes and mmfs, swept areas are energy, in Joules.
Also, magnetic flux density, B, magnetic induction, and magnetic field are generally interchangeable, as are magnetic field intensity, H, magnetic field strength, and magnetizing/demagnetising field.

27.8 Permanent magnet stability - Loss of magnetism

The ability of a permanent magnet to support an external magnetic field results from small magnetic domains *locked* in position by crystal anisotropy within the magnet material. Once established by initial magnetization, these positions are maintained until acted upon by forces exceeding those that retain the domains. The energy required to disturb the magnetic field produced by a magnet varies with material type. Permanent magnets can be produced with extremely high coercive forces H_c that maintains domain alignment in the presence of high external magnetic fields. Stability is described as the repeated magnetic performance of a material under specific conditions over the magnet's operational life.

Factors affecting magnet stability include time, temperature, reluctance changes, adverse fields, radiation, shock, stress, and vibration.

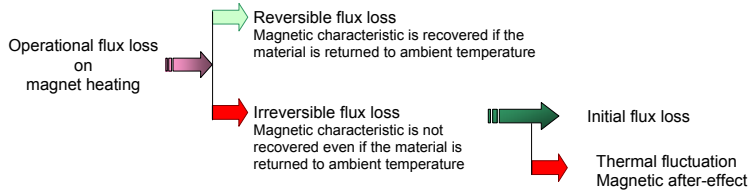


Figure 27.32. *Types of magnetic output loss.*

There are three types of loss in magnetic output due to temperature effects:

- **Reversible**: Flux output increases or decreases as temperature or mechanical stress changes. When the temperature returns to room temperature, the original flux output is observed. That is, there is no permanent change in the flux output of the magnet. If the magnetic properties vary with temperature without causing the magnet to operate at a *B-H* point below any *knee*, then the magnet only experiences a reversible loss, since the original operating condition is restored when the temperature returns to its normal level.
  It is because the temperature coefficients of B_r and H_{ci} (and H_c) are significantly different that the demagnetization curve develops a *knee* at elevated temperatures. Reversible losses cannot be eliminated by magnet stabilization and are described by the Reversible Temperature Coefficients α and β, shown in table 27.2 and expressed as %/°C. To accommodate exposure to temperatures at which mild de-magnetization occurs, magnets can be *pre-stabilized*, that is, partially de-magnetized.

- **Irreversible, Recoverable**: With material temperature change (high or low), a critical operating parameter is reached resulting in the magnet being partially demagnetized. These losses are only recoverable to the original flux level by re-magnetization, and are not recovered when the temperature returns to its original value. These losses occur when the operating point of the magnet falls below the *knee* on the demagnetization curve. An efficient permanent magnet design should have a magnetic circuit in which the magnet operates at a permeance coefficient above the *knee* of the demagnetization curve at expected elevated temperatures. This will prevent performance variations at elevated temperatures.

- **Irreversible, Unrecoverable**: The magnet is exposed to high temperature or corrosion conditions that result in a permanent metallurgical degrading microstructural change. Table 27.2 shows critical temperatures for the various materials, where:
  T_C is the Curie temperature at which the elementary magnetic moments are randomized and the material is demagnetized; and
  T_{op} is the maximum practical operating temperature in air.
  Different grades of each material exhibit values differing slightly from the values shown in Table 27.2. The maximum practical operating temperature is dependent on the operating point of the magnet in the circuit. The higher the operating point on the demagnetization curve, the higher the magnet allowable operating temperature.

Partially demagnetizing a loaded magnet by exposure to elevated temperatures in a controlled manner stabilizes the magnet with respect to temperature. The slight reduction in flux density improves a magnet's stability because domains with low commitment to orientation are the first to lose their orientation. A magnet thus *stabilized* will exhibit constant flux, with smaller variation between magnets, when exposed to equivalent or lower temperatures.

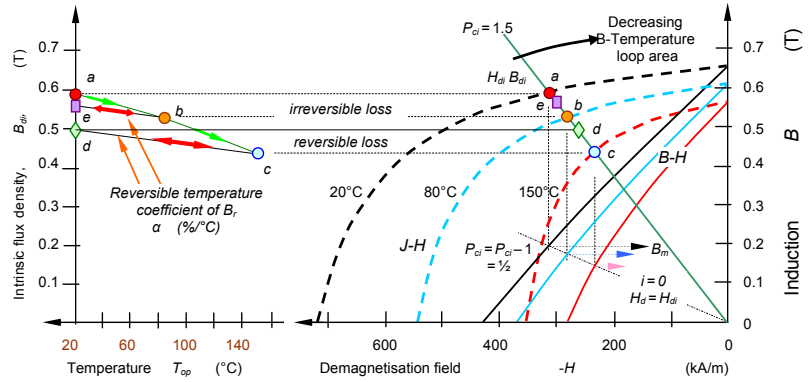


Figure 27.33. *Induction reversible and irreversible losses of a bonded magnet after a temperature increase.*

Flux losses in a magnet, due to increasing temperature, can be separated into irreversible and reversible loss components.
In figure 27.33, the reversible loss is the portion of loss at the elevated temperature which is recovered when the magnet returns to its initial temperature. The irreversible loss is the part of the flux loss at the

elevated temperature that is not recovered when the magnet returns to its initial temperature. To regain the *reversible* portion of the lost flux, the magnet will have to be remagnetized. The irreversible loss in an application can be minimized by cycling the magnet above the expected operating temperature or by magnetically stabilizing (by approximately 1.5 to 2 times the expected irreversible loss).
For a given load condition, P_{ci} =1.5 in figure 27.33, the flux change is partially reversible up to 80°C, although point b represents reduced magnet energy. On cooling to 20°C from 80°C, the operating point flux density is reduced to point e. If the temperature is then increased to 150°C, the magnet operating point shifts from point e through 'b' to point c, well below the characteristic knee, and the reversible change on subsequent cooling to 20°C is to operating point d, as determined by the reversible temperature coefficient, α, independent of the load line slope. The reversible flux density is the difference between the flux at the operating points c and d. Irreversible loss is the flux density at the operating point 'a' minus the flux density at point 'd'. Continued cycling between 20°C and 150°C will incur minimal additional irreversible (recoverable) loss. The flux-density temperature-dependant loop-area in figure 27.33 decreases as the load line permeance increases.
Characteristic parameters B_r, H_c, H_{ci}, and μ_{rc} (recoil relative permeability) specify the second quadrant of the hysteresis loop.

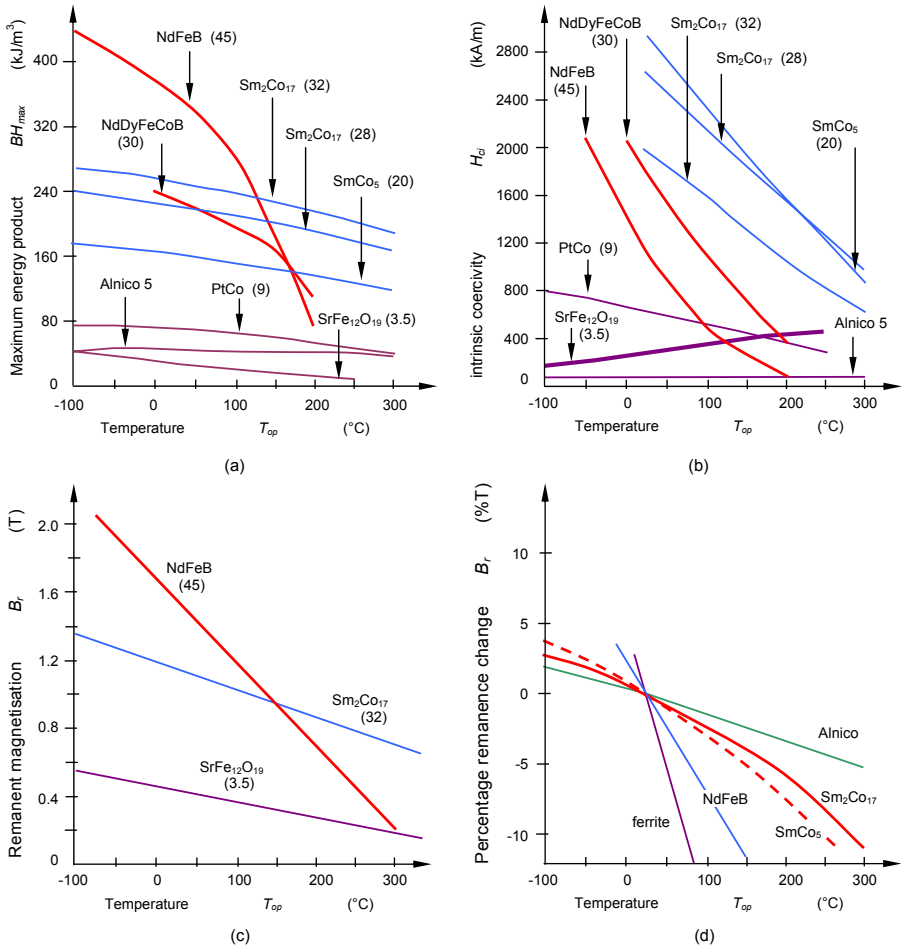


Figure 27.34. *Temperature effects on: (a) maximum energy BH_{max}, (b) intrinsic coercivity H_{ci}, (c) remanence B_r, where material room temperature maximum energy product is shown in brackets, and (d) percentage reversible changes in remanence.*

Changes in temperature affect both flux output, which is proportional to B_r, and resistance to demagnetization, which is proportional to H_{ci}. The amount to which these change are called the *reversible temperature coefficient of induction* B_r, namely α or the *reversible temperature coefficient of coercivity* H_{ci}, namely β. These are sometimes referred to as α (alpha) or β (beta) respectively and typical values for different hard magnetic material are shown in Table 27.2. Temperature variation of maximum energy product BH_{max} and intrinsic coercivity H_{ci} are shown on figure 27.34. The changes in B_r and H_{ci} as a function of temperature are not linear. A negative value indicates a loss of B_r or H_{ci} with an increase in temperature. Note that intrinsic coercivity for ferrites increases with temperature.

Irreversible-Recoverable loss occurs when the combination of: temperature extreme, H_{ci}, reversible temperature coefficient, and applied demagnetizing field exceeds the magnet's ability to remain fully magnetized. This occurs when the magnet is at an operating point below the *knee* of the curve.

In addition to the change in B_r and H_{ci} values, the demagnetization curve undergoes a subtle temperature induced change. The recoil slope of all magnetic materials increases, to varying extents, with increased temperature. For example, the recoil slope for SmCo increases from 1.082 at room temperature by 4%, up to 1.210 at 200°C.

Time
The effect of time on permanent magnets is minimal. These changes, known as *magnetic creep*, occur as less stable domains are affected by fluctuations in thermal or magnetic energy, even in a thermally stable environment. This variation is reduced as the number of unstable domains decreases. Rare Earth magnets are not as likely to experience this effect because of their extremely high coercivities. Irreversible aging losses increase with increasing temperature and decreasing permeance, with the vast majority of the decrease occurring within the first year after magnetisation. Below 50°C, irreversible loss for magnets over a ten-year period is less than ½% for low permeance coefficients, ranging from virtually zero for Samarium Cobalt materials to less than 3% for Alnico 5 materials.

Radiation
SmCo materials, and especially Sm_2Co_{17}, withstand radiation up to 40 times better than NdFeB materials. SmCo exhibits significant demagnetization when irradiated with a proton beam of 10^9 to 10^{10} rads. NdFeB will lose all of it magnetization at a dose of 7×10^7 rads, and 50% at a dose of 4×10^6 rads. Magnet materials with high H_{ci} values, operated at high permeance coefficients, P_c, should be used in radiation environments, and then shielded from direct heavy particle irradiation. Stabilization can be achieved by pre-exposure to expected radiation levels.

Shock, stress, and vibration
Rigid magnet materials are brittle in nature, and can easily be damaged or chipped by improper handling. Samarium Cobalt in particular is a fragile material. Thermal shock when ferrites and Samarium Cobalt magnets exposed to high temperature gradients can fracture within the material.

27.9 Recoil operation and associated magnetisation losses

27.9.1 Losses due to reverse magnetic fields

i. Increased reverse field – increased air gap
For dynamic systems with changing operating straight lines (for example, motors) shearing should be selected so that the permanent magnet's working point remains within the straight-line range of the demagnetization curve. The reason is to ensure stability from outside field and temperature influences, as shown in figure 27.35. The working point shifts to a larger opposing field strength, for example, from P_1 to P_2, if the air gap in a magnet system is increased. If the gap change is reversed, the original operating point P_1 can only be recovered if P_2 is within the linear section of the demagnetization curve, as in figure 27.35a. At room temperature, +20°C, the magnetic material suffers no apparent irreversible loss of its magnetization, even if continuously cycled between P_1 and P_2.
Provided the demagnetization curve is linear and M is constant, there are no irreversible losses and the operating point traverses the same characteristic *reversibly* as implied by:

$$\text{from } B_m = \mu_o\left(H_m + M\right)$$

$$\frac{dB_m}{dH_m} = \mu_o\left(1 + \frac{dM}{dH_m}\right)$$

$$M = \text{constant} \rightarrow \frac{dB_m}{dH_m} = \mu_o$$

However, as the air gap is further increased past P_2, to operating point P_3, as shown in figure 27.35b, to below the knee of the demagnetization curve, irreversible losses arise. On reducing the air gap, the

working point shifts to P_4 on an inner recoil return path with a correspondingly lower flux density. The slope of this return path is referred to as permanent permeability. If the demagnetization exceeds the intrinsic coercivity $-H_{ci}$, then the reversal of M is complete. An irreversible loss has occurred, and the magnet can only be returned to its original condition if it is fully remagnetised.

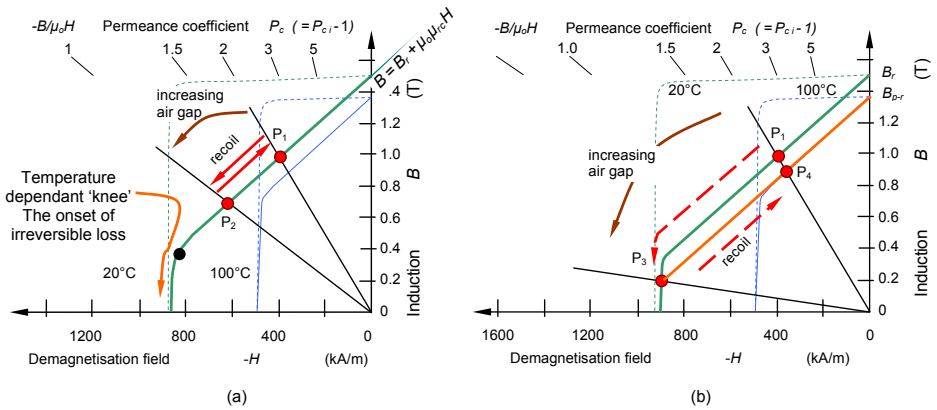


Figure 27.35. *Demagnetisation operation for varying air gap.*

Remagnetisation is not practical for a magnet installed in a magnetic circuit (a motor), but if the loss of magnetization is minimal, the device can be designed to operate within its *major* demagnetization curve at a lower flux level. In figure 27.35b, the orange line from P_3 that the operating point follows to P_4, on the original load line, is called a *recoil line*, which is actually part of a *minor* magnetization curve. The excursion of the operating point in recoil does not traverse a line, but a narrow hysteresis loop, as shown in figure 27.16, although a line of slope $\mu_o\mu_{rc}$ is assumed for calculation purposes. All recoil lines have approximately the same slope, termed *recoil relative permeability*, μ_{rc}. In many magnetic materials the slope is near constant whatever the point of origin and is approximated by the slope of the major hysteresis loop at point $H = 0$ where $B = B_r$. If a recoil line is projected to intercept the B_m axis at a reduced 'remanence' of $B_{p\text{-}r} = \mu_o M_r$, then the equation (similar to that for the demagnetization curve) describing it is:

$$\boldsymbol{B_m} = \mu_o\left(\mu_{rc}\boldsymbol{H_m} + M_r\right)$$

A permanent magnet's uniform magnetization M produces a magnetizing force H which establishes a flux density B. For operation in recoil, it is the recoil line (rather than the demagnetization curve) and the load line which are the two characteristic equations describing the magnet in a particular magnetic circuit, their intercept giving a unique magnet operating point with a specific B_m and H_m for the magnet. In terms of the B_m versus H_m diagram in figure 27.35b, the recoil line substitutes the demagnetization curve.

ii. Reverse applied field - Ampere turns bias
Irreversible loss caused by the same effect as an excessive air gap, can result if an excessive demagnetizing field is applied to a magnet via current in a coil wound as shown in figure 27.26. When a coil biasing current flows, creating -400kA/m, the operating point moves from P_1 under zero current to P_2, as shown in figure 27.36a. When the current ceases the operating point recoils, without magnetisation losses to the original operating point P_1. As shown, a higher current, equivalent to -800kA/m, causes the operating point to move past P_2, around the *knee*, down to the working point P_3. On winding current cessation, recoil to point P_4 results, such that the magnet has suffered irreversible loss of its magnetisation, with the remanence flux density decreasing from B_r to $B_{r\text{-}p}$ on figure 27.36a.

Since $H=Ni/\ell_m$, the H axis excursion of the load line is proportional to the coil's excitation i and is also inversely proportional to the magnet's length ℓ_m. Additional magnet length ℓ_m, hence additional magnet material volume, can be used to stabilize a magnet against irreversible loss in two ways:

- it raises the slope of the load line, so the magnet's operating point at any temperature or ampere turns bias, is further away from the *knee*; and
- for a given coil excitation, it reduces the rate at which the load line approaches the *knee* (with its associated loss of magnetization).

The penalty of such temperature stability mechanisms is the use of additional magnet material.
Since an irreversible loss has occurred in operating at point P_3, the magnet can only be returned to its original condition if it is fully re-magnetized.

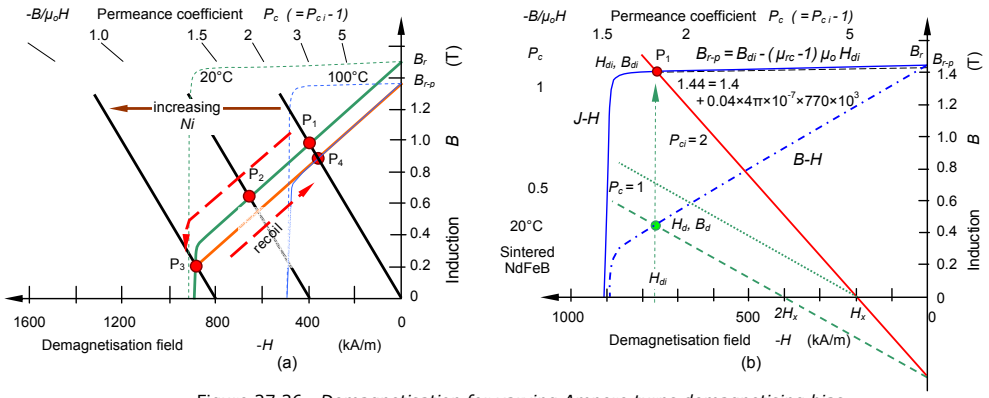


Figure 27.36. *Demagnetisation for varying Ampere turns demagnetising bias.*

27.9.2 Demagnetisation due to temperature increase

i. Increasing air gap – no external reverse field

Both induction B (and specifically remanent flux density B_r) and intrinsic coercivity H_{ci}, whence maximum energy product, are temperature (non-linear) dependant. Temperature variation of maximum energy product, remanence B_r, and intrinsic coercivity H_{ci} are shown in figure 27.34 for various hard magnetic materials. The relationships are referred to as the Reversible Temperature Coefficients, alpha α for B_r and beta, β for H_{ci}, in % change per °C, over a specified temperature range, as indicated by equations (27.3) and (27.4). Both coefficients are negative, except in the case of ferrite ($SrFe_{12}O_{19}$), where β is positive, as shown in figure 27.34b. Consequently, the demagnetization curves of permanent magnets are temperature dependent, as shown in figure 27.37. Notice that each diagonal magnet line has a bend in it, referred to as the *knee* in the curve, and the knee gets higher with increased temperature.

An increase in temperature causes the working point to shift along the load line radially towards the origin. In figure 27.37a, for a high permeance, P_{c1}, as the temperature increases the operating point progressively moves from point P_w, to P_x then to P_y and finally to P_z. Since the operating points stay within the linear region (above the *knee*) of each demagnetization curve, each reduction in flux density is reversible, that is, after cooling the flux density returns to its original value associated with operating point P_w. The magnet is operating in its safe linear region and performs as expected.

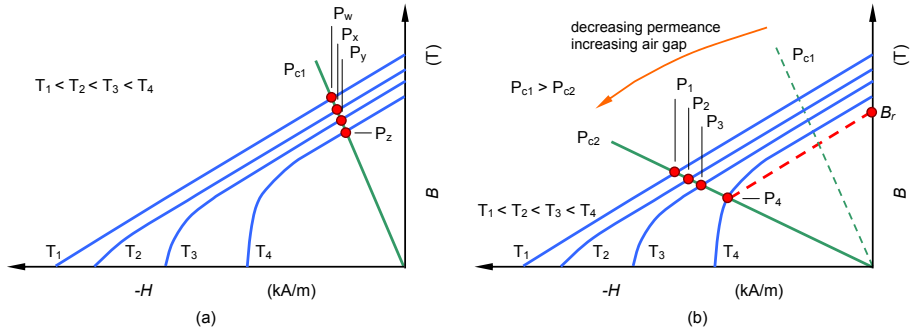


Figure 27.37. *Load line and operating point temperature dependence.*

In figure 27.37b, for the load line with decreased permeance P_{c2}, as the temperature increases the operating point moves towards the origin, from point P_1, to P_2 then to P_3 and finally to P_4. With an increased magnet temperature T_4, the operating point P_4 is below the *knee* of the demagnetization curve and the change in flux density is irreversible, that is, after cooling the flux density does not return to its original value associated with P_1. Only a portion of the magnetisation is reversible, as shown in figure 27.33. If the air gap is closed, the operating point recoils to a lower remanence flux density, B_r at T_4. As in figure 27.33, equation (27.3) can be used to estimate the remanence flux density for temperatures T_1, T_2, and T_3, based on the projected B_r for temperature T_4. With any operating point below the *knee* on the demagnetising curve, P_4, the irreversible component (demagnetisation - irreversible magnetic losses) and can only be recovered by remagnetisation. The knee flux density level increases with increasing temperature, reflecting the material's increasing vulnerability to demagnetization at higher temperatures.

To avoid irreversible changes in the flux density through temperature fluctuations, the operating point must remain within the upper linear section of the demagnetization curve over the entire temperature range in which the magnet is to operate. A permanent magnet can be completely demagnetized by heating to a temperature above the Curie temperature T_c. After cooling to the initial temperature, the initial state of magnetization can be reproduced by remagnetising, provided heating has not caused changes in the material microstructure.

ii. Reverse field – constant air gap

Consider a coil excitation that cycles the load line between zero and H_x = -320kAt/m, but with the magnet now operating at +150°C rather than +20°C. The diagram in figure 27.38b shows that, at higher demagnetizing fields, the operating point passes the *knee* and the magnet suffer an irreversible loss of magnetization. The operating point does return up along the major demagnetization curve, but follows a recoil path within (below) this characteristic. Since an irreversible loss has occurred, the magnet only returns to its original condition if it is fully remagnetized.
In figure 27.38a, N_o is the operating point on the *B-H* (normal) 20°C curve for the operating line $P_c = 1$ derived from the magnetic circuit and the corresponding point on the *J-H* (intrinsic) 20°C curve, with $P_{ci} = P_c + 1 = 2$, is point I_o. Both operating points have the same magnet field co-ordinate, $H_{io} = H_o$, when no external bias field is applied, that is, $i = 0$ in any coil.
When a demagnetizing field H_x = -320kAt/m is applied to the magnet (via current in a coil), at 20°C, the operating point moves from I_o to the operating line $P_{ci} = 2$ at point I_1 on the *J-H* curve, equivalent to point N_1 on the *B-H* curve. The field intensities for the normal and intrinsic operating points remain the same even when $i \neq 0$. The operating points shift (virtually) back to I_o and N_o as the demagnetizing (current) field is reduced to zero, that is, $i = 0$. The projected remanence flux density decreases slightly during what is effectively a stabilisation process.

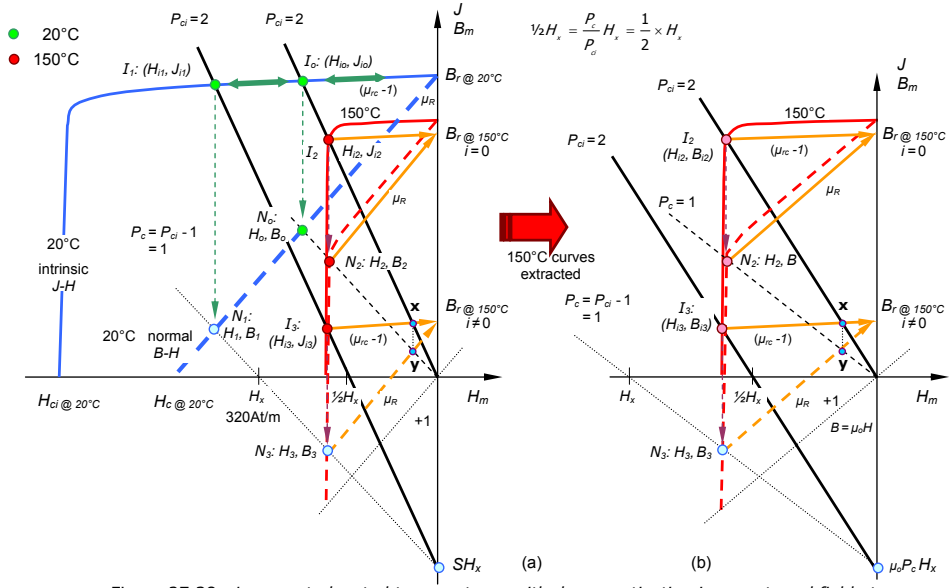


Figure 27.38. *Losses at elevated temperatures with demagnetisation in an external field at: (a) 20°C and 150°C with (b) 150°C operation extracted and expanded.*

Figure 27.38b shows the operating point trajectories at a higher temperature, 150°C, where coil excitation causes the operating points to locate well below the knee, in fact in the third quadrant. With no field bias, $i = 0$, operation at points $I_2 : N_2$ results in a small reduction in B_r. Having applied a coil current corresponding to H_x = -320kAt/m, removal of the demagnetising bias field leaves the magnet with severe irreversible loss, with the operating points having reduced from $I_3 : N_3$ to x : y, respectively. When a bias field is applied, either the intrinsic or normal demagnetising curves readily illustrate the loss.

Magnet Stability: irreversible thermal losses

The variation in a magnet's remanence B_r, in figure 27.34c is approximately linear with increasing temperature and reversible up to its specific transition temperature. Once the change becomes non-linear, an irreversible loss of B_r occurs. This transition is associated with the onset of a reversal of the material's magnetization, and is represented by a *knee* in the demagnetization curve. The load line intersection with the demagnetization curve identifies the operating point (B_d, H_d) for the magnet supplying flux to a given magnetic circuit. As with the remanence, if the operating point is above the demagnetization curve *knee*, changes in the magnet's condition are reversible, but if the operating point falls below the *knee*, irreversible loss of part of B_r occurs.

The position of the *knee*, that is, the threshold for s irreversible loss component arising, depends on temperature, and as the temperature increases, the operating point may readily fall below the *knee* of the applicable demagnetization curve.

Example 27.2: ***Magnet load and temperature dependant operating point***

Consider the following two magnets to be operating in a magnetic circuit with a load line as specified:

iii. A fully dense anisotropic neodymium-iron-boron magnet with demagnetization curves as shown in figure 27.39, is temperature cycled between 20°C and 120°C. Determine the subsequent remanence flux density at 20°C and 120°C, if the load line slope is -1.5; and

iv. A ceramic ferrite magnet with demagnetization curves as shown in figure 27.40, is temperature cycled between -60°C and 60°C. Determine the subsequent remanence flux density at -60°C and 60°C, if the load line slope is -1.

Solution

i. Sintered NdFeB

a. Reversible loss:- magnet operating temperature is cycled between +20°C and +80°C.

The second quadrant demagnetization *B-H* curves in figure 27.39 develop a *knee* in the second quadrant above about 0°C, with a knee apparent for a temperature of +20°C. The magnet's operating point at the intersection of the load line, P_{ci} = -1.5, and the +80°C curve is above the *knee*, so no irreversible loss occurs for temperatures up to about 85°C. This reversible change is illustrated on the right of the *B*-axis, with magnet flux cycling between point 'a' at 20°C, to point 'b' at +60°C and point 'c' at +80°C.

b. Irreversible loss:- magnet operating temperature is cycled between +20°C and +120°C.

The magnet begins again at 20°C, point 'a', is heated first to +60°C, point 'b', then to +80°C, point 'c', and then up to +100°C - point 'd'. However, the intersection of the load line and the +100°C demagnetisation curve is now below the *knee*, and an irreversible loss of magnet flux occurs (starting just above +80°C). When the magnet is cooled to 20°C, the magnet's operating point on the load line, point 'd*', is no longer on the 20°C demagnetization curve, but at some point within the major *B-H* curve in the second quadrant. If rather than cooling from 100°C, the magnet is heated to 120°C, the operating point moves to point 'e'. Now when the magnet is cooled to 20°C, operating point 'e*' results. The magnet is no longer fully magnetised, having suffered irreversible loss. Operation up to 120C is stabilised but it must be re-magnetized to saturate the material's magnetization once again, and to regain operation on the major *B-H* curve.

In summary:

'a' to 'b' to 'c' is a linear change and is reversible, as described in part a.
'c' to 'd' then to 'e' are non-linear changes and represent partial irreversible loss.
'e' to 'e*' is a linear change and is reversible, but with reduced magnetic properties.
'e*' to 'a' illustrates restoration of full magnetic properties after re-magnetizing the magnet.

The recoil lines in figure 27.39 from 'e' and 'e*' predict remanence fluxes of 0.87T and 1.06T at +120°C and +20°C, respectively. The magnet is now temperature stabilised for operation up to 120°C.

As the load line slope is increased, given the shape of these demagnetization curves, the transition temperature to an irreversible loss also progressively increases. A greater load line slope raises the flux produced by the magnet to the magnetic circuit and helps to stabilize the magnet against irreversible loss thermal effects.

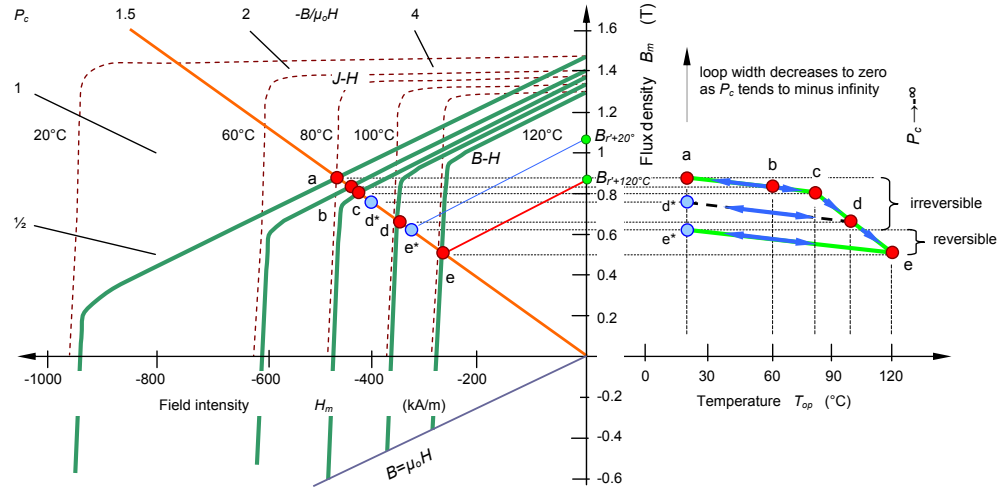


Figure 27.39. *Interpretation of irreversible and reversible temperature losses for a NdFeB magnet.*

ii. Ferrite

When considering the change in coercivity, H_{ci} decreases with temperature both for samarium-cobalt and for neodymium-iron-boron magnets; H_{ci} increases with temperature in the case of ceramic ferrites (being based only on magneto-crystalline anisotropy). This means that the *knee* of the demagnetization curve can arise as the temperature *falls*. The magnet is cycled from +20°C to -60°C and back again. Figure 27.40 illustrates that, with a load line slope of -1.5, the transition to an irreversible loss occurs when the temperature falls below about -20°C.

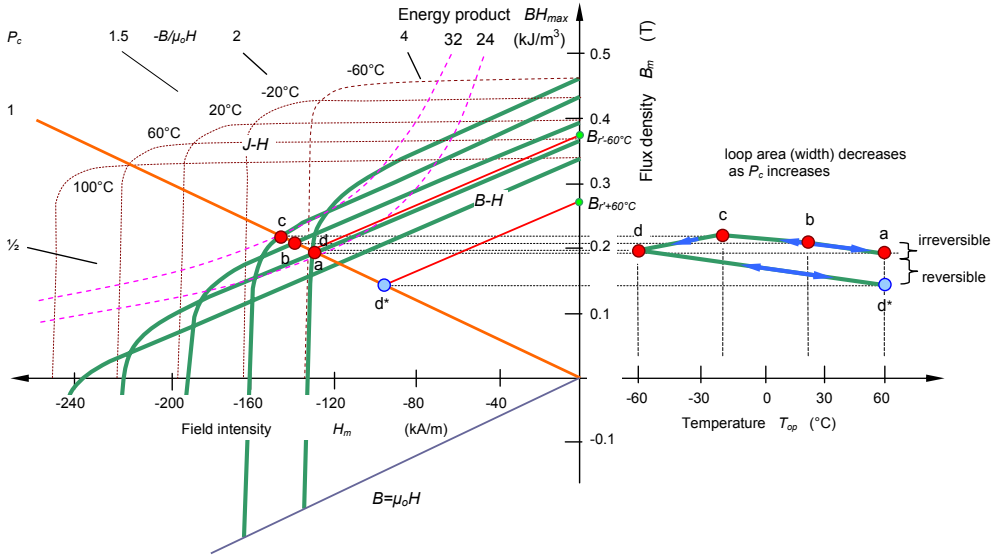


Figure 27.40. *Interpretation of irreversible and reversible temperature losses for a ferrite magnet.*

The sequence of thermal events is:

a to b to c is a *linear* reversible change between +60°C and -20°C.
c to d is a *non-linear* irreversible change from -20°C down to -60°C.
c to d to d* is a *linear* reversible change from -60°C back to +60°C, but with reduced magnetism.

The recoil lines in figure 27.40 from 'd' and 'd*' predict remanence fluxes of 0.38T (reduced from 0.46T) and 0.275T (reduced from 0.37T) at -60°C and +60°C, respectively.

♣

If operation of a magnet over its working temperature range is predicted to introduce unacceptable irreversible loss, then the magnet application should be reassessed to increase the load line slope and stabilize its operation without degrading the properties.

27.10 Energy transfer

One or more air gaps introduced into a magnetic circuit enable useful work to be performed. The mechanical energy used to separate a magnet from soft iron, there in creating an air gap, is stored as potential energy within that air gap and the magnet. This moves the point of operation on the intrinsic curve in the second quadrant. The normal curve in the second quadrant represents the energy output of the magnet and is used during magnet design. If the iron in the circuit is completely removed, the air gap becomes large and the operating point of the curve approaches H_c (the normal coercivity) in the second quadrant and the induction B approaches zero.

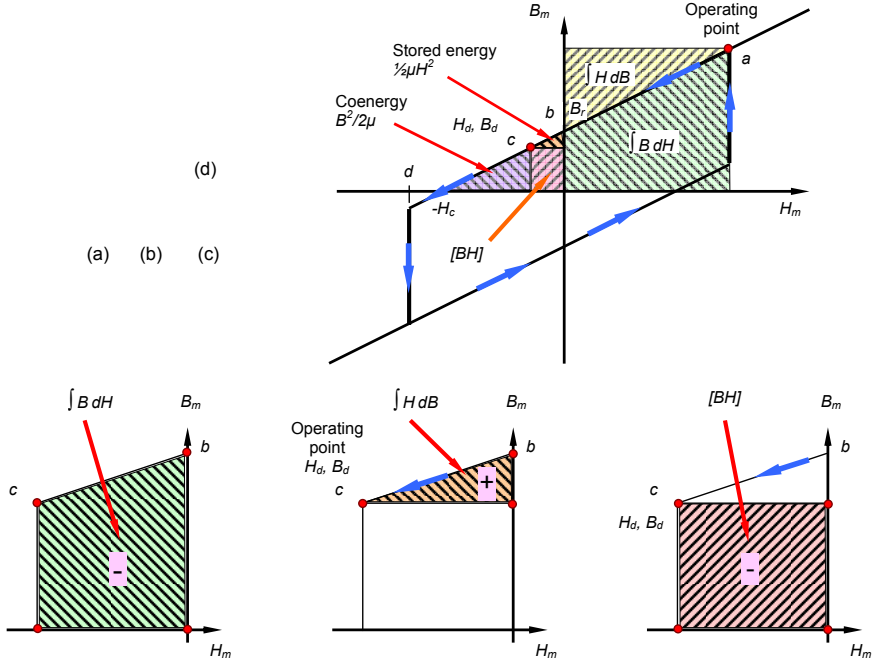


Figure 27.41. *Graphical demagnetisation change per unit volume: (a) applied field energy, (b) magnet kinetic energy, (c) total potential energy, and energy changes around the B-H loop.*

If the air gap is closed again, the stored potential energy is used to perform the work of bringing the magnet and the iron together. However, the operating point does not return to B_r. The magnet recoils along a so-called minor hysteresis loop to a point below B_r, figure 27.16c.
Repeated opening and closing of the air gap will result in the magnet cycling along this minor hysteresis loop. The average slope of the minor loop is the recoil permeability, $\mu_o\mu_{rc}$.

Section 27.6 presented the method for determining a magnet's operating point (B_m, H_m), associated with which is stored energy, which may be instrumental in the conversion of electrical and mechanical work. In figure 27.41, the energy stored in volume V_m of a magnet, leading to the change in energy is:

$$-\int_b^c BdH - \int_b^c HdB = -[BH]_b^c \qquad (27.50)$$

The first term is the work done (energy) by the applied field, the second is the internal kinetic energy stored, and the sum equals the total potential energy per unit magnet material volume.

After being magnetized to saturation, point a, the potential energy is reduced to zero at remanence (B_m = 0 at point b); the magnet conditions must move into the second quadrant to deliver its stored energy. Figure 27.41 shows how the three energy density components of equation (27.50) develop as the magnet moves from point *b* to point c. Considering the direction of integration, the three areas sum according to equation (27.50), namely, the work done as a result of the applied field $\int BdH$ equalling the change in potential energy $[BH]$ plus the release of kinetic energy from the magnet $\int HdB$. The higher the potential energy of a permanent magnet, the greater the release of its kinetic energy in establishing an external field. While magnet developments have focussed on improving the available (potential) energy density, it also requires a corresponding increase in the work to change the operating condition of the material. Consequently, the basis of most devices using high-energy magnets is a design that keeps the magnet close to a unique operating point with a minimum of dynamic operation.

The kinetic energy released by a magnet of volume V_m, which is operating at a typical point c, (H_d, B_d), in figure 27.41, is

$$W = V_m \int HdB \qquad (27.51)$$

This triangular area in figure 27.41b, the kinetic energy being released by the magnet, is the magnetic energy stored in the magnet reluctance, equivalent to $w = \frac{1}{2}LI^2$. For a given operating point, (H_d, B_d), this energy can be derived from the Thevenin or Norton equivalent circuits, specifically the magnetic energy in the magnet reluctance:

$$w = \tfrac{1}{2}\phi_m^2\Re_m = \tfrac{1}{2}\Im_r^2/\Re_m$$

Energy being delivered, is readily derived from the Norton equivalent circuit in figure 27.17c, where

$$w = \tfrac{1}{2}\phi_m^2\Re_m = \tfrac{1}{2}(B_mA_m)^2\frac{\ell_m}{\mu_mA_m} = \tfrac{1}{2}\frac{B_m^2}{\mu_m}A_m\ell_m = \tfrac{1}{2}B_mH_m \times volume$$

This energy is only released if work is done, that is, power is released or absorbed. Mechanically this could be an air gap change or electrically a voltage developed across a current carrying coil. Note coil current alone is insufficient, the voltage component can only be produced by a changing flux, from $v = N\times d\Phi/dt$, which implies a magnet operating point flux change, $d\Phi$.

Energy delivered to an air gap

The following equation may be applied to the field in any volume, even an air gap V_g, where $B_g = \mu_oH_g$, that is:

$$\begin{aligned} W_g &= V_g\int\frac{B_g}{\mu_o}dB \\ &= \tfrac{1}{2}\frac{B_g^2}{\mu_o}V_g = \tfrac{1}{2}B_gH_g\,V_g = \tfrac{1}{2}\mu_oH_g^2\,V_g \end{aligned} \qquad (27.52)$$

If this is the air gap in the magnetic circuit of figure 27.21, then equations (27.18) and (27.19) may be used in equation (27.52) to show that all the energy released from the magnet is delivered into the gap:

$$W_g = \tfrac{1}{2}B_gH_g\,V_g = -\tfrac{1}{2}B_mH_m\,V_m \qquad (27.53)$$

Note the energy density delivered to the gap is half the magnet's energy product. The work done by the externally applied field in establishing the air gap energy has two components as shown in figure 27.42a. The first component is the pu volume energy released into the air gap, equation (27.53), the second is the pu volume energy associated with establishing the operating point, namely the energy associated with increasing the air gap from zero at point b to length ℓ_g at point c.

Once the magnet operates on a recoil line of relative permeability u_R and recoil remanence B_r, this equation is modified by $B_m = \mu_o\mu_{rc}H_m + B_r$ to

$$W_g = \tfrac{1}{2}B_mH_m\,V_g = \tfrac{1}{2}(-\mu_o\mu_{rc}H_m - B_r)H_m\,V_g = \tfrac{1}{2}(-\mu_o\mu_{rc}H_m^2 - B_rH_m)V_g \qquad (27.54)$$

The further down its demagnetization curve a magnet is driven, the greater the area swept by the load line, by which the applied field within the magnet does more work. This energy is recaptured by the magnet if the gap is re-closed and the load line returns to its original position. However, if this cycle includes driving the magnet beyond a *knee* in its characteristic, a return to the original load line along a recoil line involves the irreversible loss of magnet field energy, as shown in figure 27.42. Large dynamic excursions of a high-energy magnet can be restricted by increasing ℓ_g during installation of the magnet. Subsequent cycling along the recoil line incurs no further irreversible loss, only reversible changes in energy.

While BH_{max} is the most common figure of merit for a permanent magnet, operation at this point maximizes the release of magnet energy into the gap, namely $-½V_m BH_{max}$. If the magnetization M is constant and the linear demagnetization characteristic is represented by $B_m = \mu_o(H_m+M)$, then the air gap energy, equation (27.53), may be written as

$$W_g = -½\mu_o V_m \left(H_m^2 + MH_m\right) \qquad (27.55)$$

Differentiating, BH_{max} occurs at $B_m = ½\mu_o M$, $-H_m = ½M$, for which

$$W_g^{max} = ½\mu_o V_m \left(½M\right)^2 \approx ½\mu_o V_m \left(½B_r\right)^2$$

The magnet energy released is lower either side of the BH_{max} point on the major demagnetization characteristic, and by virtue of a reduced alignment of the magnetization M, W_g is also smaller for operation on recoil lines within the major B versus H curve. The superimposed constant energy contours on the characteristic in figure 27.42 illustrate the energy penalty of non-optimal operation. Convention is to express these as constant energy product B_mH_m (as also seen in figures 27.8, 27.11, and 27.15, amongst others), rather than the actual energy density $½BH_{max}$.

A more realistic representation of the soft iron pole pieces requires the inclusion of the flux leakage coefficient k_t and mmf loss factor k_r via equations (27.19) and (27.18) within equation (27.52). Both loss factors reduce the amount of magnet energy delivered in to the gap

$$W_g = ½B_g H_g V_g = -½\frac{B_m H_m}{k_t k_r} V_m$$

By adding a coil of N turns with current i, as in figure 27.26, and using equation (27.30), the air gap energy becomes

$$W_g = -½\frac{V_m B_m}{k_t k_r}\left(H_m - \frac{Ni}{\ell_m}\right)$$

The last term represents the energy contribution from the coil. Since the flux in the circuit is $\Phi = B_m A_m$, and the total flux linkage with the coil is defined as $\lambda = N\Phi$, the gap energy can be written as:

$$W_g = ½\frac{1}{k_t k_r}\left(-V_m B_m H_m + \lambda i\right)$$

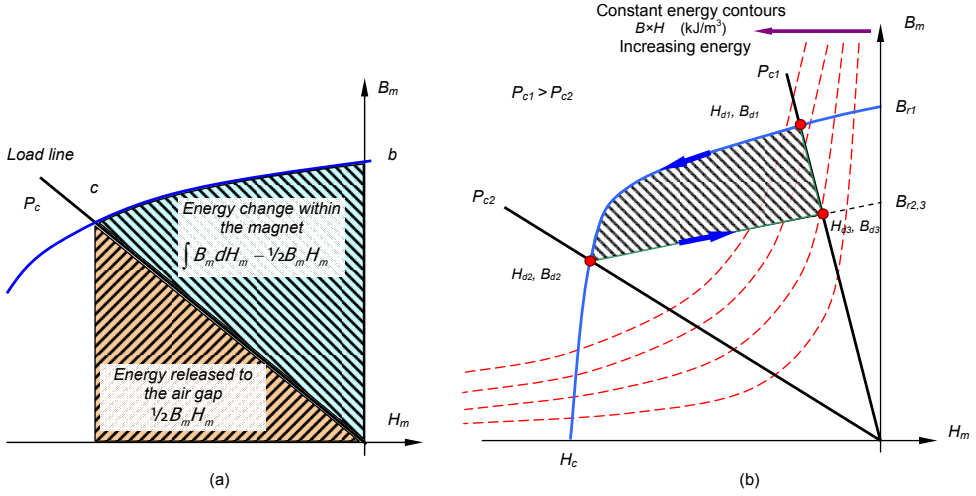


Figure 27.42. *Per unit volume change:*
(a) of the magnet's applied field energy and (b) magnetic field energy with load line slope.

Coenergy

Coenergy can be considered a dual of stored energy that is often used to compute forces on electromechanical systems with permanent magnets and current-carrying coils. Force and torque can be deduced by changes in energy with respect to position, but the process is complicated by the need to compute the energy that is sunk or sourced by the supplies driving the coils over the change in position. Conversely, computing the change in coenergy for constant currents directly yields the mechanical work done on the system.

An expedient lower bounds for coenergy computation, for the integral is $H_o = -H_c$. The coenergy density in the magnet w_m' is, for the case in which the flux density is aligned with the magnetization:

$$w_m' = \int_{-H_c}^{H} B(H)\,dH = B_r H + ½\mu H^2\Big|_{-H_c}^{H} = ½\mu\left(H + H_c\right)^2$$

The definition of magnet flux density B from equation (27.11) yields a simpler definition of coenergy within the magnet:

$$w_m' = \frac{B^2}{2\mu}$$

Stored energy and coenergy can be shown as regions associated to the demagnetization curve as shown in figure 27.43.

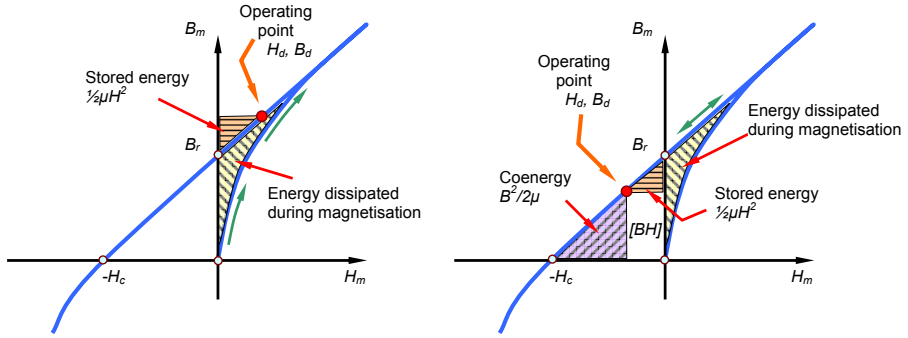


Figure 27.43. *Graphical representation of permanent magnetic energy components:*
(a) with applied bias $+H_d$ to enable operation in the first quadrant and (b) demagnetisation quadrant operation energy components.

As shown, the area swept in the first quadrant represents the energy per unit volume (the integral of $H(B)\,dB$) that was needed to magnetize the material. This energy can never be recovered, even if a large field intensity is applied to the magnet to push its operating point back into the first quadrant, as indicated in both parts of Figure 27.43. Remagnetisation only stores energy ($½\mu H^2$) in the magnet without possibility of extracting the energy associated with the magnet's magnetisation.

27.11 Force of attraction within an air gap

The equation $F = -dW/dx$ can be used to calculate the force of attraction between two pole faces bounding the air gap of figure 27.21 (figure 27.44) of area A_g and length ℓ_g. The energy in this gap is given by equation (27.52), which, with $B_g = \mu_o H_g$, becomes

$$W_g = ½\frac{B_g^2}{\mu_o} V_g = ½\frac{B_g^2}{\mu_o} A_g \ell_g \qquad (27.56)$$

Using Cartesian axes in the air gap, with flux along the gap length ℓ_g being in the x-direction, force is calculated by differentiating equation (27.56) with respect to ℓ_g, which yields

$$F_x = -\frac{dW_g}{d\ell_g} = -½\frac{A_g B_g^2}{\mu_o} = -½\frac{\phi^2}{\mu_o A_g} \qquad (27.57)$$

The definition of $-F_x$ acting into the air gap confirms this as a force of attraction, which will increase as the gap closes. Although ℓ_g, is not a parameter in equation (27.57), F_x is proportional to Φ^2, where a flux increase results from decreasing ℓ_g, which causes an increase in load line slope as the magnet's operating point moves up its demagnetization characteristic. Similarly, there are forces acting to retard

the lateral displacement of one pole piece with respect to the other (F_y, F_x), which may also be determined from equation (27.56) by splitting A_g into its y and z components and differentiating with respect to the appropriate axis.

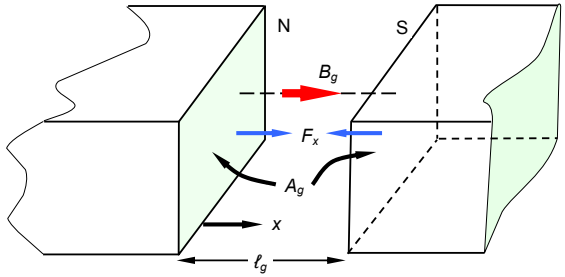
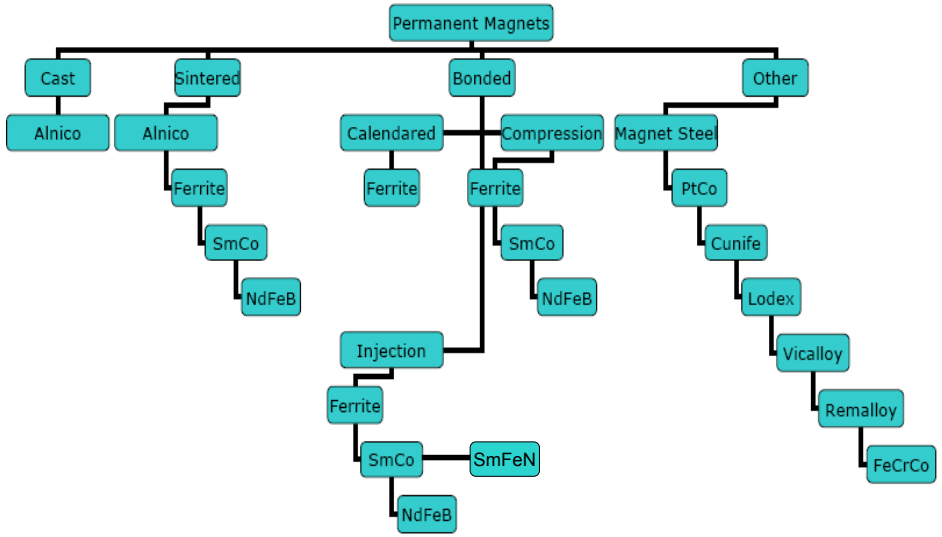


Figure 27.44. *Force of attraction between opposite pole faces bordering a uniform air gap.*

27.12 Appendix: Magnet processing and properties



Permanent magnet manufacturing Process

Sintered / fully dense, anisotropic magnets - maximum output: (Rare Earths, Ceramics, and Alnicos)

- Maximum energy product for magnet size and weight
- Restricted to simple geometries
- Brittle thus requires careful handling
- No dilution effect due to non-magnetic phase

Injection moulded - shape flexibility: (Rare Earths and Ceramics)

- Complex geometries
- Tight geometric tolerancing without finishing operations
- Relatively robust, resistant to chippage
- Insert and over moulding to reduce assembly costs

- Variety of pole configurations are possible
- Multistep and multi-component moulding to produce assemblies
- Dilution of magnetic phase produces lower energy product
- Anisotropic and isotropic powders offer a wide range of magnetic alignment and output options
- Relatively high tooling costs make them suited to high volume manufacturing

Compression bonded - low cost manufacturing: (Rare Earths and Ceramics)

- Higher loading than injection moulded, but lower than fully dense creates a compromise in energy product
- Limited to simple geometries: rectangles, cylinders, arcs – thin walled cylindrical shape possible
- Tight geometric tolerancing except in pressed thickness
- Brittle thus requires careful handling
- Isotropic powder allows complex magnetizing patterns

Casting: (Alnico)

Extruded: (Bonded NdFeB and Flexible)

Calendering: (Flexible)

Magnet Material

Neodymium-Iron-Boron - high energy

- Relatively abundant resource with large proven reserves
- Refining costs are moderate
- Manufacturing technology is established
- Highest energy product output of all commercially available PM materials
- High temperature, > 150ºC, applications require a compromise in energy product
- Tendency to corrode requires protective coating, chips, cracks, and brittle
- PrFeB variant for temperatures between a few degrees Kelvin and 135K

Samarium Cobalt - stable

- Relatively abundant resource with large proven reserves
- Manufacturing technology is established; dominated by 2-17 grades
- Second to NdFeB in magnetic output – high energy product and coercive force
- Excellent high temperature performance with grades available for use to 550ºC
- Corrosion resistance superior to NdFeB, but coatings generally advisable
- Brittle, chips, cracks easily, and hard to machine
- Refining costs are higher than for NdFeB

Ferrite (Ceramic) - low cost

- Abundant, low cost raw material
- Magnets are lowest cost option
- Manufacturing technology is well-established
- Lower magnetic output than the rare earth materials, but high coercive force
- Excellent high temperature performance with grades available for use to 250ºC
- Limited low temperature performance, generally to -40ºC
- Corrosion resistance is outstanding, but brittle and chips easily

Alnico – stable properties

- High corrosion resistance
- High mechanical strength, tough but brittle
- High temperature stability
- High cost
- Low coercive force and energy product

Table 27.6 Physical and mechanical properties of AlNiCo, FeCrCo, FeCoVCr, SmCo, NdFeB, and hard ferrite

Material and production method			Physical properties				Mechanical properties			
Material	isotropic or anisotropic	Production method	Coefficient of thermal expansion	Thermal conductivity	Resistivity	Density	Tensile strength	Compressive strength	Young's modulus	Hardness
			10^{-6}/K	W/m.K	μΩ×m	g/cc	MPa	MPa	MPa	HV
AlNiCo	i or a	casting or sintering	11	10 - 15	0.45 - 0.55		80 - 300	300 - 400	100 - 200	300 - 400
FeCrCo	i or a		10	10 - 13	0.7 - 0.8		1200 - 1400 600 - 700			300 - 350 400 - 500
FeCoVCr	a	casting	11.5		0.55 - 0.65		2000 - 5000 2500 - 3500			
$SmCo_5$	a	sintering	//6 ∟13	10 -13	0.5 - 0.6	8.2	30 -40	900 - 1000	100 - 150	500 - 600
Sm_2Co_{17}	a		//8 ∟11	10 -13	0.75 - 0.85	8.4	40 - 50	800 - 900	150 - 200	600 - 700
NdFeB	a		//3.4 ∟-5	8 - 10	1.4 - 1.6	7.4	80 - 90	1000 - 1100	150 - 200	500 - 600
hard ferrite	i or a		9 , 12, 11	4	> 10^4		50 - 60	600 - 700	15 - 200	500 - 600

27.13 Appendix: Magnetic Basics

$$B = \mu_o H$$

B = magnetic flux density or magnetic induction, Vs/m^2, 1 Tesla
μ_o = magnetic permeability of a vacuum = $4\pi \times 10^{-7}$ Vs/Am
H = magnetic field strength, A/m

If a material is present, the relation between magnetic field strength and magnetic flux density becomes

$$B = \mu_o \mu_r H$$

B can be split into the flux density in the vacuum plus the material part according to

$$B = \mu_o H + J$$
$$J = \mu_o M$$

This gives

$$B = \mu_o (H+M)$$

The definition of B yields

$$M = (\mu_r - 1)H = \chi_{mag} H$$

with $\chi_{mag} = (\mu_r - 1)$ = magnetic susceptibility. With superconductors (= ideal diamagnets), $\chi_{SC} = -1$.

27.14 Appendix: Magnetic properties for Sintered NdFeB and SmCo Magnets

Neodymium magnets are graded by the material they are made of. The higher the grade (the BH_{max}, following the 'N' in Table 27.3), the stronger the magnet.

Table 27.7 can be used to determine whether it is better to use sintered Neodymium Iron Boron or sintered Samarium Cobalt in a particular application.

Table 27.7 Typical magnetic properties for NdFeB and SmCo

	Neodymium Iron Boron (NdFeB)	Samarium Cobalt (SmCo)
High temperature applications	Higher H_{ci} materials can be used up to 200°C, moderate H_{ci} materials used up to 150°C. Low H_{ci} materials limited to 100°C	SmCo can be used at substantially higher temperatures than NdFeB. Continuous operation at temperatures above 250°C Also significantly better at temperatures below 100K
Loss of flux density at elevated temperatures	Loses 0.11% of B_r /°C. See NdFeB temperature effects on B_r & H_{ci}	Loses about 0.03% of B_r /°C
General environmental conditions	Neodymium Iron Boron is highly reactive to environmental conditions, while Samarium Cobalt magnets are very corrosion resistant.	
Humid environments	Surface treatment advisable, due to oxidation Options are nickel, IVD or polymer coatings.	Surface treatment is not required. Does not oxidize.
Hydrogen rich atmosphere	Not recommended, hydrogenation occurs, causing the magnet material to disintegrate	No known adverse effects
Cost of part	Lower cost generally	Higher cost than NdFeB because of Co content
Radiation environment	Damaged by radiation, mainly gamma rays	Higher grades are stable to radiation
Mechanical strength	Mechanically strong, not as brittle as SmCo	Brittle. Both alloys require diamond tooling, EDM, or abrasive grinding when machining.
Clean room environments	Surface treatment recommended	Surface treatment not required
Vacuum applications	Needs to be coated, with nickel or IVD recommended (metallic - do not outgas). Polymer coatings can be used, but not in an ultra high vacuum or high temperatures in a vacuum, due to outgassing.	Surface treatment is not required. However, fairly porous; parts may outgas for a limited duration before high vacuums can be achieved
High field requirements	Higher fields achievable, higher energy products (up to 440kJ/m^3)	Highest energy product is 240kJ/m^3, difficult to obtain higher.
Cryogenic temperatures	Only special formulation grades	Operates well in cryogenic temperatures.
Aerospace applications	Gaining use in aerospace applications, optionally with surface treatments. Used in aircrafts, missiles, and satellites.	Established in aerospace, military, and defence applications.
Salt, open seas, and salt spray environments	Surface treatment essential, limited life	Stable in this environment
Acid and alkaline environments	Surface treatment is necessary, limited life	Stable in this environment, however, qualification tests recommended
Thin walled, thin cross-section applications (a dimension below 1mm)	Mechanically stable, parts as thin as $\frac{1}{8}$mm	Poor in thin cross-sections. Under ½mm is not recommended
Single piece large parts	Better than SmCo, larger blocks can be sintered	Larger blocks (over 75mm in any dimension) are challenging
Nickel plating as per military specifications	Not available – proprietary plating to specifications, only electrolytic nickel	Electroless and electrolytic nickel plating
Plating as per military specifications	Not available. Plating as per proprietary specifications.	Not typically used.
Radial ring (for true radially oriented field)	Possible.	Not available
Sensitivity of flux density, B_r, and coercivity, H_c, to temperature changes	Temperature coefficient of B_r ranges from, negative, 0.11%/°C to 0.13%/°C. Higher coercivity materials (>160kA/m) are closer to negative, 0.11%/°C Temperature coefficient of H_c ranges from, negative, ½%/°C to ¾%/°C. The higher the intrinsic coercivity, the lower the temperature coefficient of H_c	Temperature coefficient of B_r ranges from negative, 0.03%/°C to 0.04%/°C. Series Sm_2Co_{17} less sensitive to temperature changes (about negative 0.03%/°C) than $SmCo_5$ (about negative 0.04%/°C) Temperature coefficient of H_c ranges from negative, 0.15%/°C to ¼%/°C. Series Sm_2Co_{17} less sensitive to temperature changes (about negative 0.15) than $SmCo_5$ (about negative ¼)

Based on table 27.7, in summary, do *NOT* use Neodymium Iron Boron magnets under the following conditions:

- In an acidic, alkaline, or organic solvent (unless the magnet is hermetically sealed)
- In water or oil (unless hermetically sealed, or a limited life results)
- In an electrically conductive liquid, such as electrolyte containing water
- In a hydrogen-containing atmosphere, especially at elevated temperatures. Hydrogenation, a process where the H_2 molecule reacts with the NdFeB, causing rapidly deteriorate
- Environments containing corrosive gasses, such as $C\ell_2$, NH_3, $N0_X$, etc.
- In the path of radioactive rays

27.15 Appendix: Magnetic Axioms

1. Flux lines, like electrical currents, follow the path of least resistance. In magnetic terms, this means that flux lines will follow the path of greatest permeance (lowest reluctance). Flux lines will always follow the shortest path through any medium. They therefore can only travel in straight lines or curved paths, and they can never take true right-angle turns.
2. Flux lines repel each other, hence never cross, if their direction of flow is the same.
3. For unsaturated ferromagnetic materials, flux lines will always leave and enter at right angles.
4. All ferromagnetic materials have a limited ability to carry flux. When saturated, they behave as though they do not exist (like air, aluminium and so on). Below the level of saturation, a ferromagnetic material will substantially contain the flux lines passing through it. As saturation is approached, because of axioms one [1] and two [2], the flux lines may travel as readily through the air as through the material (which appear like air when saturated).
5. Flux lines will always travel from the nearest north pole to the nearest south pole in a path that forms a closed loop. They need not travel to their own opposite pole; although they ultimately do if poles of another magnet are closer and/or there is a path of lower reluctance (greater permeance) between them.
6. Magnetic poles are not unit poles. In a magnetic circuit, any two points equidistant from the neutral axis function as poles, so that flux will flow between them (assuring that they meet the other conditions stated above).

Table 27.8 The relationship between magnetic parameters in cgs and SI units.

Quantity	symbol	Gaussian	Gaussian (cgs units)	S.I.	S.I. Units	Conversion factor (cgs to S.I.)
Magnetic Induction	B	gauss	G		Tesla, T	$1G = 10^{-4}$ T
Applied Field strength	H	Oersted	Oe		Am^{-1}	$10^{-3} / 4\pi$
Magnetisation	M		emu cm^{-3}		Am^{-1}	10^3
Magnetisation	$4\pi M$		G		-	-
Magnetic Polarisation	J		-		T	-
Specific Magnetisation	s		emu g^{-1}		JT^{-1} kg^{-1}	1
Induction in free space		B=H	G	$B = \mu_o H$	T	$1G = 10^{-4}$ T
Induction in medium		B=H+4πM	G	$B = B_o + \mu_o M$	$B = \mu_o (H + M)$	$1G = 10^{-4}$ T
Permeability of free space	μ_o	unity	Dimensionless	$4\pi\ 10^{-7}$	H m^{-1}	$4\ \pi\ .\ 10^{-7}$
Relative Permeability	μ_r		-		Dimensionless	-
Susceptibility	χ		emu cm^{-3} Oe^{-1}		Dimensionless	$4\ \pi$
Maximum Energy Product	BH_{max}		M G Oe		kJ m^{-3}	1 MGOe $=10^2 / 4\ \pi$
G = Gauss, Oe = Oersted, T = Tesla						

Reading list

http://www.magnets.bham.ac.uk/magnetic_materials/
http://www.mceproducts.com/
http://www.dextermag.com/Permanent-Magnet-Materials.aspx
http://www.arnoldmagnetics.com/
http://www.intemag.com/magnet_materials.html
http://www.magnetweb.com/
http://www.femm.info/wiki/PMEnergy

28

Contactors and Relays

A relay is an electrically operated switch. Generally, a relay is an electromagnetic device, within which an electro-magnet is fixed to cause controlled movement either by magnetic attraction or magnetic repulsion. Other hardware attached to the moving magnetic portion of the component, such as relay contacts, will cause switching of electrical circuits.

Differentiation between contactor (or relay), circuit breaker, and mechanical switch is as follows:

A *relay* is an electromagnetic device for remote or automatic control that is actuated by variations in conditions of an electric circuit and which, in turn, operates other devices (such as contacts) in the same or more often, a different (isolated) higher power circuit. A high power-handling relay is termed a contactor.

A *circuit breaker* is a switch that automatically interrupts an electric circuit under an infrequent abnormal condition; e.g., a fault condition such as an overload or rupture of either high voltage or high current or both.

A *mechanical switch* is a device for making, breaking, or changing the connections in an electrical circuit, usually mechanical and operated manually.

28.1 Mechanical requirements for relay operation

A relay comprises a coil (of insulated copper wire), a plunger, and contacts, all housed together within a case or body. Electrically energising the coil creates a magnetic field that moves the plunger. The movement of the plunger causes a moving low resistance contact to move toward (or away from) a non-moving fixed contact. It is the coming together of these contacts (make) or separation of them (break) that has the effect of switching. The fixed contact may be mounted on compliant springs or fixed brackets. The movable contact (single throw, *ST*) is mounted on some form of spring that can be deflected or on a hinge arm. The force and travel needed for such motions serve a number of purposes.
Relay coils are wound for a particular voltage, commonly 24V dc, a safe voltage for control automation, 115V ac is often used in North America and 230V ac is found in EU systems. A coil will activate its plunger as the voltage increases from zero to its normal rated level, *nominal voltage*, and this is know as the *pull-in voltage* (often around 80% of nominal). As power is removed from the coil the plunger returns to its rest position as the voltage drops below the *drop-out voltage* (often around 20% of nominal).
Before the armature is actuated for a relay with double throw contacts, *DT*, the movable contacts must be held against the normally closed fixed contacts by a spring force sufficient to establish good electrical contact. When the armature is actuated, a number of things happen. Each movable contact is pushed or pulled away from the corresponding, normally closed, fixed contact. This requires a force sufficient to overcome one or more springs. Also, there is friction between the contacts if they slide before they separate, and in any actuator pivots. As contact motion takes place, various springs deflect according to Hook's law, and inertial forces must be overcome.

After the accelerated motion of contact transfer, the movable contact impacts and decelerates as it reaches its normally open fixed contact. Both contacts must deflect or deform to some degree as the desired contact force builds ups. Over-travel can be employed to provide contact cleaning action, through sliding, and to compensate for contact wear or erosion. When there are multiple sets of contacts (poles), allowance is made for manufacturing tolerances for the various stages of travel required for and their associated spring forces.

The word *normally* (*N*) refers to a de-energized relay condition (no electrical power into the coil).
The subsequent word *open* (*O*) or *closed* (*C*) refer to the position of the contacts in a de-energized condition, viz., *normally open NO* or *normally closed NC*.

28.2 Relay Contacts

28.2.1 Contact characteristics

Contact characteristics that affect switching performance are:

- Electrical conductivity
- Thermal conductivity
- Hardness, limit of elasticity: Young's Modulus
- Resistance to erosion, welding or electrical sticking, cold welding, mechanical wear, oxidation, and atmosphere contamination (chemically active).
- Tendency to bounce on impact, gaseous absorption, catalytic polymerization of hydrocarbons, metal transfer at contact closure and arcing at opening.

Besides the physical and chemical properties of the contact metal, there are some geometrical and dynamic considerations:

- Shape of contacts
- Force between contacts
- Amount of slide or *wipe*
- Amount of rolling or twisting motion
- Supporting structure resiliency and its tendency to enhance or inhibit bounce or chatter.

When contacts meet, the metal at the point of the contact deforms until the actual touching area supports the contact force and provides metal-to-metal contact, unless some foreign material interferes. Deformation is at the point of contact, either in the elastic or plastic modes, contributes to the amount of contact bounce. Microscopically, many actual points of contact (often referred to as *a-spots*) form the electrical conductor and carry the current. The contact interface is also subject to mechanical abrasion and metal *galling* as it rubs, and *cold welding*. The surface absorbs a monomolecular layer of volatile molecules in direction proportion to the molecular weight and concentration of the volatile material and the ambient pressure and inversely proportional to the temperature. Water vapour is also a common substance forming thin absorbed layers.
Each metal has particular pertinent chemical properties. Silver and its alloys, which have excellent electrical and thermal characteristics, tend to combine chemically with gaseous compounds of sulphur, the halogens (fluorine, chlorine, bromine, and iodine), and silicones, to form high electrical resistance, in the form of hard coatings. Unlike other noble metals (gold, platinum, rhodium, iridium, palladium, and ruthenium, all of which are used in contacts), silver has no catalytic effect (polymerization) in the sense of changing, under sliding pressure, the absorbed hydrocarbon molecules into a solid hydrocarbon material. Arcing, however, can accomplish the precipitation of solid carbon or carbonaceous products, usually in a ring around the actual point of contact.

Molybdenum, tungsten, nickel and mercury are used alone or as alloying or sintering ingredients. Cadmium oxide, tungsten carbide, tin, magnesium, and carbon are sometimes added to silver to inhibit sticking or welding, particularly in high-current relays or contactors. When contacts are within an inert gas, like nitrogen, other contact materials are applicable.

28.2.2 Contact materials

Low voltage and current contact material operating boundaries are summarised in figure 28.1.

Fine Silver, Ag
Fine silver has the best electrical and thermal properties of all metals. However, it is affected by sulphidation, which forms a film on its surface that increases contact interface resistance. Thus contact pressures must be sufficient to break through the film. While such pressures have no appreciable effect on silver-cadmium contacts, they do result in increased material wear of fine silver contacts. Also, an interface voltage of several hundred millivolts can occur with fine silver contacts because of the sulphide film. Controlled arcing burns off the sulphidation, and contact over-travel wipes away the residue. Breaking through this film generates electrical noise, consequently, fine silver contacts are not used for low-level switching, such as audio circuits. Therefore, fine silver and silver alloy contacts are for use in circuits of 12V, 0.4A, or more.

Gold-Flashed Silver, Au flash 3µm Ag
For relays that are inoperative for long periods before initial operation, sulphidation of silver contacts can result in an impregnable contact interface resistance. Instead of specifying silver contacts for such conditions, gold-flashed silver contacts are specified. Gold flashing on each contact results in minimal sulphidation/corrosion/oxidation during storage, and provides good electrical *make* upon contact. Because gold has a low boiling temperature, any flashing burns off after just a few switch cycles if arc voltage and current limits are exceeded. The silver under-layment is then exposed, which may develop a sulphide film. Generally, gold-flashed contacts should not be subjected to arcing. Gold-flashed silver has the same qualities as 10µm plate Au, but is less durable.

Gold Overlay, plated 10µm Au
A common contact for use in dry and low-load level switching circuits (>1mA/100mV) is gold overlay. The overlay is of sufficient thickness so as to not wear through to the base metal unless subjected to arcing conditions. This plating is removed by friction and erosion after around 1 million switching cycles in *dry circuits* (no current is switched).

Silver Nickel, AgNi 0.15
Depending on the application, material transfer may be prevalent with fine silver contacts. Typically, material tends to accumulate in the centre of one contact, while the loss of material on the other contact leaves a hole, or *pit*, leading to premature contact failure. In such applications, it is desirable to use fine grain silver contacts, in which alloying with 0.15% nickel gives the contacts a fine grain structure. As a result, material transfer is evenly distributed across the entire contact surface, resulting in longer contact life. Minimum contact load is 20V/50mA for a single contact.

Silver Cadmium Oxide, AgCd0
Silver cadmium oxide contacts are used for switching loads that produce a high-energy arc. Such contacts are less electrically conductive than fine silver contacts, but have superior resistance to material transfer and material loss due to arcing. They exhibit greater interface resistance between mated contacts, and a slightly greater contact assembly temperature rise. It is used for high ac loads because it is more resistant to welding at high switching current peaks. Material evaporates/wears evenly across the surface. It is not recommended for strong dc breaking arcs because of the resultant wear (one-side reductions). The minimum arc voltage rating of silver cadmium oxide is 10V, minimum contact load is 20V/50mA, and, like fine silver contacts, the silver in this alloy will oxidize and sulphidate. Therefore, an arc is necessary to keep the contacts clean. This contact material is being replaced by Silver Tin Oxide ($AgSn0_2$) and AgMe0.

Silver Tin Oxide, $AgSn0_2$
Tin oxide makes silver more resistant to welding at high making current peaks. It has a high burn out resistance when switching high loads and a low degree of material migration under dc loads. Minimum contact load is 20V/50mA. Such properties are useful where high inrush currents occur, such as lamp loads including fluorescent lighting. Silver Tin Oxide is frequently chosen as the replacement relay contact material for Silver Cadmium Oxide, which is being withdrawn.

Silver Tin Indium Oxide, AgSn0In0
Silver tin indium oxide contacts, although not readily available, exhibit better resistance to arc erosion and welding than silver cadmium oxide contacts. They are less electrically conductive and are harder than silver cadmium oxide contacts. They have greater interface resistance between mating contacts and, therefore, a greater voltage drop and temperature rise. Silver tin indium oxide is more expensive than silver cadmium oxide, and the relay is limited to use in applications such as incandescent lamp loads and capacitors where there is an inrush current during contact bounce. For low and medium power resistive and inductive loads, silver cadmium oxide is most commonly used. Properties are similar to Silver Tin Oxide but it is more resistant to inrush. Minimum contact load is 12V/100mA.

Silver Copper Nickel, Ag (97-98%), Cu+Ni
The copper and nickel contents give the hardness. Single contact minimum load is 20V/50mA. Silver copper nickel contacts are for use in high inrush dc applications such as incandescent lamps and capacitive loads. These contacts exhibit good resistance to welding, with a long contact life, but tend to oxidise at higher temperatures.

Gold Silver Nickel Alloy
Gold silver nickel alloy contacts are used for switching loads generally of less than 1A, and are characterized by less electrical noise on make and break than fine silver contacts. Gold diffused silver contacts offer characteristics similar to gold silver nickel alloy, but are less expensive. Other gold-based alloy contact materials include AuPd (Pd <5%) and AuCo for galvanic deposited gold layers.

Palladium, Pd + (Cu/Ni/Ag, Rh coated)
Palladium contacts do not sulphidate or oxidize, and so offer extremely low electrical noise levels. They have an electrical life expectancy of approximately 10 times that of fine silver contacts. However, because of relatively poor conductivity properties, load currents are limited to about 5A. Palladium contacts require 0.15mm to 0.30mm over-travel to insure good wiping action. Because of this, they are used primarily on telephone-type relays, that is, relays on which the contact arms are parallel to the length of the coil, and on which such over-travel is easy to obtain. Also, palladium contacts should be bifurcated (multiple contacts operating in parallel) to help ensure circuit continuity on contact closure.

Tungsten, W
More resistant to welding at high loads than hard silver, with a high burnout resistance, tungsten is a good standard contact material, used on some heavier duty relays. Minimum contact load is 20V/50mA for a single contact. Tungsten contacts are for use in high voltage applications, usually where highly repetitive switching is required. Tungsten has a melting temperature of 3,380°C, which gives it excellent arc-erosion resistance. Tungsten may develop unwanted oxide films, especially when used as the anode contact in some dc applications. Therefore, tungsten is often used as the cathode contact, and a palladium alloy is used as the anode contact. Such a combination also minimizes contact interface resistance and material transfer.

Platinum Iridium, PtIr
Platinum has a high corrosion resistance, being much less prone to erosion in comparison with silver. In pure form, it is low in hardness and therefore is generally alloyed with other materials. Platinum-iridium combines hardness with excellent resistance to the formation of arcs.

Table 28.1: Contact materials for low voltage relays

Contact material	Typical features	*V, I, P* values
$AuAg_8$ (Gold F)	For low-resistance applications at low load Low constant contact resistances For measuring currents, dry switching	µV - 24V µA - 0.2A < 5W
Rh	For high-resistive applications at low load Galvano-technical contact coating in µm range for reed contacts with higher endurance	< 150V < 2A
PdNi (Ni content 20 to 50%)	Galvano-technical contact coating with similar features to Rh Powder metallurgical contacts are also possible Laminar creep of material	< 150V < 5A
Ag-pure and Ag fine grain (Ni content 0.15%)	Most common contact material; universally applicable Sulphur sensitive, therefore often flash-golded Also suitable for alternating current	1V - 150V 50mA - 100A > 1W
AgPd (Pd content 30 to 50%)	Important material in communications technology Good burn-up resistance Sulphur insensitive Slightly higher contact resistances than with Ag	1V - 150V 50mA - 5A
AgNi (Ni content 10 to 20%)	Important material for inductive loads Suitable for make currents up to 50A Good burn-up resistance Low susceptibility to welding Higher contact resistances than with Ag	6V - 380V 10mA - 100A
$PdCu_{15}$	For lamp loads in the automotive field (pulsed operation) Burn-up resistant, long endurance Laminar creep of material Higher specific insulation resistances that are not constant	6V - 24V 5A - 20A
AgCd0 (Cd content 8 to 15%)	Material for alternating current Burn-up resistant Low susceptibility to welding	12V - 380V > 0.5A > 10 W
$AgSn0_2$ (Sn content 8 to 15%)	Material for direct and alternating current Burn-up resistant Low susceptibility to welding Environmentally friendly (replacement for AgCd0)	> 12V - 380V > 0.5A > 10W
W	For high make currents (as pre-travel contact) For higher switching rates Burn-up resistant, low susceptibility to welding Subject to corrosion	> 60V > 1A > 50W

Mercury, Hg
Mercury has a melting temperature of -38.9°C. Thus, as used in relays, it is in a liquid state. Mercury clings to the surface of any clean metal, and is used as the contacts in mercury-wetted reed relays. It has good electrical conductivity and, being liquid, there is no material transfer build-up from contact to contact. Any such material transfer is negated by the fact that when the contacts open and the mercury returns to the pool in the bottom of the relay, fresh mercury takes its place at the next switch operation. Mercury has a boiling temperature of 357°C. Because of this, mercury contacts cannot switch currents of more than a few amperes.

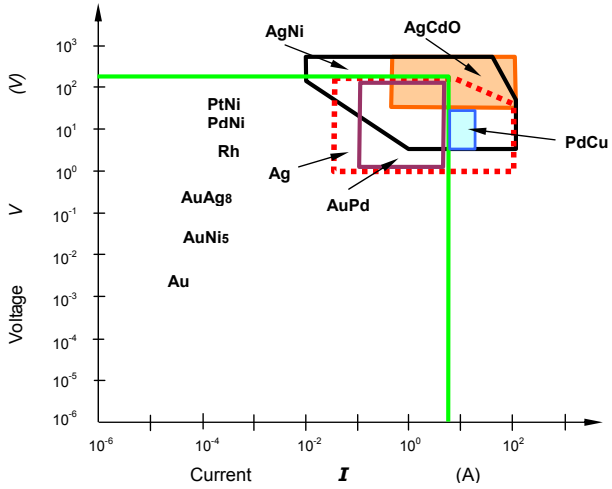


Figure 28.1. *Low voltage and current contact materials.*

28.2.3 Contact Life – material loss and transfer

The electrical life expectancy of general purpose and power relays is normally rated to be 100,000 operations minimum, while mechanical life expectancy may be in excess of one million operations. The reason electrical life is rated so low compared with mechanical life is because contact life is application dependent. The electrical rating applies to contacts switching their rated loads. Rated electrical life also takes into consideration arc destruction of the contacts. By the use of appropriate arc suppression, contact life may be lengthened.
Contact life is terminated when the contacts stick or weld, or when excessive material is lost from one or both contacts and a good electrical *make* is not possible. These conditions are the result of cumulative material transfer during successive switching operations, and of material loss due to splattering. Material transfer occurs due to Joule I^2R heat. Material loss is due primarily to splattering of the molten and boiling metal as contacts bounce on *make*.
In dc applications, metal migration is predictable in that one contact is always negative, and the other, positive. In ac applications where switching is at random, either contact may be negative or positive when arcing occurs. Migration will not be in the same direction each time the contact *breaks*, and material loss from either contact should not be significant, unless load conditions cause splattering.
Controlled arcing of short duration can be beneficial in achieving the rated life of the contacts, because such arcing burns off any deposits on the contacts that might prevent electrical *make*. Such control is achieved by arc suppression. Unless arcing and/or contact over-travel cleans the contacts, films may develop on the contact surfaces, or foreign matter may collect. For this reason, it is best to apply general purpose and power relays only in applications where the load voltage (or counter emf) and current are in excess of the arc voltage and current ratings of the contacts.

A method of quenching an arc between separating contacts is with an *RC* network placed directly across the contacts. As the contacts just begin to separate and an arc ignites, load current feeding the arc is shunted into the capacitor through the series resistance, depriving the arc of some of its energy. As a result, arc duration will be shortened and material loss will be reduced.

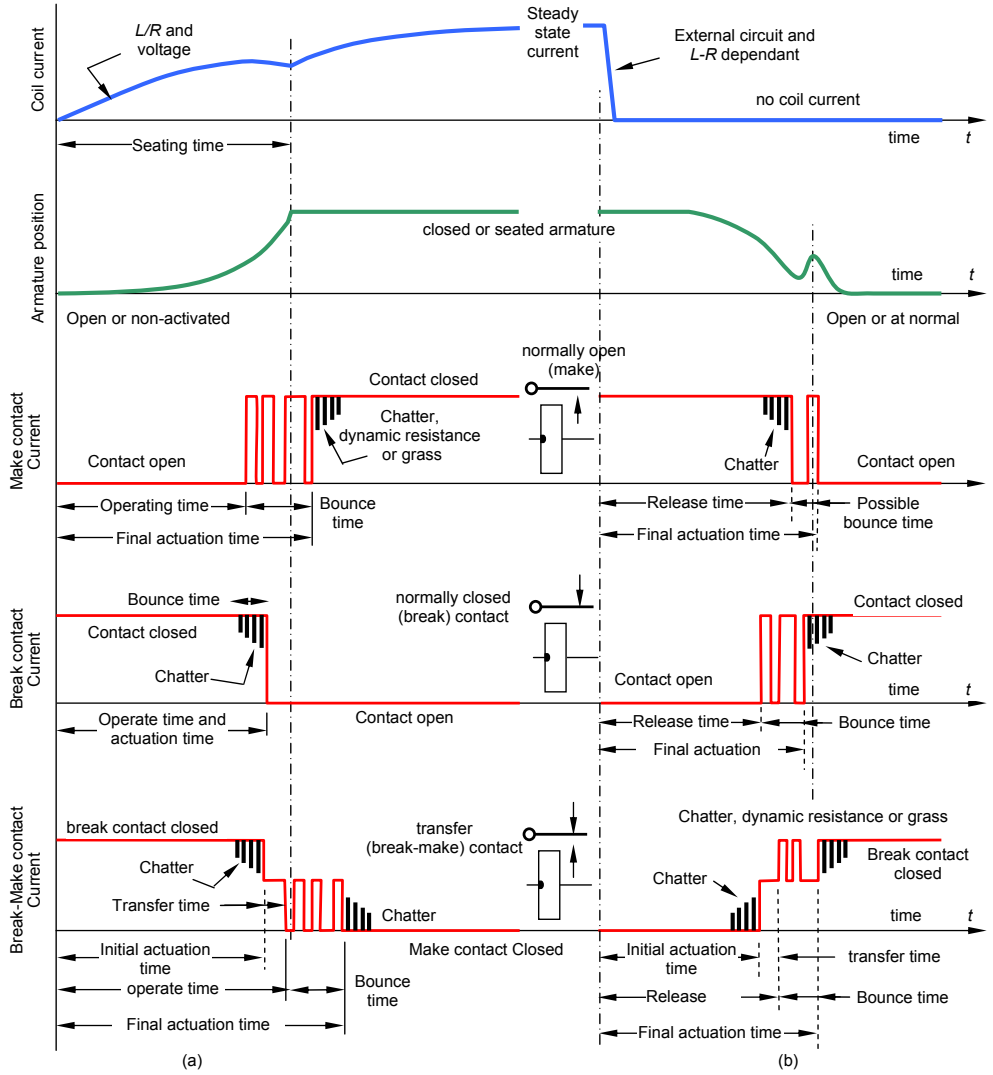


Figure 28.2. *Voltage and current waveforms typical of relay (a) pick-up and (b) drop-out. Contact currents for non-inductive load*

28.3 Defining relay performance

There is a sequence of events in relay pick-up (operate) and drop-out (release) with respect to current rise and decay. The events are defined in terms of duration of coil current, armature motion, and contact actuation. Figures 28.2 (a) and (b) show contact performance as a series of time domain waveforms, for a relay with a normally open contact, a normally closed contact, and a transfer (break-make) contact. Figure 28.3 shows the relationship between parameters defining relay performance and their definitions, which follow.

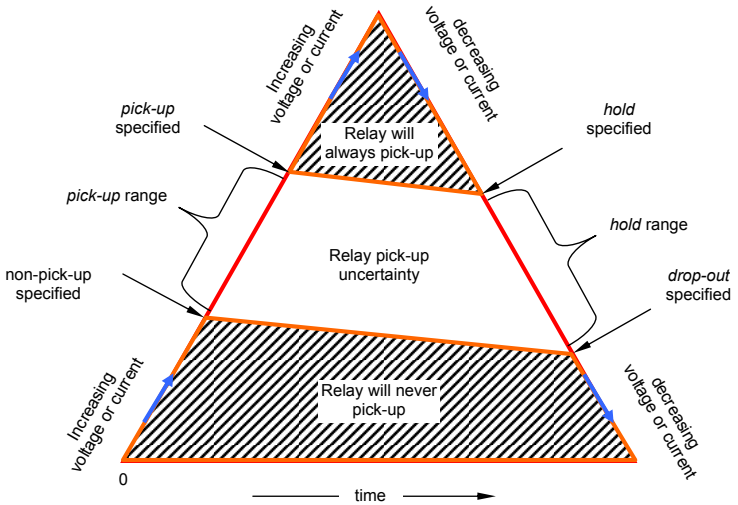


Figure 28.3. *Graphical presentation of relay performance related definitions.*

Pick-Up (figure 28.2a): Upon coil energization, current begins to rise exponentially at a decreasing rate, but no armature movement occurs until the power develops sufficiently to operate the contact spring load. This period is sometimes referred to as *waiting time*. Contact actuation occurs during the armature movement. The *final actuation time* exceeds the *initial actuation time* by the amount of the contact bounce. For normally closed contacts, operate time and initial operate time are identical. On break-make contacts, the time interval between initial opening of the normally closed contact and closure of the normally open contact is called *transfer time*.

Drop-out (figure 28.2b): On de-energization of the coil, the magnetic flux does not cease immediately. The length of time it persists depends upon the release characteristics of the coil (fast-to-release, slow-to-release, and the like). The sequence of events described under pickup is essentially reversed under dropout. A normally open contact may be momentarily re-closed as a result of armature rebound off the backstop. This effect, which is not always present, depends on many factors, such as contact spacing, contact spring load, backstop design, etc.

Closing Arc
As a contact is activated, and two points come together to carry current, an arc forms. This causes material to evaporate, and if high transient currents are present (say starting a motor or a fluorescent lamp) then large portions of the contact surface may melt causing the contacts to weld. The process is reinforced by contact bounce. As the contacts close, the arc is suppressed; it appears as a peak.

Opening Arc
To break a circuit, a contact will open. As it does, the effective contact surface is reduced due to the decreasing contact force and movement. The current flow is the same, and therefore the current density in the remaining pathway increases up to the melting point. An explosion like process can occur as the contact material springs out. An arc may be produced by:

- resistive and capacitive loads in conjunction with high voltages
- inductive loads

Permanent or sustained arcs are produced mainly by dc currents. Alternating current quenches the arc when the current crosses the zero current level. Arcs are influenced by contact material, a reduction in arcing voltage and arcing current and the speed of the breaking elements. When switching high dc loads, a larger contact gap and blow out magnet may be critical.

Relay switching performance is affected by high ambient temperatures (as shown in figure 28.4), humidity, dust, and contaminant gases. A relay itself creates heat and oxidants as it operates. The other influencing factors on relays electrical service life is the arc produced when the contacts open and close. Contact friction, clearance mechanical quality, etc. are of lesser significance.

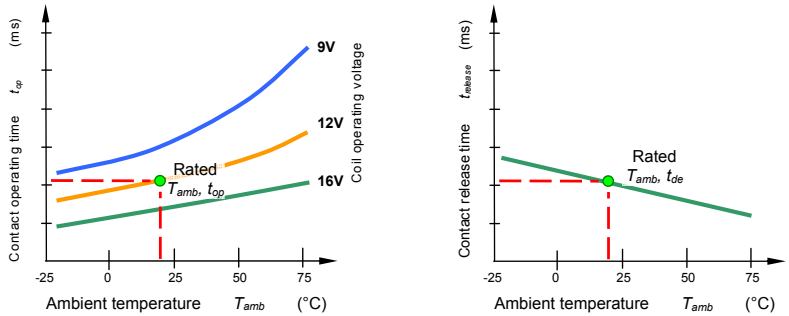


Figure 28.4. *Contact operate and release time dependence on coil temperature.*

28.4 AC and DC relay coils

i. AC across a DC coil relay
In theory, ac can be used to operate a dc relay but this is not so in practice. Since alternating current decreases to zero every half-cycle (100 times per second for 50Hz voltage or 120 times per second for 60Hz voltage), the relay armature tends to release every half-cycle. This continual movement of the armature not only causes an audible buzz, but will cause the contacts to open and close as the armature moves.
In order to operate a relay from ac, a device known as a shader ring (or shader coil) is used on top of the core. Because of the shader ring, the magnetism developed in part of the core lags the magnetism of the remainder of the core. That is, there is a slight phase displacement between the magnetism of parts of the core. Thus, as unshaded-core magnetic energy decreases to zero every half-cycle, the magnetic energy decreases to zero every half-cycle, the magnetic energy still present in the shaded portion of the core holds the armature sealed. By the time the energy in the shaded portion decreases to zero, coil and unshaded core magnetic energy have begun to increase once again as current increases in value.

ii. DC across an AC coil relay
An ac relay may be operated from dc provided two precautions are taken.

i. The first precaution is to provide some type of a residual break between relay core and armature to prevent the armature from sticking as a result of any appreciable residual magnetism remaining in the core after coil power is removed. AC relays are so constructed that when the armature is in its seated position, it physically (magnetically) touches the core. (On dc relays, a small copper pin in the armature effectively prevents the armature from coming in magnetic contact with the core.) As long as the ac relay is operated from AC voltage, there is no problem with residual magnetism holding the armature seated after release of coil power. But when an ac relay is operated from dc voltage, there is a danger that residual magnetism may hold the armature seated. At the very least, the presence of residual magnetism in the core causes a reduction in the dropout voltage of the relay. To negate the effects of residual magnetism, a small piece of mylar tape may be stuck to the top of the ac relay core. This tape is extremely durable, and should last for perhaps hundreds (if not thousands) of operations. The tape is 0.05mm to 0.1mm thickness.

ii. The second precaution to be taken is to ensure the dc voltage used is less than the ac voltage rating of the coil. To use an ac coil on dc requires lowering the amount of dc voltage to that value where coil power is within maximum limits. When using rectified ac to operate any relay coil, it is best to use capacitive filtering to reduce the voltage ripple to less than 25%

28.5 Temperature consideration of the coils in dc relays

Relays and temperature are intertwined, where a rise in temperature causes a rise in electrical resistance. When a relay is exposed to various temperatures, its operating characteristics, specifically here, pull-in and dropout, change dependent upon the temperature. The most notable changes occur in the pick-up voltage V_{PI} and coil resistance R_C. The coil winding of a relay is produced with copper wire and thus the coil resistance varies with the temperature coefficient of copper. For the temperature range that a relay will normally be exposed to, the resistance change in copper is of the form:

$$R_1 = R_o \times (1 + \alpha \times \Delta T) \qquad (28.1)$$

where: R_1 = Resistance at temperature T_1, Ω
R_o = Resistance at temperature T_o, Ω
$\Delta T = T_1 - T_o$, change in temperature from T_o, °C
α = Slope of a line from a point (-234.5, 0) through the point (T_o,1)
α = 0.003929 at T_o = 20°C or 0.003853 at T_o = 25°C, °C^{-1}
T_1 = New operating temperature, °C
T_o = Reference temperature, where 20°C or 25°C are typically used references, °C

For a dc relay, the magnetic force developed is proportional to the Ampere-turns developed in the coil. Since the mechanical forces are nearly constant over the normal temperature range (and the number of turns is fixed), the pick-up current, I_{PI}, will be constant. If pick-up current is constant and coil resistance varies, the pick-up voltage, $V_{PI} = I_{PI} \times R_C$, varies directly as the coil resistance. This leads to a simple mathematical method to determine coil resistance and pick-up voltage V_{PI1} at any temperature if a reference point V_{PIo} is known.

$$V_{PI1} = V_{PIo} \times (1 + \alpha \times \Delta T) \qquad (28.2)$$

While temperature changes affect relay parameters, the power dissipated within the relay also affects the temperature in most applications. The power dissipated within the relay comprises a number of components. Heat generated in the relay coil when voltage is applied to it. This heat creates a temperature rise in the relay coil and package. The temperature rise is dependent upon several factors such as the volume of copper wire used, insulation thickness, insulation type, bobbin material, bobbin thickness, terminal size, conductor size, and several other factors that are design related. Each of these factors will either enhance or resist the flow of generated heat from the coil assembly and into the ambient air. For a given relay design, these factors can be accumulated into the 'coil to ambient thermal resistance' of the relay, in °C/W. The thermal resistance is analogous to the electrical resistance and the temperature rise created by coil power dissipation follows the equation:

$$T_{RC} = R_{\theta CA} \times P_d = R_{\theta CA} \times \frac{V_A^2}{R_C} \qquad (28.3)$$

where: T_{RC} = Temperature rise caused by coil power dissipation, °C
$R_{\theta CA}$ = Thermal resistance from coil to ambient, °C/W
P_d = Steady-state power dissipated in the coil, due to coil resistance R_C and coil voltage V_A, W

For normal relay temperature ranges, this relationship is nearly linear and consistent under the following conditions:

- The relay is in still air and not subjected to significant airflow or the value of θ_{CA} was determined with an airflow identical to the end application.
- All power calculations deal with the coil resistance at the final coil temperature T_C attained. If only room temperature coil resistance were used, the resulting non-linearity would result in significant errors at higher temperatures.

The value for thermal resistance is specified when the relay carries no load current. The final coil temperature can be calculated using manufacturer's parameter data under no load relay conditions.

Under contact load conditions, the contact power dissipation may be treated as a separate heat source that adds heat into the relay package. Its effect on coil temperature is dependent upon many factors including package size, contact to coil distance, contact terminal size, connecting wire size, shared thermal paths, etc. These factors can be lumped into a contact to coil thermal resistance, which leads to

$$T_{RL} = R_{\theta CC} \times P_K = R_{\theta CC} \times I_L^2 R_K \qquad (28.4)$$

where: T_{RL} = Temperature rise caused by load dissipation in the contacts, °C
$R_{\theta CC}$ = Thermal resistance from the contacts to the coil, °C/W
P_K = Power dissipated in the contacts, W
R_K = Resistance of contact circuit, assumed temperature independent, Ω
I_L = Load current flowing through the created by the contacts, A

Based on practical data, the contact temperature rise can be approximated by

$$T_{RL} = k_R I_{IL}^{1.85} \qquad (28.5)$$

The final coil temperature rise above ambient, is the sum of the two heat source components, self-heating and that due to contact heating, namely

$$T_C = T_A + T_{RC} + T_{RL} = T_A + R_{\theta CA} \times \frac{V_A^2}{R_C} + k_R I_{IL}^{1.85} \qquad (28.6)$$

Example 28.1: *Relay coil thermal properties*

For a dc relay under the following no-load conditions:
T_o = 20°C, $V_o = V_{PI}$ = 6.8V, R_o = 90Ω, V_A = 13.5V(the applied coil voltage)
$R_{\theta CAJ}$ = 40°C/W, T_A = 85°C, I_l = 0A (no load current)

determine:
i. Cold-start pick-up voltage (with the coil previously un-energized) and coil resistance at T_A
ii. Final steady-state coil temperature T_C and resistance for an applied coil voltage V_A
iii. Hot-start pick-up voltage (after coil energized at V_A) at T_A and V_A; and
iv. for a 20A contact load, with k_R = 0.29 modelling the relay contacts, determine the new steady-state thermal operating conditions.

Solution

The pick-up current is virtually independent of thermal conditions since it is based on magnetic circuit Ampere-turns. Therefore the expected coil current is V_{PI} / R_o = 6.8V/90Ω = 75.56mA, at all operating temperatures.

i. The resistance at an operating temperature of 85°C is given by

$$R_1 = R_o \times (1 + \alpha \times (T_1 - T_o)) = 90\Omega \times (1 + 0.003929 \times (85°C - 20°C)) = 90\Omega \times 1.2554 = 113\Omega$$

The voltage at this temperature is obtained using the same scaling factor since the necessary current is assumed independent of temperature.

$$V_1 = V_o \times 1.2554 = 8.54V$$

ii. Since $T_C = T_A + T_{RC}$ and $P_d = V_A{}^2 / R_C$ then

$$T_C = T_A + R_{\theta CA} \times \frac{V_A^2}{R_C} = 85°C + 40°C/W \times \frac{13.5V^2}{R_C}$$

Since R_C is temperature dependant, an iterative solution is necessary to determine T_C and R_C. After several iterations T_C =140°C, which leads to a coil resistance at this temperature of

$$R_C = R_o \times (1 + \alpha \times (T_1 - T_o)) = 90\Omega \times (1 + 0.003929 \times (140°C - 20°C)) = 90\Omega \times 1.4715 = 132.4\Omega$$

iii. The hot-start relay pick up voltage is therefore

$$V_1 = 6.8V \times 1.4714 = 10.0V$$

Thus the necessary ampere-turns relay coil current is 10.0V/132.4Ω = 75.58mA, independent of temperature.

iv. For a 20A load current through the contacts:

$$T_C = T_A + R_{\theta CA} \times \frac{V_A^2}{R_C} + k_R I_{IL}^{1.85} = 85°C + 40°C/W \times \frac{13.5V^2}{R_C} + 0.029 \times 20^{1.85}$$

Again, since R_C is temperature dependant, an iterative solution is necessary to determine T_C and R_C. After several iterations T_C =146.5°C, which leads to a coil resistance at 146.5°C of

$$R_C = R_o \times (1 + \alpha \times (T_1 - T_o))$$
$$= 90\Omega \times (1 + 0.003929 \times (146.5°C - 20°C))$$
$$= 90\Omega \times 1.4970 = 134.73\Omega$$

The hot-start relay pick up voltage is therefore

$$V_1 = 6.8V \times 1.4970 = 10.18V$$

Thus the necessary ampere-turns relay coil current is 10.18V/134.73Ω = 75.56mA, independent of temperature.

♣

The values obtained in example 28.1 apply to dc relay coils operated continuously at these values. Intermittent duty (with short, that is, less than one minute, on-times and longer off-times) may result in substantially lower temperatures. Therefore, if a specific duty cycle is given for the relay operation, testing at these conditions could yield acceptable results for final coil temperature when the continuous duty temperatures calculated in example 28.1 would not. The methods discussed are applicable to standard dc relays and while the coil resistance formula is applicable to polarized dc relays (one that utilizes a permanent magnet) and ac relays as well, the pick-up voltage equations will not work in such cases. With a polarized dc relay the temperature induced change in magnetic force of the magnet must be considered. This is normally such that it reverses part of the change in pick-up voltage caused by the copper wire resistance. In the case of ac relays, the inductance contributes a significant portion of the coil impedance and is related to the turns in the coil. Since the inductance varies only slightly with temperature, the pick-up voltage exhibits less variation over temperature than for dc coil relays.

28.6 Relay voltage transient suppression

Voltage suppression is applicable to relay coils and relay contacts. The circuitry used is similar for each suppression case.

Although coil voltage suppression is used extensively, relays are normally designed and specified without taking into account the dynamic impact of suppressors. The optimum switching life (for normally-open contacts) is therefore obtained with a totally unsuppressed relay and rated electrical life factors are then based on this premise. Improper relay coil suppression has the typical symptom of random *tack welding* of the normally-open contacts when switching an inductive load or high inrush currents like with a lamp load. The successful *breaking* of a dc load requires that the relay contacts move to open with a reasonably high speed.
When an electromechanical relay is de-energized rapidly by a mechanical switch or semiconductor, the collapsing magnetic field produces a substantial voltage transient ($V = Nd\varphi/dt$) in an effort to release the coil stored energy ($W = ½LI^2$) and oppose the sudden change of current flow. A 12V/28V dc relay coil, for example, may generate a voltage of over 1kV during unsuppressed turn-off. This relatively large voltage transient can create EMI, semiconductor breakdown, and switch wear problems. It is thus common practice to suppress relay coil voltages with other components which limit the peak voltage to a controlled defined level. The measure of successful coil suppression depends on the degree to which the method affects the operation of the relay contacts. Improper or excessive suppression can cause the relay to suffer from a long release time, slow contact transfer, and contact bounce on break. All of these conditions will increase contact arcing when load switching, which reduces relay life.

28.6.1 Types of transient suppression utilized with dc relay coils

Coil de-activation
A typical relay will have an accelerating motion of its armature toward the un-energized rest position during drop-out. The velocity of the armature at the instant of contact opening will play a significant role in the relay's ability to avoid *tack welding* by providing adequate force to break any light welds made during the *make* of a high current resistive load (or one with a high in-rush current). It is the velocity of the armature that is most affected by coil suppression. If the suppressor provides a conducting path, thus allowing the stored energy in the relay's magnetic circuit to decay slowly, the armature motion is retarded and the armature may even temporarily reverse direction. Any direction reversal and re-closing of the contacts (particularly when combined with inductive loads) can lead to random, intermittent *tack welding* of the contacts such that the relay may free itself if operated again or even jarred slightly.
The basic techniques for suppression of transient voltages across relay coils are based on the suppression device in parallel with the relay coil or in parallel with the switch used to control the relay. It is normal to have the suppression parallel to the coil since it can be located closer to the source of the

problem, the relay (except in the case of PC board applications where either may be used). Better switch transient voltage protection is afforded when the suppression circuit is across the controlling switch, as shown in figure 28.5. Suppression used in parallel with the switching element is likely to be either a Zener diode or a series resistor-capacitor snubber. The Zener diode control is most advantageous since it does not significantly reduce relay endurance.

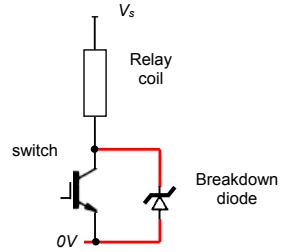


Figure 28.5. *Indirect coil voltage suppression.*

When the suppression is in parallel with the relay coil, any of the topologies in figure 28.6 may be used.

i. A reversed-biased rectifier diode.
ii. A bilateral transient suppressor diode that is similar in *V-I* characteristics to two Zener diodes series connected cathode to cathode (or anode to anode).
iii. A metal-oxide-varistor (MOV).
iv. A reverse-biased rectifier diode in series with a Zener diode such that their anodes (or cathodes) are common and the rectifier prevents coil-activated current flow. The Zener voltage is two or three times the level of the nominal voltage of the relay.
v. A reversed-biased rectifier diode in series with a resistor.
vi. A resistor, when loss conditions permit its use, is often the most economical suppression.
vii. A series resistor-capacitor snubber. Generally the least economical solution, figure 28.6e.
viii. A bifilar wound coil with the second winding and series diode used as the suppression circuit. This is not practical since it adds significant cost, losses, and size to the relay.

i. Diode clamped coil
A diode as in figure 28.6a clamps the cut-off spike to approximately 0.7V. However, the energy maintained by the continued current flow increases the release time. Some relays can reverse their armature movement direction when returning to the rest position due to a current flow increase. This can causes the make contact to close again under certain circumstances and can lead to an increased arc duration. This results in reduced endurance, hence the clamped diode configuration is not recommended for higher load currents.
The use of a rectifier diode alone to provide the transient suppression for relay coils may be cost effective and eliminates any transient voltage, but its impact on relay performance can be devastating, since the flux producing Ampere-turns decreases slowly, increasing the de-energized time. Problems of unexplained, random *tack welding* can occur. In some applications, this problem is only a minor nuisance or inconvenience and the relay can be cycled until the proper response is obtained. In some applications, the first occurrence of welding may cause a complete system failure or even present a hazardous situation.
Energize time of a relay can be increased and high voltage transients eliminated with an inductor/diode combination placed between the power supply and the relay, as seen in figure 28.6b.

ii. and iii. Silicon transient suppressor diode or MOV
Based on armature motion impact and optimizing for normally open contacts, the best suppression method is to use a silicon transient suppressor diode, as shown in figure 28.6c. This suppressor will have the least effect on relay dropout dynamics since the relay transient will reach and be maintained at a predetermined voltage level and permits coil current to flow into a low dynamic impedance. This results in the stored energy being quickly dissipated by the suppressor. Bi-directional transient suppressor diodes permit the relay to be non-polarized when installed internally. If a uni-directional transient suppressor is used, it must be used with a series rectifier diode to block normal current flow and it has little advantage over the use of a Zener diode, as in figure 28.6d. The transient suppressor should be selected such that its pulse energy rating exceeds any anticipated transient, such as coil turn-off or motor ‘noise’. MOVs produce virtual identical waveforms as silicon suppression diode, but MOV clamping properties can deteriorate with continuous electrical stressing.

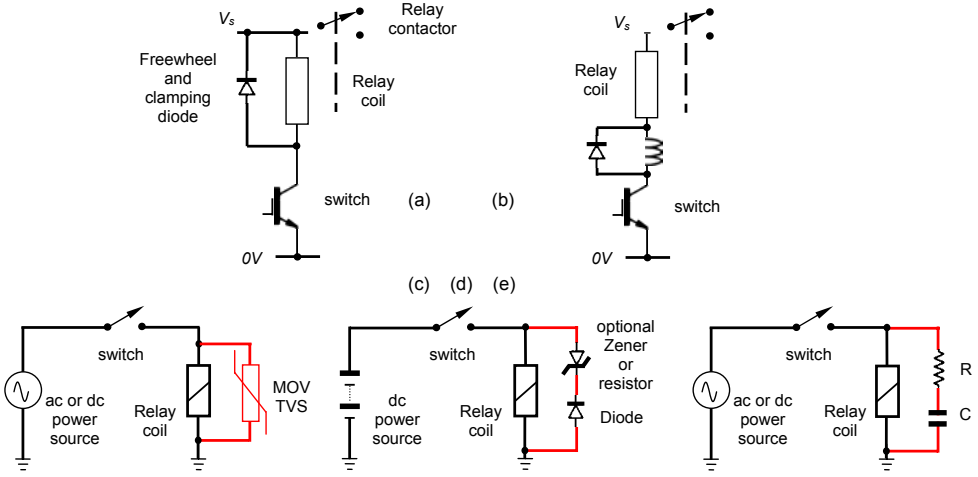


Figure 28.6. *Various relay coil voltage suppression techniques.*

iv. Zener + diode

If back EMF suppression of the relay coil is needed, use a Zener-Zener or diode-Zener series (figure 28.6d) combination with a Zener voltage at least twice the coil source voltage. The use of a reversed-biased rectifier diode in series with a Zener diode will provide the best solution when the relay can be polarized, since the Ampere-turns is reduced rapidly.

In relay carry-only applications, the release time may not be important and less expensive coil suppression techniques can be used. However, if the release/reset time is important, or if the contacts are to interrupt a load, then the Zener-Zener or diode-Zener combination may not be applicable.

v. Parallel resistance

The parallel resistance should be so rated that its value corresponds to approximately six times the coil resistance. In this way, the external cut-off spike is limited to three times the operating voltage. At a nominal voltage of 12V, the external spike can be contained to less than 36V. At a cost increase, coil activation losses can be reduced by adding a series diode, as indicated in figure 28.6d.

Table 28.2: The effects dc coil suppression on relay drop-out time

Suppression technique	Drop-out Time	Theoretical Transient
	ms	V
Unsuppressed	1.5	undefined
Diode + 24V Zener	1.9	-24.8
680Ω resistor	2.3	-167
82Ω resistor	6.1	-20.1
Diode	9.8	-0.8

These suppression techniques are based on normally-open contact performance, and must be qualified for normally-closed contacts. When the primary load is on, the normally-closed contacts (and a small load or none on the normally-open), it may be desirable to use a rectifier diode alone as relay suppression (or perhaps a rectifier diode and a lower value of series resistor). The retarded armature motion that adversely impacts normally-open contact performance will typically improve normally-closed contact performance. Improved performance results from less contact bounce during closure of the normally-closed contacts. This results from the lower impact velocity created by the retarded armature motion and can be utilized to improve normally-closed contact performance on certain relays. Table 28.2 show how increased coil reset voltage decreases relay drop-out time.

Pulse Width Modulation (PWM) and relay coils

A method to regulate the power consumption of a source supplying as relay is with a dc current driver, since the main electrical parameters of a relay (pull-in, pull-through and holding currents) are to a certain extent temperature independent. Since relay coils are usually voltage driven, those characteristics translate into the temperature dependent voltages for pull-in, pull-through and holding, due to the temperature dependence resistance of the coil copper wire.

Once the relay has pulled in, it maintains this status (armature maintains the contact position on the core) unless the coil current falls below the holding current, as shown in figure 28.7. For shock and vibration tolerance, there is an overhead current required, which depends on the relay type, relay parameters, temperature variation, and shock and vibration requirements.

PWM controlled drivers regulate the effective applied voltage by changing the fixed frequency duty ratio of dc voltage. Inductive systems like relay coils, in the presence of parallel connected freewheel components, respond to a negative going voltage edge with an exponential current decay.

Relay coil inductances are relatively high, with low resistance, which results in comparatively long time *L/R* constant hence small current ripple. But the R and L values are not constant. The relay coil inductance depends on the coil current (saturation) and status of the relay (armature open or closed). The copper coil resistance is highly temperature dependant.

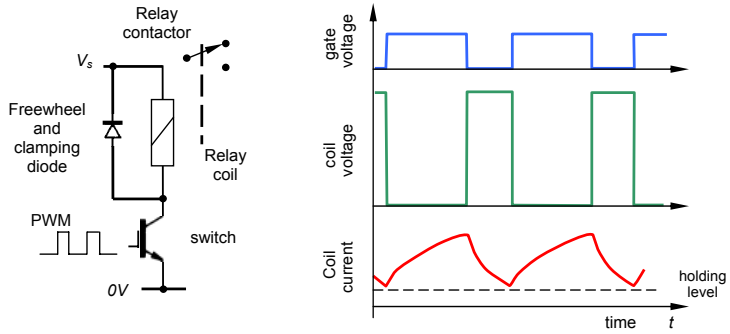


Figure 28.7. *Coil current response to a PWM voltage, with a parallel freewheel diode.*

Alternatively, using a series diode-Zener increases the ripple current, which must not fall below the minimum holding current level, as occurs in figure 28.7.

Coil activation

The definition of operate time is the interval between the application of the nominal coil voltage and closing of all normally open contacts (or opening of all normally closed contacts). This includes:

- Time for the coil to build up the magnetic field due to the increasing current.
- Transfer time of the moveable contact.
- Bounce time after the initial make or break.

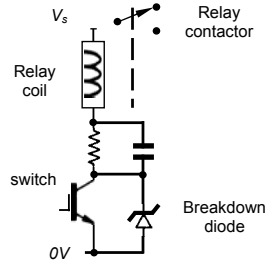


Figure 28.8. *Reduction of relay activation time*

Operate time is essentially a function of the coil power (specifically current, Ampere-turns) and inductance (*L/R* time constant). Standard circuit techniques can be used to alter relay timing characteristics. More than half of the switching time is taken to build up the coil field, thus the basic scheme for reducing operate-time is to apply more voltage across the coil. Faster relay operating speed can be accomplished by overdriving the coil with a higher than nominal voltage. For example, a 28V dc coil should not exceed 35V dc for continuous duty. To prevent overheating, the coil voltage should be reduced to the nominal value (or above the holding current level) shortly after the relay operates or a resistor equal to or greater in value than the coil resistance should be placed in series with the coil to keep total power applied at the specified level. Doubling the nominal voltage and adding an external resistor equal to the coil resistance can reduce the operate time by 40%.
A fast operating speed can be achieved by using an over-voltage pulse that decays to normal operating potential in a few milliseconds, as shown in figure 28.8, which achieves this with a simple parallel *RC* network placed between the power supply and the relay coil.

28.6.2 Relay contact arc suppression protection with dc power switching relays

Transient and arcing occurs during relay *make* (close), *break* (open) and contact bounce. Arcing can drastically lower the life expectancy of relays. Over 70% of relay failure occurs at the contacts. The most prevalent relay failure mechanisms are increased contact resistance, contact contamination, and material loss. Relay contact life expectancy is commonly a function of how much arcing can be withstood before failure occurs. Arc duration is often determined by the contact separation speed. Arcing, particularly during switching of inductive loads in high-voltage circuits, can be extremely destructive. To achieve maximum contact life, reliable arc suppression is important. It is difficult to prevent all arcing, but employing an arc suppression circuit will extend contact life.
Vacuum and hydrogen gas filled relays and contactors are often use in direct current electrical systems. Switching a direct current load is onerous on a relay. The relay or contactor needs to clear the maximum fault current, which is usually several times higher than the normal load. Unlike ac power, where both voltage and current regularly pass through zero allowing the arc formed during switching to naturally extinguish, a dc load can only be interrupted by forcing the arc voltage higher than the effective source voltage. A number of mechanisms are used to increase the arc voltage, ranging from arc chutes or multiple contacts, to magnetic blow-out, which lengthens the arc path.
Power switching relays are designed to interrupt rated power. However, reactive inductive loads can result in significant voltage overshoot, which can be suppressed by a variety of measures, usually more robust than the method used for relay-coil voltage suppression.
Figure 28.9 shows methods used to reduce the load on contacts by limiting the peak voltage transient developed across the relay contacts when interrupting inductive loads. The same circuits across the load will protect the load and contacts from voltage overshoot.

- Figure 28.9a shows a metal oxide varistor (MOV) across the power contacts. This circuit is suitable for most general-purpose ac and dc applications and MOV selection depends on transient energy, etc. MOVs are compact in size and low in cost. They protect from high breaking voltages with minimal additional drop out delay. They have limited switching frequency and are optimised for a specific voltage. Suppression diodes perform similarly. See chapter 10.X.
- Figure 28.9b shows an MOV in series with an SVP (Surge Voltage Protector - spark gap). The MOV absorbs transient overshoot energy, the SOV provides excellent dielectric isolation once the circuit is open. This solution is also compact in size and low cost. See chapter 10.X.
- Figure 28.9c illustrates the use of a traditional RC snubber, which will suffice for low power and energy ac and dc situations, but suffers a size and cost penalty at high power. RC elements are bidirectional, with minimal drop-out delay. A low overvoltage can be achieved, but is not suitable for low voltages. A disadvantage is a capacitor can lead to high making currents.

The main drawback of any relay contact protection circuit that restricts the arc voltage, is increased release time. Protection (rectifier) diodes have a breaking voltage peak of 0.7V, and no effect on making behaviour but do delay drop out by a factor of 3 or 4 times.

Vacuum and hydrogen gas-filled relays have some significant inherent advantages in switching high current dc loads. These include:

- Long load life due to ability to use high temperature contact materials
- Low contact resistance due to the elimination of contact oxidation/contamination
- Light weight and small size due to small contacts and short contact gaps
- Low coil power due to optimized magnetic circuits and small size
- High integrity, durable ceramic to metal hermetic seals

Hermetically sealed dc relays rated at 320V dc are available for lower dc voltage power applications, for example in electric vehicles and aircraft (see section 28.13).

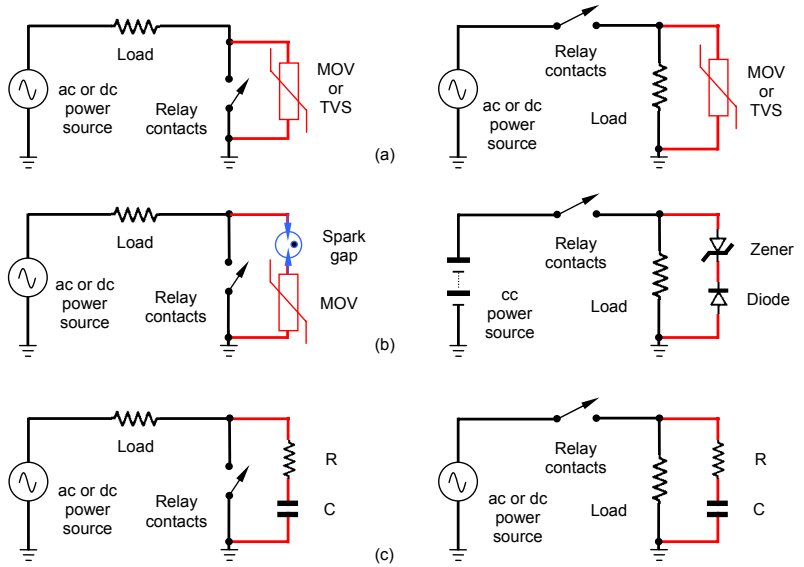


Figure 28.9. *Commonly used relay contact (direct and indirect) protection circuits.*

28.7 DC power switching

i. Low-voltage power relays
Relay loads are classified into four ranges. The load and voltage range definitions are:

- Dry circuit is defined by $0 \le V \le 30$mV and $0 \le I \le 10$mA. The softening voltage U of the contact material is not reached.
- Low load $U \le 20$V and $I \le 100$mA.
- Intermediate load $U \le 50$V and $I \le 1$A.
- High load $U \le 250$V and $I \le 1$A.

Dry circuit loads: No current is switched. The contacts carry current only after they are closed or before they are opened. The currents may be high, but are not switched. The maximum voltage applied is less than the softening voltage of the contact materials. Usually loads up to a maximum of 30mV/10mA are considered dry circuit loads. Since there is no arcing, contact resistance is kept low by using gold plating or gold alloy contacts. An increase in the contact resistance can only occur due to corrosion or polymerization. Switches with gas-tight housings can perform more than 200 million operations, switching dry circuit loads, without a change in contact resistance.

Low Loads: During contact break, the temperature in the constriction area increases to the melting and then boiling temperature of the contact material. Even though the open circuit voltage is lower than the minimum arc voltage, short arcs with low energy still occur due to inherent inductance and capacitance. Consequentially, carbon forms on the contacts (due to atmospheric hydrocarbons) but the arc does not have enough energy to remove the carbon. The sliding of the contact surfaces causes polymerization of the organic compounds with the result that deposits with high, unstable resistance are left on the contacts. This is a particular problem with metals of the platinum family. Typically, carbon or carbonized contacts have contact resistance between 2 and 4 Ohms. The solution is to use gold or gold alloy platings on the contacts.

Intermediate Loads: For intermediate loads, short arcs and discharges from cables are the most common effects. Current is below the minimum level for even momentary arcing when the contacts are open. Loads of 50 to 400mA at 26V are typical for this range. Some arcing can occur during the *make* or *break* of the contacts, but extinguishes itself by contact transfer completion. This arcing is

usually just enough to carbonize any organic vapours present. The carbonized material eventually deposits on the contacts and contact resistance increases, possibly leading to failure. The objective is to minimize the amount of organic vapours by the choice of insulation, potting compounds, and cleaning agents.

High Loads: Arcing Contacts. Arcing is detectable with all normal operational high loads. Typically, the voltage must be greater than 12V and the current must be more than 400mA for arcing to occur with most metals. The arcing actually serves a useful function, as it cleans the mating surfaces. Depending on the voltage and current, switched contact erosion or material transfer is the main effect. Contact materials such as silver-cadmium oxide, minimize electrical erosion at these load levels.

Mixed level switching: In different applications, monitoring the contact position in the load circuit is required. The synchronised failsafe coupling between two different contacts is realized in safety relays by applying forcibly guided contacts. For example, one contact is used to switch the load, and another is in the load circuit. The power in the load circuit depends on the application and variants between mW and kW (kVA). The control circuit is normally on the logic level of devices such as FPGAs or ASICs. A frequent issue is the separation of the contacts connected to power from those connected to the control circuit. The switching arc produces a splash of oxides and carbon particles which disperse in the area, for example, undesirably on to the contacts through which the control signal is carried.

ii. Higher-voltage power relays

Depending on the type of switching, higher-voltage power relays can be divided into two categories

Cold-switching of high voltages 12 to 70kV (no current during contact movement)
Hot-switching with voltages up to 25kV (current flow, hence arcing, during contact movement)

There are two types of relays for hot-switching

Make only relays, with pulse currents of up to a few kilo-amperes, with durations less than a few milliseconds
Power switching relays, for current up to 150A.

Misapplication of high voltage relays occurs most commonly in power switching applications. Since most relays are used only to isolate the load, it cannot be assumed that the rated carry current is the power switching rating.

High voltage applications (typically above 1,800V)

Make only applications - select a SF-6 gas filled relay whenever possible, since such relays are designed specifically for high voltage applications, but generally are not suitable for *breaking* the load.
Make and Break applications - select a vacuum relay that has contacts made of a material with a high melting point, such as tungsten or molybdenum. Some vacuum relays have copper contacts for high current carry applications, and are not suited for power switching.
In a vacuum relay, part of the contact material vaporizes during power switching and deposits itself on the inside walls of the relay. When this occurs, the dielectric stand off voltage decreases (leakage current increases) with the number of power switching cycles, possibly making it unsuitable for the application. The power switching rating of a vacuum relay is therefore dependent on the power to be switched, the number of cycles, the dielectric stand off voltage, and the maximum circuit leakage current allowed.

Medium voltage applications (typically < 1,800V)

Both vacuum and hydrogen filled relays are suitable for power switching applications at 1,800V dc and below. Vacuum relays typically have a longer life cycle rating than hydrogen filled relays, but do not carry or interrupt as much current. The most important parameter is the current the relay will be required to switch during 'abnormal' switching conditions. The relay may have to interrupt the entire current capacity of the system before a circuit breaker or fuse has time to function. An incorrectly selected relay could be vaporized during opening, resulting in electrical shorts within the system.
Whenever a relay is power switched, an arc is generated. The arc duration and the current and voltage levels are critical factors in determining relay life and reliability. Whenever a relay is required to switch a load, there are several precautions that should be taken to ensure a satisfactory result. These are:

1. Circuit load elements can generally be characterised as basically capacitive, inductive, or reactive, even though they may be comprised of both active (tubes and solid-state devices) and passive elements (capacitors, resistors, inductors, etc.). Circuits with capacitive or inductive elements are more difficult to switch due to the reactive stored energy. Switching these different types of loads has a specific effect on relay voltage.

Resistive loads

Circuits primarily resistive have minimal effect upon the voltage across hv contact terminals. In resistive loads, the duration of the arc is primarily determined by the speed at which the contacts separate, a shown in Figure 28.10a. The interruption of an ac load is easier on the contacts than a dc load since the ac interrupts itself each half cycle as the current goes through zero. Resistive loads are the standard against which other load types are measured, that is, relay load switch ratings usually assume a resistive load.

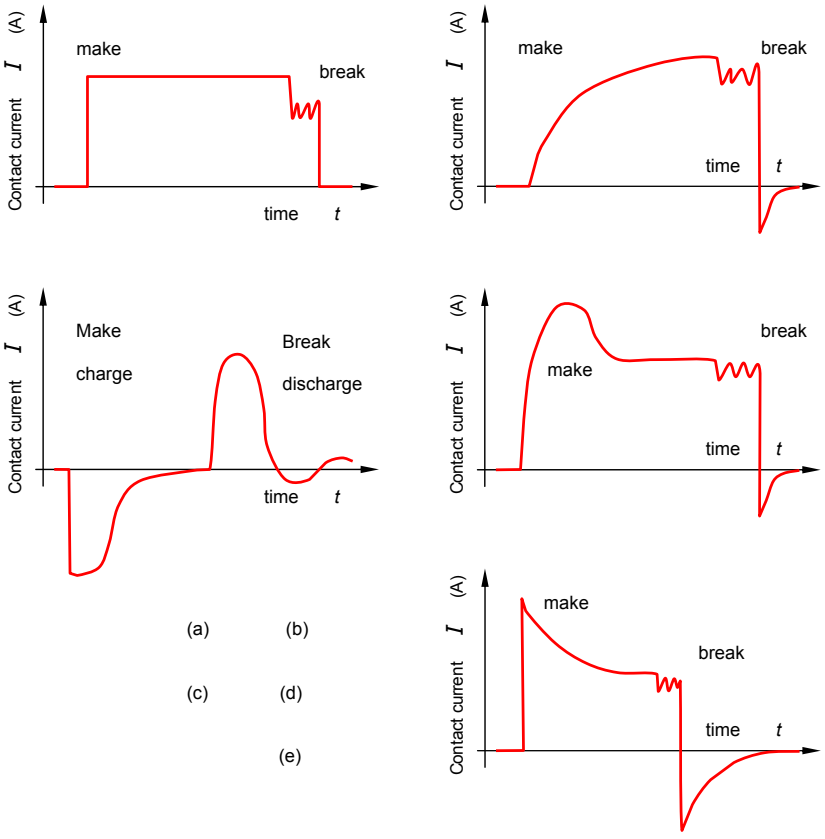


Figure 28.10. *Typical load profiles:*
(a) resistive load; (b) inductive load; (c) capacitive load; (d) lamp load; and (e) motor load.

Inductive loads

With inductive load elements, a high momentary voltage transient occurs when the circuit current is interrupted, which decays rapidly to the open line voltage. Inductive loads in high voltage circuits can be destructive. The release of stored energy when the load current is interrupted serves to maintain the current, as shown in Figure 28.10b, and cause voltage spikes that can damage associated circuit components, including the relay. Inductive loads in ac circuits are less stressful than in dc circuits. However, in both cases, the inductive load should be suppressed at its source with an appropriate protective device. If the inductive load is properly clamped, it becomes, in effect, a resistive load.

Capacitive, lamp, and other high in-rush loads
When circuits with large capacitive elements break, a negative bias voltage appears equal to the stored energy of the capacitor. This stored energy can cause a momentary high current surge upon contact make.
When switching on a lamp, charging, or discharging a capacitor, the inrush current may be many times the steady state current, as shown in Figure 28.10 parts c and d. The primary concern with high inrush loads is that contact bounce and associated arcing can cause the relay contacts to weld when making the load. For this reason, power-switching type relays specifically for lamp or capacitor charge or discharge applications should be used. Normally the maximum interrupt rating is used to determine the correct relay for lamp applications. SF-6 gas-filled relays are usually the best choice for capacitor discharge applications. A typical contact current profile for the high inrush associated with electric motors, is shown in figure 28.10e.

2. Are voltage spikes present in the circuit? Minimal inductance can generate extremely high voltage spikes that can damage circuit components. For this reason, steps should be taken to clamp any inductance at its source.

3. Select a ground isolated relay for high voltage load switching, whenever possible. Relays which do not have an internal ground plane are known as ground isolated and when such a relay is used for load switching, the potential for ground faults is virtually eliminated. If a ground-isolated relay is not available, locate the relay on the ground side of the load, as shown in figure 28.10. Then the load will limit the fault currents in the event of an internal arc-over to ground in the relay.

Power switching applications for high-voltage relays
High-voltage power switching applications are those that require the relay to make and/or break the load - hot loads. In most applications, it is important to know the highest potential fault current to be encountered and how many times the relay or contactor will be required to clear the fault, since these specify the required relay.
Load switching in ac circuits is less stressful on the relay due to the natural arc extinction that occurs as the current periodically passes through zero. Because of this, available relay ratings are much higher for switching ac circuit loads.
Switching of direct current loads creates specific problems for relays and requires relays, contactors, and power controllers that have been specially designed to handle the arcing problems of dc switching. When load switching occurs at voltages above 1000V, the typical power switching lifetime derating curves in Figure 28.11a are applicable.
Higher current affects relay ratings more than higher voltage, and the life expectancy for double throw, DT, relays is lower due to greater contact bounce.

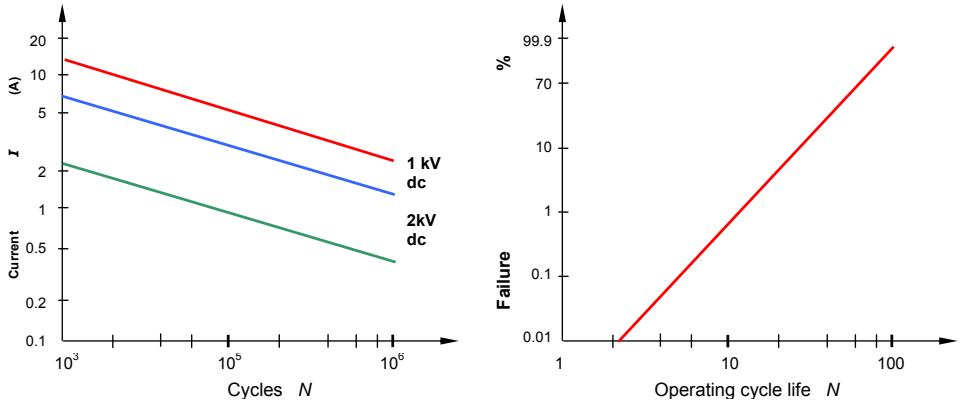


Figure 28.11. *(a) Typical contactor lifetime derating with increased load switched current and (b) reliability analysis Weibull data plot of end of life failures.*

Weibull plotting is used to predict product reliability. It is a simple and efficient way to predict reliability from a small number of life tests and it is widely used for this purpose. The cumulative percent failure is plotted against life. The Weibull scales ($\log_e$-$\log_e$ scales) are designed so the failure data of a wide

variety of manufactured devices will tend to lie in a straight line. As the chart in figure 28.11b shows, this data can then be interpreted to estimate the likelihood of failure at a given life.

28.8 The physics of vacuum high-voltage relays

In terms to contact loads, understanding can make relay selection less problematic, allowing the correct selection for a given application. Table 28.3 provides a comparison that indicates the best dielectric and contact materials for specific applications.

Table 28.3: High-voltage relay performance comparison with different dielectric and contact materials

Application	SF_6 Gas	SF_6 Gas	Vacuum	Vacuum
	Tungsten/Molybdenum	Copper (special applications)	Tungsten/Molybdenum	Copper
Carry Only (dc)	Good But the gas increases the contact resistance resulting is less current being carried than in vacuum	Better than hard contacts but the gas increases the contact resistance resulting is less current being carried than in vacuum	Good But not as much current as copper contacts	Best
Carry Only (RF)	No The gas will interfere with the RF carry capabilities	No The gas will interfere with the RF carry capabilities	Good But not as much current as copper contacts	Best
Make and Break	Good for make but only low currents on break	Good for make but only low currents on break	Best	Fair Extremely low currents only
Make Only	Best	Better But not as good as hard contacts	Good But not as much current as copper contacts	Fair Extremely low currents only
Long periods of on-use or where Low and stable leakage current is needed	Best Only relays termed 'Make only'	Better than hard contacts but the gas increases the contact resistance resulting is less current being carried than in vacuum	Good Generally will 'burp' when HV is applied	Good Generally will 'burp' when HV is applied

A vacuum is an ideal dielectric. The dielectric strength of a vacuum is about 8 times greater than that of air. Since there is no oxidation in a vacuum, low resistance copper contacts (rhodium for reed relays) are used that allow the relay to carry significantly more current than traditional air exposed relays. Vacuum relays with copper contacts are termed *carry only* relays.
Since most high-voltage arcs are initiated by the ionization of the insulating medium, a hard vacuum, which, by definition, is the absence of any such media, produces the greatest possible isolation between contact electrodes. It is possible to obtain dielectric strengths of 100kV per mm contact gap in a vacuum relay. A vacuum dielectric has the additional advantage of providing an inert environment in which high-voltage contacts can operate completely oxide-free. Thus, vacuum relays typically have a contact resistance, which is lower and more stable than other relay types.
Voltage breakdown can occur even within an absolute vacuum when the contact material itself becomes the source of ionized material to support an arc. Because (in make and break modes) soft contact materials like copper and rhodium vaporize easily as the contacts switch and deposit on the inner walls of a vacuum relay, a 'plating out' of the walls occurs over time, resulting in dielectric breakdown.
Therefore, high strength and high work function materials like refractory metals (tungsten, molybdenum, etc.) are commonly used for contacts in order to raise the electrostatic field strength necessary to cause voltage breakdown. In addition, refractory metals have high melting temperatures, which reduce contact damage from arcs and result in longer life.

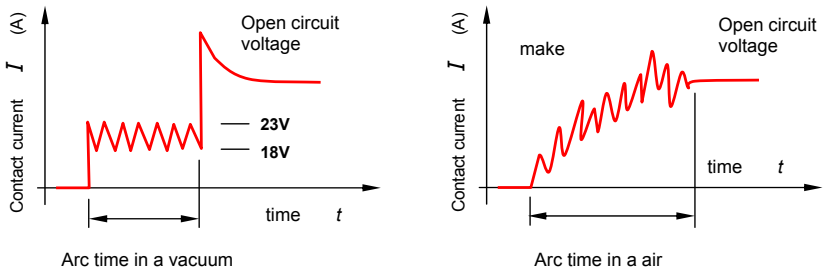


Figure 28.12. *Comparison of Arcs in air and a vacuum.*

In load switching, an arc will always be created at the point when the contacts are close enough to allow voltage breakdown. Vacuum arcs are sustained at the relatively low voltage of 18 to 23V compared to arcs in air which are more erratic and range over a wider voltage, as seen in Figure 28.12. Vacuum arcs tend to be more easily controlled and extinguished than arcs in air. The high-pressure region formed around a vacuum arc has a strong tendency to dissipate or blow out into the surrounding low-pressure vacuum. This phenomena, along with the ability to use contact materials like pure tungsten and molybdenum which have high melting points, means that vacuum relays and contactors typically experience much less contact erosion and have a longer life than comparable air-break devices. The resultant vacuum relays are termed *make and break*.

28.9 Gas filled relays

28.9.1 SF_6 as a dielectric

Some relays contain a proprietary gas mixture consisting of, mainly, sulphur hexafluoride at several atmospheres of pressure. Pressurized sulphur hexafluoride has good insulating qualities and comes close to achieving the same standoff voltage as a vacuum. These gas-filled relays are usually recommended when an application involves changing or discharging a capacitor especially when the voltage is greater than 1kV. SF_6 under pressure has many advantages over a vacuum because the leakage current is stable over long periods of non-operation and because of the way the gas performs during switching. SF_6 is an excellent insulator but when the relay is switched, ionization of the gas causes electrical continuity to occur before mechanical continuity is achieved. If the relay bounces the SF_6 becomes easily ionized and carries the arc current. This makes the relay electronically bounceless and dramatically reduces contact wear, which contributes to the long life that these relays exhibit in capacitive make and break applications, and reduces electrical noise during switching. These SF_6 gas filled relays with hard contact materials are termed *make only* relays. However, this tendency to ionize makes gas-filled relays unsuitable for applications which require interruption of a load.

There are two significant shortcomings to SF_6 gas-filled relays:

- Due to a film that forms on the relay contacts, SF_6 gas-filled relays have a higher and less stable contact resistance at low voltages than vacuum relays. Contact resistance is typically between ½Ω and 1½Ω when measured at 28V dc on new relays. However, even in applications up to 100V, when the current is low, higher contact resistance may occur. When low contact resistance is important, and the voltage and/or current is low, a vacuum or other type of gas filled relay should be used instead of a SF_6 gas-filled relay.
- Due to the ease with which sulphur hexafluoride ionizes, gas-filled relays cannot normally be used to interrupt loads.

SF_6 gas filled relays do not emit hazardous X-rays because the electrons collide with the gas molecules and are unable to accumulate sufficient energy to create significant radiation.

SF_6 gas filled relays are recommended for many non-RF high voltage applications. For non-RF applications and for relays over 10kV, SF_6 gas filled relays are the most forgiving of all the high voltage relays. Because they have SF_6 gas inside rather than vacuum, the leakage current is generally lower and more repeatable over long periods of non-operation. Because of the gas, they are most tolerant should the contacts have to *make* an abnormal load. If the load is not RF, first consider relays rated for *make only*.

Sealed high-voltage relay applications include high in-rush capacitive make and capacitive discharge such as - found in ESD test equipment, cable test equipment, heart defibrillators, and for applications where no high voltage is applied for long periods where low and or stable leakage current is needed. The sealing processes for vacuum relays can be used to back-filled and pressurized with SF_6 gas, *make and break* relays with hard contacts. Where the inrush is not too high and where higher carry current is required, back-fills with pressurized SF_6 gas of *carry only* relays is used.

28.9.2 Hydrogen as a dielectric

Some relays use a proprietary gas mixture consisting of, primarily hydrogen at various atmospheres of pressure. These mixtures do not have the same high dielectric and low leakage current as a vacuum or SF_6, so are not normally used for high voltage applications above 3 kV, but are ideal for dc *make and break* load switching applications. The gas mixture, combined with the use of external magnets to control the direction of the arc, cools and extinguishes the arc in a predictable manner. Because there is no oxygen in the mixture, more conductive contact materials such as copper can be used that provide the lowest possible contact resistance. The gas mixtures and magnets provide high current interrupt capabilities up to 3,500A at 320Vdc with switching capabilities as high as 1800V dc. They also offer the ability to handle highly inductive dc load switching. Hydrogen and nitrogen gas relays are compared in Table 28.4.

Table 28.4: H_2 and N_2 gas and contact material performance comparisons for switching applications

Application	H_2 (Hydrogen)		N_2 (Nitrogen)	
Contacts	Copper	Molybdenum/Copper	Copper	Molybdenum/Copper
Carry only	Best	Good pure copper is better	Better	Fair pure copper is better
Make only life	Good Molybdenum/Copper is better	Best	Fair Molybdenum/Copper is better	Better
Make and break life	Good Molybdenum/Copper is better	Best	Fair Molybdenum/Copper is better	Better
high overload make and break	Good Molybdenum/Copper is better	Best	Fair Molybdenum/Copper is better	Better

28.10 High voltage relay designs

Typical high voltage relay designs for high-voltage switching are sealed, providing rugged, small, and efficient high-voltage designs for demanding switching applications.

i. Internal hinged Armature Style

This traditional design approach provides high mechanical reliability and is adaptable to a number of contact configurations. The contact is actuated by the movement of the spring-loaded armature when the coil is energized. The coil assembly is external to the vacuum package and readily replaceable. When power is applied to the external coil of these relays, a magnetic filed is transferred through a pole that runs through the centre of the coil to the armature, that is located inside the vacuum or gas filled sealed ceramic envelope switching chamber. The armature moves the common contact to the normally open contacts. A spring inside the sealed chamber returns the moving contact to the normally closed contact when coil voltage is removed.

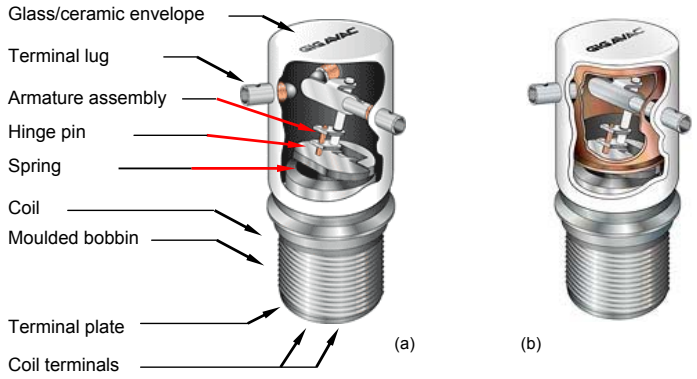


Figure 28.13. *Internal hinged armature style relay:*
(a) double throw relay design but (b) with built-in internal copper metal shield for power switching.

Figure 28.13a is a typical design used in many high-current, high-voltage relays. This is a single pole double throw relay, SPDT. Depending on the switching application, various contact materials are used inside the sealed chamber. Tungsten/molybdenum is used for *making* or *breaking* loads. Copper contacts have lower contact resistance and are used for higher current *carry only* applications such as for RF.
Figure 28.13b shows the same design but with a built-in internal shield that extends relay life. When power switching a load using a vacuum relay, even hard contacts vaporize, and the material becomes deposited and plate-out the internal walls of the ceramic envelope. Over time, these deposits reduce the isolation voltage, which causes the relay's end of useful life. This plate-out condition is solved by incorporating an internal metal shield as shown. The deposits hit the shield rather than the ceramic wall, resulting in a relay life many times longer than relays without the shield.

ii. Diaphragm Style

This simple, low-cost design approach makes use of a thin molybdenum diaphragm that allows contact movement to be transferred into the vacuum package enclosure from the external actuating assembly.
Figure 28.14 parts a and b show diaphragm style relays. The contacts are sealed in a chamber at the top of the relay. The chamber is sealed with a braze joint at the top, and with a thin molybdenum diaphragm below. The external high voltage connections are integral to the braze seal. The relay armature is below the sealed chamber as shown on the cross sectional view in figure 28.14c. When power is applied to the non-polarised coil, the armature moves, and a ceramic insulating rod that is attached to the diaphragm moves the common contact to the normally open contact (a small rod) inside the sealed chamber. Figure 28.14a is a single throw, normally open configuration, where the top contact, A3, is open and the moving contact, A2, is below.
Figure 28.14b is a double throw relay. The normally open contact is at the top, the normally closed contact is in the centre, and the moving contact is at the bottom. For this relay, the sealed chamber extends from the top of the relay down to the diaphragm that is the moving contact. Both the normally open and normally closed contacts are in a common sealed chamber.

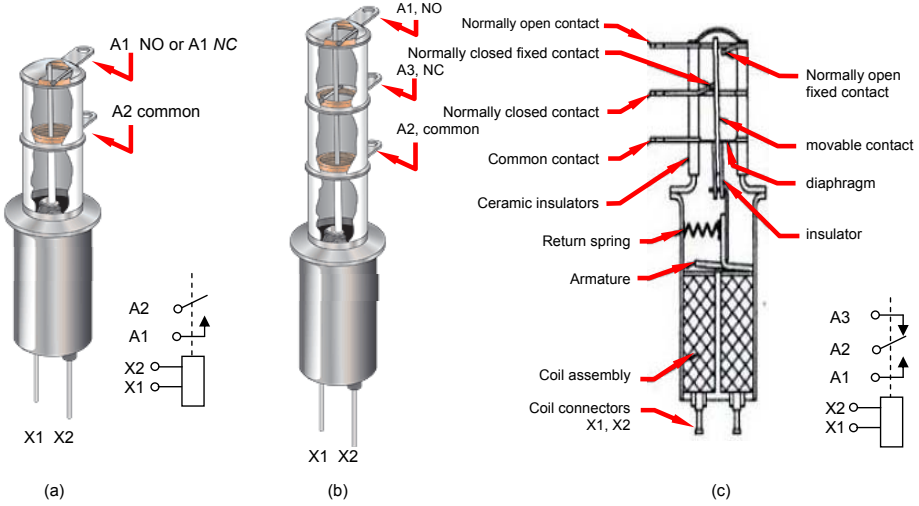


Figure 28.14. *Diaphragm style single-pole relay design: (a) single throw; (b) double throw relay design; and (c) cross sectional view.*

iii. Package encapsulation design

Figure 28.15 is a diaphragm style relay, where the relay in figure 28.13 or 28.14 is packaged inside a cup that provides more mounting and high voltage terminal options. Because the contacts are in a vacuum, they can withstand higher voltages than the distance between the external terminals. By potting the relay inside the cup, the high voltage capabilities are greatly improved.

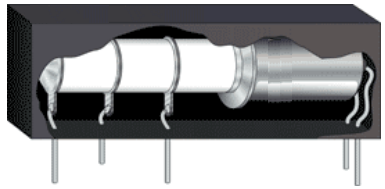


Figure 28.15. *Encapsulated high-voltage single-pole double-throw diaphragm relay.*

iv. Sealed Relay and Contactor Designs

Impervious ceramic sealing technology is a low-cost, light, rugged, high temperature, high quality hermetic seal for relays and contactors. The ceramic header is welded, without a thick layer of epoxy, to a can with high-temperature plastics inside for the arc management. The ceramic seal allows high pressure sealing for a vacuum, or hydrogen or suitable inert gasses that facilitate high-current interruption. In turn, the controlled internal environment allows the use of low-resistance, high-conductivity, high contact pressure, copper contacts. Formerly, sealed relays and contactors used all ceramic envelopes which are expensive, glass to metal seals which do not provide true seals over long periods of time, or epoxy and plastic seals that do not provide the high temperature ratings or the micro-sealing needed to exploit higher performance back fill gasses.

The aerospace industry has adopted 270V dc high-voltage systems and many industrial applications such as solar power, fuel cells, micro-turbines, hybrid and electric vehicles, etc. have also involve dc systems.

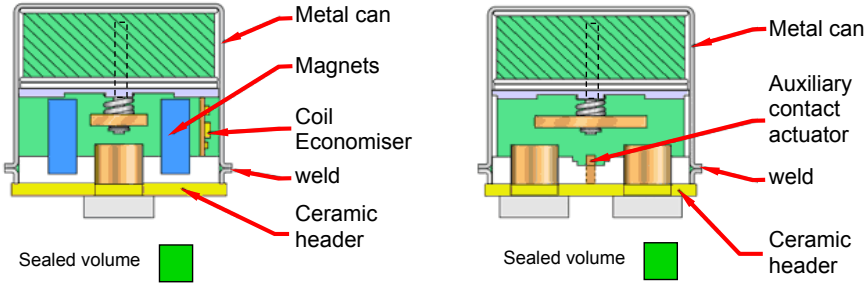


Figure 28.16. *Sealed ceramic encapsulated dc relay.*

vi. High-Voltage Reed Relays

The principal function of a high voltage reed relay is to isolate a high voltage, with the use of evacuated reed switches. These are available with tungsten or rhodium contacts, depending on the switching requirements of the application. High-voltage reed relays are intended for use in dc or ac (50Hz to 60Hz) applications. (RF reed relays are not specifically considered).
The reed switch consists of two ferromagnetic blades (generally composed of iron and nickel) hermetically sealed in a glass capsule, as shown in figure 28.17. The blades overlap inside the glass capsule with a gap separating them, and make contact with each other when a suitable external magnetic field is applied. The contact area on both blades is plated, welded or sputtered with a hard metal, usually tungsten for low frequency switching or rhodium or ruthenium for lower contact resistance. These hard metals potentially offer long life times if the contacts are not switched with heavy loads. The gas in the capsule is usually nitrogen or some equivalent inert gas.
Some reed switches, to increase their ability to switch and standoff high voltages, exploit an internal vacuum. The reed blades act as magnetic flux conductors when exposed to an external magnetic field from either a permanent magnet or an electromagnetic coil. Poles of opposite polarity are created and the contacts close when the magnetic force exceeds the retarding spring force of the reed blades. As the external magnetic field is reduced, the force between the reeds falls below the restoring force of the reed blades, so the contacts open.
The common blade (or armature blade), the only moving reed blade, is connected to the normally closed blade in the absence of a magnetic field. When a magnetic field of sufficient strength is present, the common blade is attracted to the normally open blade. The normally open and normally closed blades

always remain stationary. All three reed blades are ferromagnetic; however, the contact area of the normally closed contact is a non-magnetic metal which has been welded to the ferromagnetic blade. When exposed to a magnetic field, both the fixed reeds assume the same polarity, which is opposite that of the armature. The non-magnetic metal interrupts the magnetic flux on the normally closed blade so that the armature experiences an un-interrupted flux path to the normally open blade, and to which it is attracted. If the attractive force is of sufficient magnitude between the normally open and armature, the contacts close.
Using the proper design, materials, placing an electrostatic shield around the reed switch internal to the coil and driving the shield, will allow coupling or transmission of small signals (nV signals or fA currents) through the relay with little or no interference. This is virtually impossible with other technologies except at high cost. Thermoelectric voltage cancellation, with two series differentially connected contacts, is necessary at low signal voltage levels.
Contact Materials – Rhodium versus Tungsten
Rhodium offers superior low contact resistance, which, coupled with copper plated reed switch technology produces low loss RF reed relays, with exceptional current carry performance. Rhodium contacts are offered for high voltage applications, where low contact resistance and good current carry performance are required, provided the switching voltage is below 1000V dc or ac peak.
Tungsten contacts are used exclusively for high voltage 15kV isolation (stand-off) reed relays, which are high voltage switching contacts able to switch voltages up to 12.5kV dc or ac peak at low current, 5A, 200W. Operate and release times are both about 2ms. Tungsten is a good general purpose switching contact material (10^9 dry switching and 10^8, 200W switching operating lifetime, which reduces with breaking current) but a higher contact resistance, 100mΩ, means it is not suited for RF applications. Low capacitance of less than ½pF between contacts, is typical.
In a reed relay, the reed switch uses an electromagnetic coil for activation and is shown in its simplest form in figure 28.17. Reed relays require little power to operate and are generally gated using transistors, TTL directly or cmos drivers. Reed relay contacts, when switched dry, (zero current closure or less than 5V @ 10mA), function well into the billions of operations.

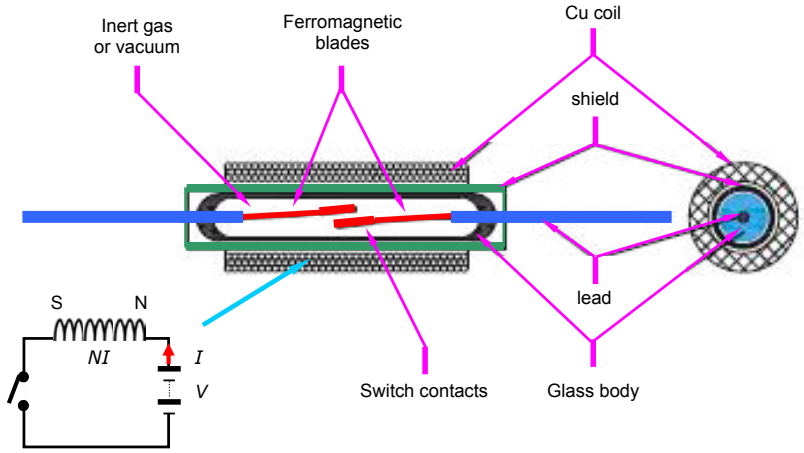


Figure 28.17. *A reed relay consisting of a wound coil with a reed switch.*

28.11 Contact ratings

There are two elements in establishing contact ratings for a contactor.
1 - *Contact carry-current rating* – This rating is a matter of heat dissipation. The contact carry current rating is based on four parameters:

- ambient temperature,
- heat generated because of the contact resistance,
- internal resistance of the contactor current carrying mechanisms, and
- size of the cables connected to the terminals.

By increasing the wiring diameter, removing more heat from the terminals, or if the ambient temperature is reduced, the more current that can be carried. The maximum allowable terminal temperature is important, otherwise there is no indication of how hot the terminals and connecting wire may get. It is

needed to select compatible wire insulation, and to determine how much heat the contactor may dissipate in a sealed enclosure. Worst case, an applied current could exceed the contactor terminal rating, causing melt down of the contactor, leading to an electrical short or possibly fire.

2- *Power switching rating* – This rating is determined by contact wear and the resultant loss of dielectric withstanding voltage or insulation resistance, caused by the depositing of vaporized contact material near the contacts.
The ability of a vacuum relay to switch both resistive and inductive loads greatly simplifies circuit design problems. In power switching applications non-isolated relays (which includes all relays not identified as ground isolated) must be used with caution when the relay mounting is at ground potential and the circuit to be switched is at a high potential. Fault conditions may cause internal arc over to the grounded housing. Ground isolated relays can be used within their voltage ratings without concern for ground faults because the switching part of the relay is completely isolated from ground.
Vacuum relays are made in models designed to be switched *hot* or *cold*.

Hot switching often entails contact arcing upon opening and during contact bounce. AC and dc circuits have extra considerations when switched 'hot'.
Cold switching is where the circuit is switched with no load through the relay terminals, so the relay acts either as an insulator or a conductor.

In the *make* mode the contacts conduct the full load current, and contact current handling capacity is limited by heating caused by contact resistance. Special low resistance copper alloys are used for most cold switching relays to assure high current handling capabilities.
In the *break* mode, the relay must perform as a high voltage insulator. Stand-off voltages are highest at dc and low ac frequencies, and reduce at higher frequencies due to RF heating of the insulator. Ceramic insulators provide the best withstand capabilities for high RF applications.
Many dc applications involve controllers or inverters that have large dc capacitors across the dc link. Unless these capacitors are pre-charged, the contactor essentially experiences undefined current when the contacts first close, limited only by the internal resistance and stray inductance of the system. If there is no pre-charge of the capacitor, the contacts can weld on the first or second cycle. The difference in expected life rating with a pre-charge of 90% and 80%, is a factor of 100. If possible or appropriate, an ac contactor should be used on the three-phase ac input to any rectifier stage, thereby avoiding the need for less robust dc relay technology.

28.12 High voltage relay grounding

It is normal practice to ground the base of all high voltage relays for electrical safety.

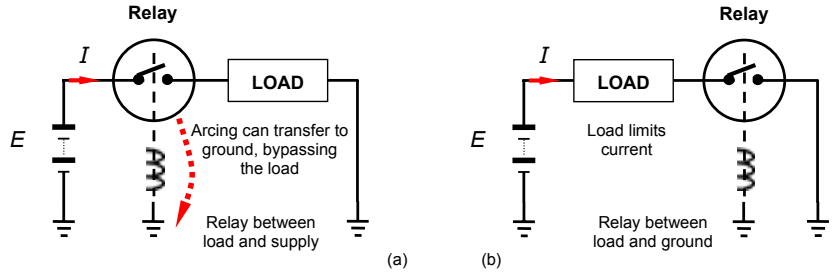


Figure 28.18. *Contact position relative the load and dc source: (a) relay arcing shorting source to ground and (b) any relay arc draws controlled current through the load from the source.*

The topological position of a relay in a series circuit can determine its maximum capabilities. For example, a relay with an internal ground plane within the vacuum envelope will break much more power when one contact is at ground than when the contacts are between the dc power supply and load, as shown in the two parts of figure 28.18. When a *hot* circuit is switched, an arc is usually created. This arc can transfer to ground when a relay with internal grounding is placed between the load and the power supply as in figure 28.19b. This ground fault or breakdown results from ionized gas and vaporized metal from the contacts that bypass the load and conducts between the high voltage lead and the ground plane of the actuator. The only limit to the resultant current surge is the inherent current limitation of the dc power supply itself. When one contact is at ground potential, however, the series load limits the surge current, as shown in figure 28.19b. To eliminate this type of breakdown problem, relays with ground

isolation from the vacuum enclosures should be used. Vacuum relays operated *cold* may be installed in any circuit location, as the relay does not interrupt power, but acts either as a low loss conductor or as a high voltage insulator.
The mentioned grounding requirements are specified by the particular relay contact style.

i. *Diaphragm style relays* need not have their base grounded. This is because there is no ground plane inside the sealed switching chamber that an arc can strike during hot switching, and because the external distance to ground, combined with the added insulation of the coil, is greater than the breakdown voltage between contacts. These relays can be used in hot switching applications on either the high-side or low-side of the load, as shown in figure 28.19.

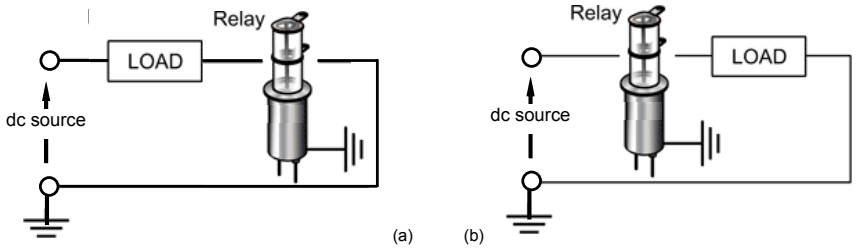


Figure 28.19. *Diaphragm style relays where grounding of the case is not necessary but is recommended for safety while the contacts can be on either side of load: (a) relay on the ground-side of the load and (b) relay on the dc source-side of the ground connected load.*

ii. *Internal armature style relays* must always have their relay base grounded, as shown in figure 28.20a, unless the voltage across the contacts is less than the specified dielectric voltage breakdown between the coil and case. When hot switching voltages above the coil to case dielectric voltage rating, the relay *must* be on the ground-side of the load as seen in figure 28.20a and the case *must* be grounded. For hot switching voltages lower than the coil to case dielectric voltage rating, the relay can be on either side of the load, that is, as in figure 28.20a or figure 28.20b, and the case does not have to be grounded but is recommended to be grounded for safety.

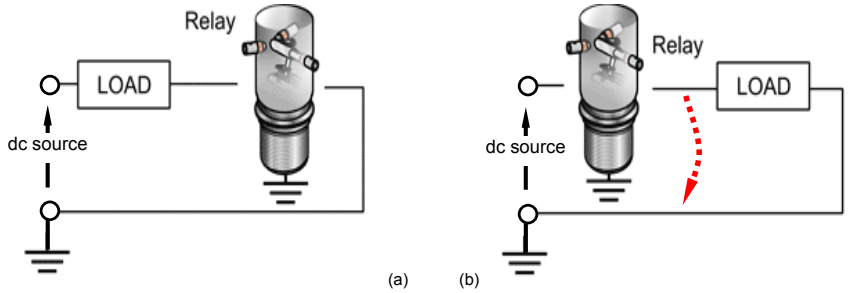


Figure 28.20. *Internal armature style relays should always have their case base grounded, plus: (a) should always be on ground-side of load if the source voltage is greater than the coil to case dielectric rating and (b) on the non-ground-side of load if the source voltage is lower than the coil to case dielectric rating.*

In figure 28.20a, if arcing transfers to ground, the load limits the current. In figure 28.20b (and also figure 28.19), if arcing transfers to ground, current bypasses the relay and load, and is only controlled by the source impedance. Power control applications often must utilize figure 28.20b, although figure 28.20a is recommended for high-voltage circuit applications.

28.13 A LV voltage, 750V dc, high-current, 350A dc, make and break relay

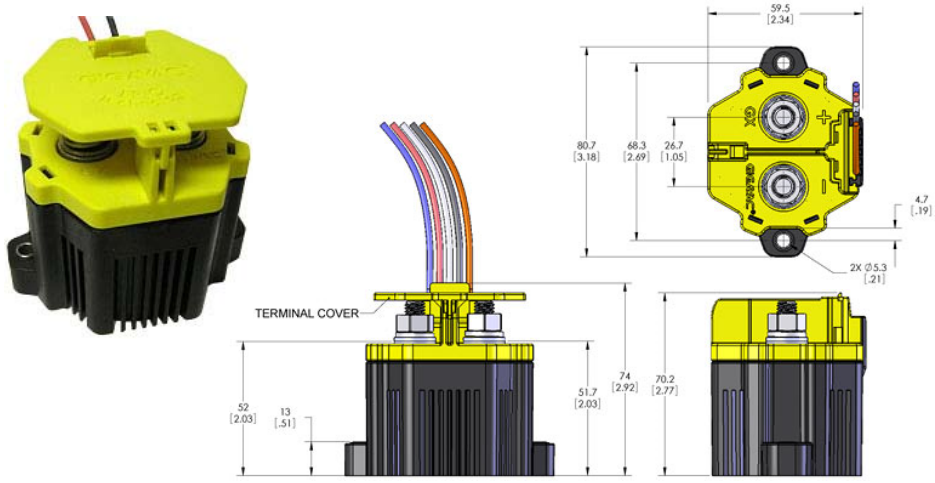


Figure 28.21. *Typical 750V dc 350A dc relay.*

Table 28.5: 750V dc, 350A make and break relay characteristics

Specifications	unit	value
Contact Arrangement (main)	Form X	SPST-NO
Contact Arrangement (auxiliary)	Form C	SPDT
Mechanical Life	cycles	1 million
Contact Resistance Maximum @ rated carry current Typical @ rated carry current	 mΩ mΩ	 0.4 0.15 to 0.3
Operate time, 25˚C Close (includes bounce) Maximum Close (includes bounce) Typical Bounce on close, Maximum Release time (includes arc time at maximum break current)	 ms ms ms ms	 20 13 7 12
Insulation Resistance @ 500V dc	MΩ	100
Dielectric withstand voltage at sea level (leakage < 1mA)	V rms	2,500
Shock (peak) 11ms ½ sine	G	20
Vibration, Sinusoidal (80 to 2000 Hz, peak)	G	15
Storage Ambient Temperature Range	˚C	-70 to +175
Mass	kg	0.44

Figure 28.21 pictures a 750V dc 30A make and break relay based on the relay construction, with the features as outlined in section 28.10iv. General properties are outline in Table 28.5, while make and break performance effect on lifetime is shown in figure 28.22. The maximum make current is 650Adc, at which level contact welding may occur. End of life is when the insulation resistance between terminals haves, that is, falls below 50MΩ @ 500V dc. Electrical life rating is based on a resistive load with 27μH maximum inductance in the circuit.

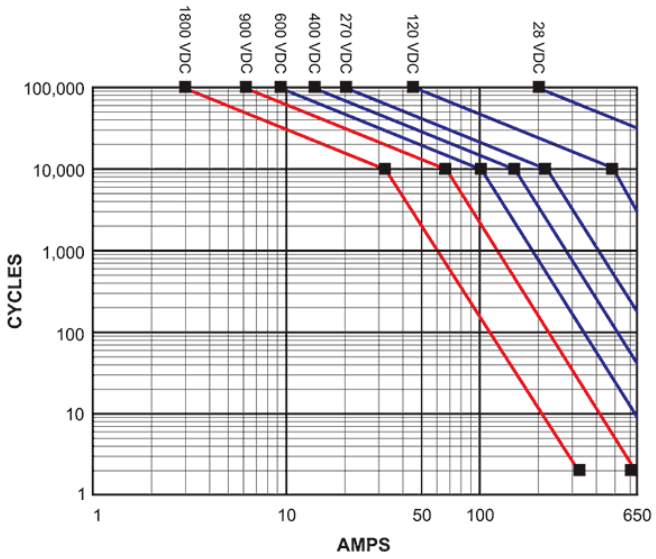


Figure 28.22. *HV dc relay type upper V-I resistive load make and break limits and life cycling ratings.*

28.14 X-ray emissions in vacuum relays

Above 15kV, all components that operate in a vacuum, including vacuum relays, can produce X-rays that are hazardous. When extremely high voltages are applied, charge carriers in the electrical field are accelerated and can cause radiation when they impact on the electrodes. This is one reason for using SF6 gas filled relays because the electrons collide with the gas molecules and are unable to accumulate sufficient energy to create significant radiation. Gas-filled high voltage relays can be operated safely at high-voltages without any concern for X-rays.
Although many relays rated for use above 15 kV are gas-filled relays, when vacuum relays are used over 15kV, the equipment should be shielded with lead Pb that is at least 1.6mm thick. If shielding is not possible, then appropriate X-ray warnings should be labelled and a radiation X-ray monitoring programme should be implemented.

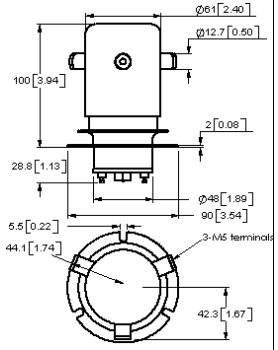
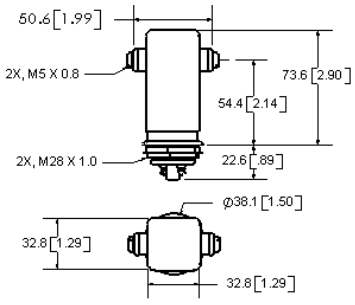
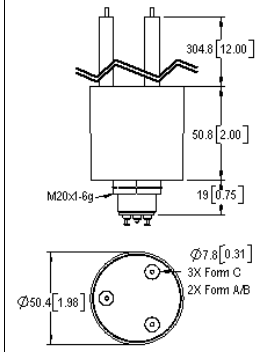
28.15 Power reconstitution conservation method

Vacuum relays may show signs of 'gassiness' after a relatively short period of non-use. A trace of gas released from an adsorbed state on the relay internal surface is usually responsible. This trace can normally be eliminated by the use of a high voltage processing procedure.
High-voltage process procedure is as follows:

- Connect a variable high voltage ac or dc power supply in series with a 10MΩ resister, a micro-ammeter, and the relay normally open contacts (with the relay on the ground-side of the supply).
- Immerse the relay in a dielectric fluid, such as transformer oil, but Fluorinert FC-77 is cleaner since it evaporates quickly from the relay surface.
- Raise the source voltage slowly. If the peak voltage is made equal to the maximum specified test voltage and less than 5μA of current is drawn (or no glow is visible in a darkened room), then the vacuum is *hard* and no reconstituting processing is necessary.
- If a glow occurs at a lower than maximum specified test voltage, hold the voltage just above the glow initiation level until the glow disappears; raise the voltage again to the onset of glow, or until the maximum specified test voltage is reached. If a dc supply is used, reverse the polarity and repeat the process.

Processing is up to 20% above the maximum specified test voltage. Typical processing times range from one minute to several minutes for high-voltage relays. It is not necessary to high-voltage process gas-filled relays.

Table 28.6: Range of available high-voltage dc relays

	25kV @ 150A	28kV @ 110A	70kV @ 10A
	Make and break, hot	Carry only (no load switching), RF	Make only, capacitor charging and discharging, with high inrush
			
			
	Ceramic, vacuum dielectric, tungsten	Ceramic, vacuum dielectric, Cu	gas filled, SF6
	SPDT	SPST NO / NC	SPST NO / NC, STDT, Latching
$I_{c\text{-}b}$	15uA	15uA	15uA
DC or 60 Hz	25kV,150A	28kV,110A / 25kV,55A	dc 70kV,10A
2.5 MHz	15kV,120A	22kV,60A / 22kV,30A	60Hz 30kV,10A
Coil hi-pot	500V rms @ 60Hz	500V rms @ 60Hz	500V rms @ 60Hz
$C_{c\text{-}c}$	5pF	2.5pF	-
$C_{c\text{-}gnd}$	5pF	2.5pF	-
$R_{c\text{-}c}$	3mΩ	5mΩ / 10mΩ	2mΩ
t_{op}	100ms	18ms / 18ms	20ms
t_{rel}	15ms	10ms / 20ms	15ms
life	1,000,000 cycles	2,000,000 cycles	500,000 cycles
mass	1kg	0.342kg	0.336kg
T_{op}	-55°C to +125°C	-55°C to +125°C	-50°C to +85°C
	G52 (H-25:- 50kV,10A, 25mΩ 60ms / 60ms, +85°C)	KC20 / KC30	K70A/B/C G71L

28.16 MV AC vacuum interrupts for contactor, switch, and circuit-breaker application

Vacuum interrupters with a contact gap of up to 7mm exhibit higher dielectric strength compared to SF_6. With wider contact gaps have a higher dielectric strength when SF_6 is used as the medium.
At 16mm (corresponding to the distance between contacts in a 36kV vacuum circuit breaker), the measured dielectric strength is approximately 200kV, as shown in figure 28.23. This is slightly higher than the rated lightning impulse withstand voltage of a standard 38kV breaker. For higher system voltages, therefore, two or more vacuum interrupters would have to be connected in series, which is marginally economical. Therefore SF_6 circuit breakers are typically used at voltages higher than 72.5kV.

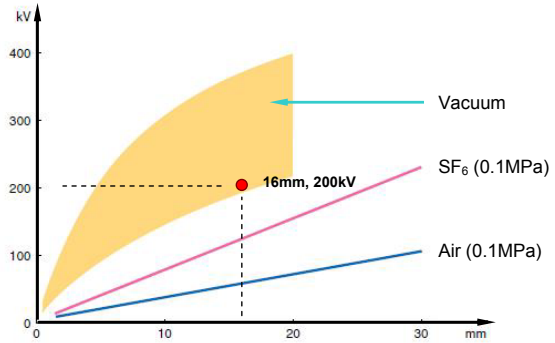
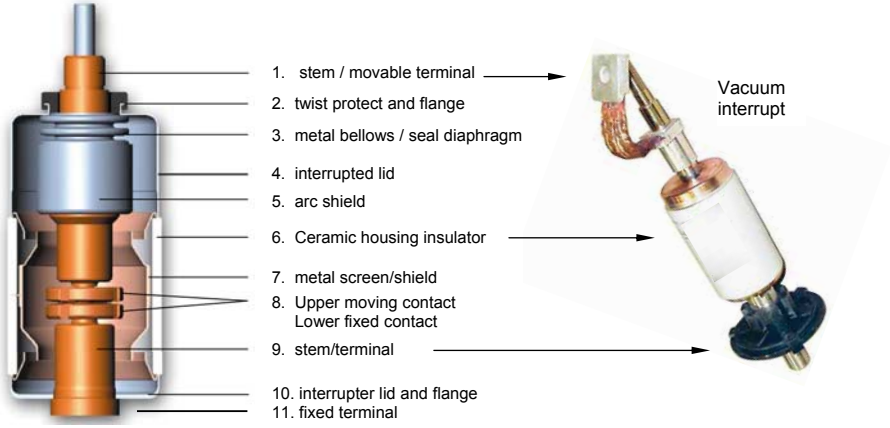


Figure 28.23. *Dielectric strength comparison between different medium, breakdownm voltage versus gap distance.*

Medium voltage contactors are apparatus suitable for operating in ac power applications. The contactors in figure 28.24b consist of a moulded resin monobloc containing vacuum interrupter Al_2O_3 ceramic modules, figure 28.24a, moving parts, electromagnet, the multi-voltage control feeder, and auxiliary accessories. The monobloc is the support for the assembly of the fuse-holder frame. Closing of main contacts is carried out by means of the control electromagnet. Opening is affected by means of a special counteracting spring.



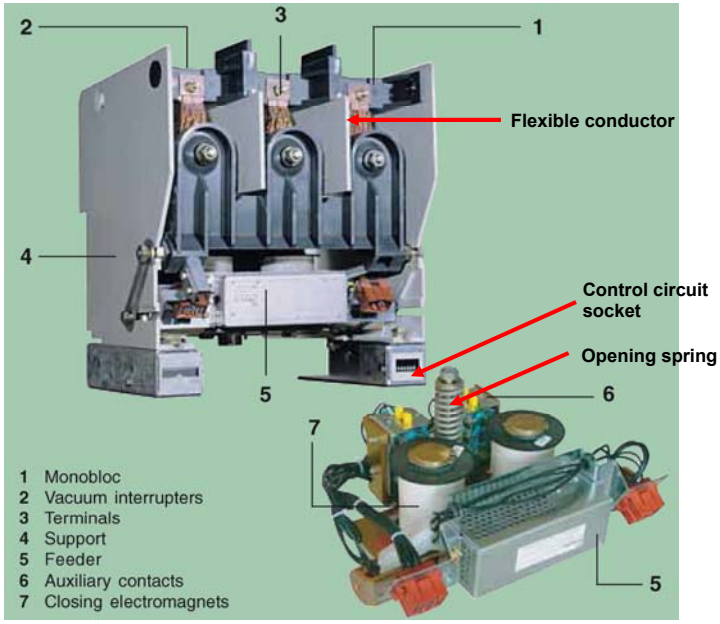


Figure 28.24. *Typical three-phase ac vacuum circuit breaker: (a) vacuum interrupter and (b) MV mechanical three phase contactor.* (source: www.ABB.com).

28.16.1 Basic Interruption Principle

To ensure minimum maintenance and long life, the main contacts of the contactor operate inside the sealed vacuum interrupters (with a vacuum level of 13×10^{-5} Pa).
On opening, there is rapid separation of the fixed and moving contacts in each contactor interrupter in a three-phase contactor. Overheating of the contacts - melting, generated the instant they separate, causes the formation of metallic vapours at the point of final contact which allows an electric arc to be sustained up to the first zero current passage associated with the ac supply. At zero current, cooling of the metallic vapours allows recovery of high dielectric rigidity able to withstand high values of return voltage. The contacts of the interrupters for contactors are made of WCAg alloy or more commonly used CuCr alloy (or also previously CuBi alloy). The hard-metal component, tungsten carbide WC, makes the contacts resistant to erosion caused by arcing. As a result, contacts have an electrical lifetime of several hundred thousand switching cycles at rated current. Also, with WCAg the chopping current is in the region of 0.5A, compared with 3 to 5 A for CuCr, which is used widely in circuit-breakers.
In motor switching versions, the value of the commutated current is less than 0.5A with extremely limited over-voltages. Importantly, high-speed interruption reduces the level of fault damage to equipment.
These interrupter components give high dielectric strength and rapid recovery after arc extinction. This high dielectric strength exhibits moderately low arc energy during high current interruption, thus minimizing contact erosion. The vacuum medium provides quick controlled arc extinction due to high velocity radial diffusion of vaporized special metal alloy contact surfaces during contact separation. This allows rapid recovery of dielectric strength and minimizes over voltages. The interruption capability is not affected by adverse conditions such as altitude, extreme temperature or humidity.
When the current is small and the contact area is large, the arc spreads by itself, and interruption should be successful. If the current to be interrupted increases, on interruption the arc area contracts and remains station at a certain point. Therefore an arc control mechanical structure is used to increase interruption capacity. Two basic structures are used. Spiral contacts generate a radial magnetic force, RMF, in figure 28.25a to rotate the arc, and axial magnetic field, AMF, contacts in figure 28.25b diffuse the arc by axial magnetic force.

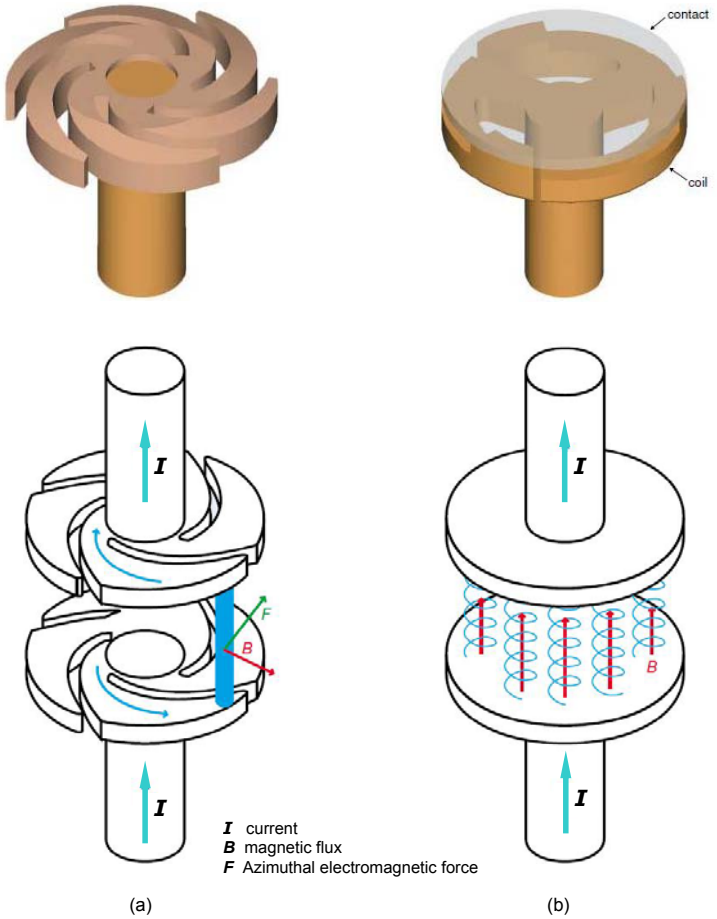


Figure 28.25. *Basic mechanical contacts, with arcing operation based on:*
(a) a spiral contact producing a radial magnetic force, RMF and (b) slots, axial magnetic field, AMF.

i. Quadruple axial magnetic field contact – AMF

Constriction of vacuum arcs at high current levels depends on the contact material and the electrode principle. By applying a radial magnetic force/field, RMF, the constricted arc column is forced to move rapidly around the contact surface. Axial magnetic field AMF contacts within vacuum interrupters, however, prevent the vacuum arc from constricting due to the reduced charge carrier movement perpendicular to the magnetic flux and the magnetic field. This applies especially to the electrons, which have a smaller mass than the ions. The electrons gyrate around the magnetic lines of force, so that the contraction of the arc is shifted towards the higher currents. The diffused arc results in reduced energy impact on the electrodes which is also indicated by the small regular arcing voltage. Therefore, AMF contacts provide excellent high short-circuit current behaviour.

Depending on the design, the local axial field distribution is different. For single-pole arrangements the direction of the AMF is the same within the whole inter-electrode gap. For multiple-pole arrangements the polarity of the field changes. The main reason in designing AMF contact systems is to achieve an AMF distribution that uniformly spreads the thermal stress of the vacuum arc over the entire contact surface.

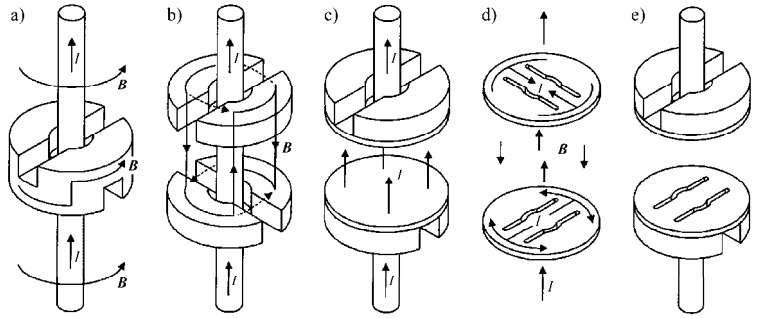


Figure 28.26. *Basic principle of a quadruple axial magnetic field contact:*
(a) ferromagnetic circuit arrangement; (b) magnetic circuit in an open position; (c) contact system, without electrical pole face slots; (d) electrical pole face electrode slots; and (e) electrical plus magnetic combined quadrupolar contact systems.

In AMF contact designs, the magnetic field is generated by a coil arrangement behind the electrode faces. Compared to RMF spiral contacts, where the current flows directly from the stems through the electrodes in closed position of the vacuum interrupter, the continuous current performance is reduced due to the increased resistance of the coil construction. The heat generated by the resistive losses are virtually all dissipated via the oxygen-free high conductivity copper stems to the outside of the vacuum interrupter. The impact of thermal radiation is minimal. The AMF is generated by a hybrid principle. The first contribution is produced by a magnetic circuit, the second by slots incorporated into the electrodes. Both AMF generating measures do not disturb the direct current flow from the copper stems through the closed electrodes. The continuous current performance is therefore comparable to RMF spiral contacts, which is of vital importance for high continuous current applications.

According to Ampere's law, the magnetic field surrounds the copper stem during current flow. By arranging four ferromagnetic pieces as shown in figure 28.26a, the magnetic flux, *B* is guided in the created magnetic circuit as indicated by the arrows. The magnetic flux is forced to penetrate four times the plane between the ferromagnetic pieces and perpendicular in the poles, to the current flow, *I*.

In the case of separating the ferromagnetic pieces as shown in figure 28.26b, the magnetic flux has to cross the gap between the ferromagnetic pieces in an axial direction four times. Due to the increasing distance between the ferromagnetic pieces in axial direction and therefore the increased magnetic reluctance of the magnetic circuit, a part of the magnetic flux does not penetrate the gap in the axial direction. Depending on the magnetic reluctances of the different circuits, a specific part of the magnetic flux closes in the azimuthal direction as indicated in figure 28.26b by the dashed circumferential arrows. This phenomenon progressively decreases the axial magnetic flux density with increased gap distance.

Figure 28.26c shows the principle of a magnetic circuit applied to a contact system of a vacuum interrupter. After opening the electrodes in order to interrupt a current, the described quadrupolar AMF forces the arc into a diffuse mode. By incorporation of slots into the electrodes, as seen in figure 28.26d where the plates have a relative rotation of 90°, a part of the current is diverted from flowing directly through the contact plate to the vacuum arc position on the contact plate. Due to the forced diverted current loops, an auxiliary quadrupolar AMF is generated reinforcing the AMF provided by the magnetic circuit. Figure 28.26e displays the principle of the complete hybrid quadrupolar contact system. The slots of this hybrid principle impact on AMF performance. In a closed position, the dc current flow from the copper stems through the electrodes, results in a continuous current performance similar to RMF spiral electrodes. Compared to RMF spiral electrodes, additional losses are evoked by the eddy currents and the hysteresis within the ferromagnetic pieces.

The thickness of the contact plate is important in dimensioning the contact design. An increased thickness results in a weakening of the AMF, due to the increased reluctance, hence less magnetic flux penetrates the electrodes and the inter-electrode gap. Flux, leakage, is diverted through the gap between the two ferromagnetic materials behind each electrode. However, an increased plate thickness increases thermal capacity during, after, and in-between short-circuit current operation.

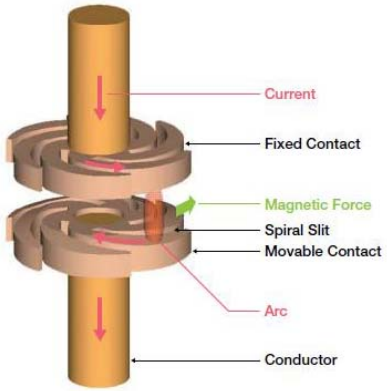


Figure 28.27. *Basic principle of a spiral radial magnetic force contact, showing a current filament flow during interruption.*

ii. Spiral radial magnetic field contact – RMF

The spiral contacts in a vacuum generate a radial magnetic field (RMF), as shown in figures 28.25a and 28.27, which causes an azimuthal electromagnetic force to act on the contracted vacuum arc. The contracted arc moves over the contact's surface at a speed of 70 to 150 m/s. This high velocity ensures that there is less contact erosion and also significantly improves the current interrupting capability. The resultant arc voltage is higher with RMF contacts.

The advantage of the RMF contact system is its simple physical structure, while another advantage of the spiral contact is that in the closed state the current can flow through the contacts directly via the stem, thereby ensuring lower power losses for the vacuum interrupter at nominal current. In most AMF contact based contactors, the axial magnetic field is generated by a coil located behind the contacts. As a result, the resistance of the interrupter is increased and the resultant additional resistive losses reduce the nominal current performance. The only practical way in which a vacuum interrupter can dissipate the generated heat is via the copper conductors, since convection is not possible in a vacuum and radiation is minimal because the larger surface areas are facing and the low emissivity of the contact metals.

In the short-circuit current range above 63kA, the more complex AMF contact system is superior to conventional RMF contacts.

28.16.2 Medium-Voltage AC Vacuum circuit breaker characteristics

i. Vacuum contactors

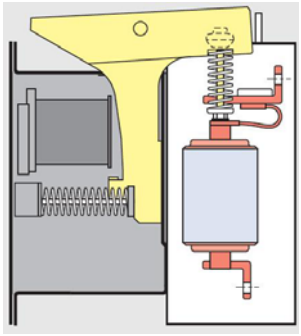
HV contactors and relays require only a 2mm to 12mm contact opening to obtain voltage withstand of 20kV to 70kV pk (or more for series contacts) Operate times of 1 to 16 milliseconds are obtained using the simple operating mechanisms to move the lightweight contacts the short travel required.

Contact resistance is low, usually less than 10μΩ to 1mΩ, depending on the type of contact used. Higher continuous currents are also available with the use of high current shunting switches with capacities of up to 63kA. The contacts are readily adaptable to series connection for higher voltages, with a maximum operating voltage of 30kV to 45kV peak per contact, enabling withstand voltage capabilities to in excess of 300kV. A typical three-phase MV ac contactor and its key characteristics are shown in figure 28.28.

Vacuum contact mechanical life is generally at least 10,000 operations to several million operations, depending on interrupter type, speed of operation, and contact opening distance. Contact electrical life is 1 to 10 operations at maximum interrupt to several million at lower currents; and is much more dependent on closing current than on interrupt currents up to several times rated load current. Closing bounce and amp-seconds during arcing are the main electrical life determining factors.

ii. Operate voltages

Single contact, normally open, normally closed, latching and trip free type HV vacuum contactors and relays are used at 208V rms to 30kV pk, and 50 to 1,200A continuous. In most standard applications, voltage ratings are raised in multiples of 30kV (operating) above 15kV (operating) by placing contacts in series. For example, two 15kV rated type contacts in series per phase are suitable for 45kV, three contacts in series per phase are suitable for 75kV applications.



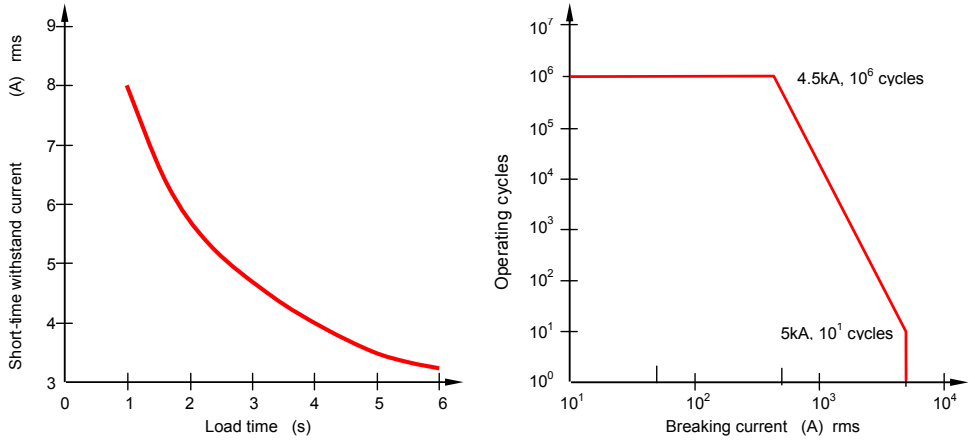


Figure 28.28. *Vacuum contactor, rated voltage=12kV, rated normal current=450A, rated make current=4500A, impulse withstand voltage=75kV.*

iii. Current ratings

Continuous current ratings are from 50A to in excess of 1,200A rms, with up to 36kA rms continuous with the use of a shunt switch to carry the continuous current while using the vacuum interrupter's capability for the actual interruption. This switchgear has 60Hz interrupt ratings of 2kA to 28kA, 10 cycle momentary ratings of 5kA to 60kA and capacitor discharge to 1,000kA. 50Hz maximum interrupt ratings are derated 10%.

iv. AC/DC current interruption

DC: The single pole HV vacuum contactors and relays can be used on both ac and dc current interruption. The units dc rated vacuum contacts can normally be used for limited current interruption to 10A dc. Some ac rated contacts can interrupt dc at much higher currents with carefully controlled displacement pulses to create current zeros, and slow recovery voltage rates similar to those of 60Hz to 400Hz waveforms. Mechanism utilize special tungsten contacts for limited dc current interruption. In some cases, the interruption can be carried to as high as 20 to 40A dc. When interrupting dc, transient suppression such as non-linear resistance or a capacitor in series with a 1Ω/kV inrush limiting resistor should be used in parallel with the vacuum contacts and the load.

For higher dc current interruption, that is, several hundred to several thousand amps or more at 15kV dc per contact, or 40kV dc for 2 contacts in series, ac rated contactors with copper alloy vacuum contacts can be satisfactory. This is possible if multiple displacement-pulses are applied with a controlled rate of recovery voltage that approximates 60Hz current zero and recovery voltage characteristics by means of a resistor/capacitor and switching network or if in combination with an inductive resonant circuit.

AC: For ac interruption, the contacts are designed with copper alloy combinations to limit current chopping to less than a 1A to 8A level to minimize switching transients. AC current interruption generally occurs at the first current zero after contact separation. If the contacts are not sufficiently apart for the rate of recovery voltage, or arc energy is high, current may carry over to the next current zero before clearing. At higher contact voltages and currents, the contacts must not bounce after opening, which would temporarily reduce the contact spacing.

v. Inductive load switching

With highly inductive loads, wherever possible, transient suppressing non-linear resistance or protective capacitors should be placed across the line as close to the load equipment terminals as possible. Standard lightning arresters are not as effective as a protective shunt capacitor with approximately 2Ω/kV of inrush damping resistor in series directly connected across the load. This is essential for repeatedly switched inductive loads such as arc furnaces, motors, and many low current, transient generating inductive loads. Most switching devices create over voltages on switching and for iron core reactive loads, normal switching over voltages of 2 to 2½ times operating voltage occur with any type of interrupter. Dry type transformers and air core inductive loads can generate higher over voltages if there is insufficient shunt capacitance.

vi. Closing and inrush currents

When closing on to transformers and other iron core inductive loads, normal inrush currents of 5 to 10 times rated load current are expected, depending on the magnetic retentivity from the previous interruption. If iron core inductive loads can be re-energized on the opposite polarity from which they were de-energized, , thereby avoiding core saturation, then inrush is minimized, otherwise, it is limited primarily by the circuit and winding resistance and leakage. Repeated high inrush closing causes mechanical stress on the transformer windings as erosion of the vacuum contact (which can be over 10 times as great on closing as on interrupting even the same current). Therefore, transformers that have marginal insulation and mechanical bracing can deteriorate with repeated switching, regardless of type of switch.

vii. Contact erosion

Contact erosion and resultant internal vaporized metal deposit distribution generally determines the end of electrical life. A 2mm to 6mm of total erosion is generally the life limit. With the proper selection of contact material and with currents under 600 to 1kA, erosion should be small. At currents between 1kA and 3.5kA, erosion should be moderate. At currents approaching the maximum interruption rating, life may only be 1 to 100 operations. Even at low to moderate currents, closing generally causes 2 to 10 times as much erosion as interrupting the same currents. Thus, for long-life current closing, the closing current should be limited. Step-start dual closing contactors which use inrush current limiting resistors can be used. Voltage zero and current zero sensing devices for closing or opening are viable.

viii. Actuator types and contactor configurations

Normally open, normally closed, double throw, and latching type configurations are available. Standard actuator voltages are 115V, 60Hz for smaller units, 230V, 60Hz is recommended for larger, heavy-duty solenoid actuated units. 208V, 480V, 24V, 60Hz, 50Hz, and 24V dc, 100/125V dc, 400Hz, and other voltages are available. Actuator voltage is specified plus the applicable HV contact operating voltage and current, basic impulse level (BIL), insulation level for MV contacts and MV contacts to actuator or ground, maximum rms current interrupt, 1 cycle (16.6ms) momentary rms current, 10 cycle rms momentary current, or peak capacitor discharge current and the RC time constant of the current decay to 35% of peak. Other information may include type of load, number of operations per year, maximum current levels on closing and opening, speed of opening and closing, and number and type of auxiliary contacts required.

ix. Auxiliary contacts

Two auxiliary SPDT contacts are usually standard.

x. High voltage protection

Safety regulations require high voltage and line to ground current protection. In 2 to 16 ms the contactor's high speed trip, driven by a relay driver, can close the contacts and divert the fault current, or open the contacts and interrupt the load current, or both, thus minimizing damage to the controlled equipment. When proper ground fault or leakage current sensing is used, this may be fast enough to reduce personnel injury from accidental electrocution because of contact with one line and ground or voltages developed in the ground circuit from ground current.

xi. Uses

Basic MV contactors and relays with sealed vacuum contacts are available in single pole, two-pole, and three-pole variations. Models include normally open, normally closed, or double throw units. Latching actuators are also available. The contactors and relays are used in many high voltage power supplies for capacitor bank charging and discharging, current transfer, tap or load selection, or sealed arc interruption. The units are suited for high-speed interruption of up to 10A dc or 28kA ac. They are also reliable for high-speed crowbar or fault diversion to protect sensitive electronic devices. These various functions are accomplished with 1ms to 4ms contact closure or separation. A high-speed vacuum power interrupter can crowbar in 2ms, divert the fault, then interrupt up to its maximum rated current in approximately 3ms to 16ms, depending on type of system and current zero timing.

28.16.3 Altitude derating

Insulating properties of air decrease as the altitude increases. This phenomenon must be taken into account during the design stage of insulating parts of equipment that will be installed at over 1000m above sea level. In such cases, a correction coefficient is used from the graph in figure 2729, which is calculated from:

$$k_a = e^{\frac{m\times(h-1{,}000)}{8150}}$$

h altitude in metres;
m value considered constant for simplification and equivalent to 1 for power frequency, lightning withstand impulse, and between phases.

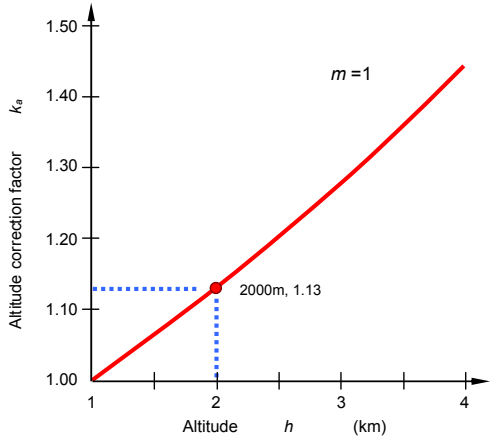


Figure 28.29. *Graph for determining the voltage correction factor k_a according to the altitude.*

Example 28.2: ***Vacuum circuit breaker altitude properties***

A vacuum circuit breaker is installation at an altitude of 2000m. If it used at
- Rated voltage, 7kV
- Power frequency withstand voltage 20kV rms
- Impulse withstand voltage 50kV p

determine its necessary contactor voltage requirements.

Solution

From the graph or from the following equation:

$$k_a = e^{\frac{m\times(h-1{,}000)}{8150}} = e^{\frac{2{,}000-1{,}000}{8150}} = 1.13$$

For the above parameters, the apparatus will have to withstand the following values, based on test values at zero altitude, that is, at sea level.

Power frequency withstand voltage equal to:
20 x 1.13 = 22.6kV rms
Impulse withstand voltage equal to:
50 x 1.13 = 56.5kV p.

For installations at an altitude of 2000m above sea level, with an operating voltage of 7kV, apparatus with a rated voltage of 12kV characterized by power frequency withstand voltage of 28kV rms and with 60/75kV p impulse withstand voltage must be used.

♣

28.18 Corona

Corona (also known as partial discharge) is caused by the electric field next to an object exceeding the breakdown gradient for air, or whatever the field source is immersed in. Since the magnitude of the field is inversely proportional to the radius of curvature, sharper edges break down sooner. The corona initiation voltage is typically 30 kV/cm radius. Dust or water particles on the surface of the object reduce the corona starting voltage, by providing local areas of tighter curvature, and hence higher field stress.
Corona hissing or cracking can often be heard, particularly with a safely placed stethoscope or ultrasonic detector. In addition, the ozone produced by the corona can sometimes smelt.
The presence of corona can reduce the reliability of a system by degrading insulation. While corona is a low energy process, over long periods of time, it can substantially degrade insulators, causing a system to fail due to dielectric breakdown. The effects of corona are cumulative and permanent, and failure can occur without warning. Corona causes:

- Light
- Ultraviolet radiation
- Sound (hissing, or cracking as caused by explosive gas expansions)
- Ozone
- Nitric and various other acids
- Salts, sometimes seen as white powder deposits
- Other chemicals, depending on the insulator material
- Mechanical erosion of surfaces by ion bombardment
- Heat (although generally minimal, and primarily in the insulator)
- Carbon deposits, thereby creating a path for severe arcing

The simplest case to analyze is that of a sphere where the magnitude of the electric field at the surface of the sphere in free space is the voltage/radius. If the sphere is near another conductor, the field is no longer uniform, as the charge will redistribute itself towards an adjacent conductor, increasing the field.
Since corona is a breakdown phenomenon, it follows Paschen's law: the voltage is a function of potential difference, *pd*. Double all the dimensions and halve the gas pressure, and the corona voltage will be much the same.

Arcing occurs instead of corona when the voltage is too high. Corona will not form when

For Concentric Cylinders in Air:
$R_o / R_i < 2.718$
For Parallel Wires in Air:
$x / r < 5.85$
For Equal Spheres in Air:
$x / R < 2.04$
Arcing is difficult to avoid when $x / R < 8$

where
R_o = radius of outer concentric sphere
R_i = radius of inner concentric sphere
R = sphere radius
r = wire radius
x = distance between wires or between spheres

Corona surface factor

Table 28.7 gives empirically determined correction factors for various surface conditions. These factors are multiplied by the corona starting voltage (or field) to determine the corrected effective voltage.

Eliminating or reducing corona

Electrically, reduce or eliminate unwanted voltage transients, which can cause corona to start.

Corona can be avoided by minimizing the voltage stress and electric field gradient. Initially maximize the distance between conductors that have large voltage differentials. Use homogeneous insulators of void free solids, such as properly prepared silicone and epoxy potting materials. Smoothly radiusing the corners of objects at high voltages relative to nearby objects will reduce the local field strength. Put the sharp corner in contact with a substance with a higher breakdown strength than air. Corona can be reduced by making the high field occur within a substance with a higher breakdown than the surrounding air.

Table 28.7: Correction factor for Corona breakdown

Condition of Conductor	m_o
New, unwashed	0.67 - 0.74
Washed with grease solvent	0.91 - 0.93
Scratch-brushed	0.88
Buffed	1.00
Dragged and dusty	0.72 - 0.75
Weathered (5 months)	0.95
Weathered at low humidity	0.92
For general design	0.87 - 0.90
7 strand concentric lay cable	0.83 - 0.87
19, 37, and 61 strand concentric lay cable	0.80 - 0.85

Covering sharp corners with an insulating film increases the corona inception voltage at the points with high E-field stress. Generically known as *corona dope*, this is an enamel or polystyrene paint or gels that can be applied. Clear acrylic spray paint is a generic possibility, although the coating is quite thin.
Potting the entire assembly in an insulator, room temperature vulcanising RTV silicone, achieves the same result. Immersing the assembly in oil or other insulating fluids is also effective. All of the potting and immersion techniques depend on removing the air or gas bubbles to be effective. Commercial manufacturers create a vacuum on the container while the assembly is being potted to facilitate the removal of the air bubbles.
An approach to reducing corona on wires is to surrounding the conductor by a semiconducting film or layer of greater radius. This effectively increases the radius of the object, and hence lowers the field strength. Minimal copper may be needed to carry the required current (often micro or milliamps), but a large diameter conductor is required to reduce the corona.
Field grading rings are often used on high voltage equipment to control the electric field distribution. Rather than rely the field that would exist in free space between two charged conductors, a series of other conductors are interposed at intermediate voltages. The intermediate voltages are derived from a capacitive or resistive divider. A capacitive divider may be a simple as the inter electrode capacitances of the grading rings themselves.
Running the system in a tank at high pressure, or in an insulating gas, increases the corona starting voltage.

28.19 Appendix: Contact metals

Table 28.8. Rare and precious metals used for mechanical contacts

Metal or alloy	*Melting point* (°C)	*Vickers hardness (annealed)*	*Density* (kg/m^3)	*Resistivity at 20°C* (Ω -m×10^8)
Light-duty contacts				
Gold	1064	20	19 200	2.2
Platinum	1770	65	21 450	10.6
10% Iridium–platinum	1780	120	21 600	24.5
20% Iridium–platinum	1815	200	21 700	30.0
25% Iridium–platinum	1845	240	21 700	32.0
30% Iridium–platinum	1885	285	21 800	32.3
25% Iridium–ruthenium–platinum	1890	310	20 800	39.0
7% Platinum–silver–gold	1100	60	17 100	16.8
30% Silver–gold	1025	32	16 600	10.4
30% Silver–copper–gold	1014	95	14 400	14.0
10% Silver–copper–gold	861	160	13 700	12.5
Rhodium	1960	40	12 400	4.9
Iridium	2447	220	22 400	5.1
Palladium	1554	40	12 000	10.8
40% Silver–palladium	1290	95	11 900	35.8
40% Copper–palladium	1200	145	10 400	35.0
Medium–duty contacts				
10% Gold–silver	965	30	11 400	3.6
20% Palladium–silver	1070	55	10 700	10.1
10% Palladium–silver	1000	40	10 600	5.8
5% Palladium–silver	965	33	10 500	3.8
Fine silver	961	26	10 500	1.6
0.2% Magnesium–0.2% nickel–silver	961	140	10 400	2.8
1% Graphite–silver	961	40	9 900	1.8
2% Graphite–silver	961	40	9 700	2.0
Standard silver	778	56	10 300	1.9
10% Copper–silver	778	60	10 300	2.0
10% Cadmium oxide–silver	850	50	9 800	2.1
10% Nickel–silver	961	40	10 300	2.0
15% Cadmium oxide–silver	850	60	10 000	2.3
20% Copper–silver	778	85	10 200	2.1
20% Nickel–silver	961	48	10 100	2.1
Cadmium–copper–silver	800	65	10 100	4.2
50% Copper–silver	778	95	9 700	2.1
Heavy–duty contacts				
10% Cadmium oxide–silver	850	55	10 000	2.1
15% Cadmium oxide–silver	850	65	9 800	2.3
40% Tungsten carbide–silver	960	90	11 900	2.5
45% Tungsten carbide–silver	960	95	12 200	2.8
50% Tungsten–silver	960	125	13 600	2.8
50% Tungsten carbide–silver	960	160	12 500	3.0
55% Tungsten–silver	960	140	13 400	3.0
60% Tungsten carbide–silver	960	200	13 200	4.8
65% Tungsten–silver	960	185	14 800	3.3
73% Tungsten–silver	960	220	15 600	4.0
78% Tungsten–copper	1080	240	15 200	6.1
68% Tungsten–copper	1080	160	13 600	5.3
60% Tungsten–copper	1080	140	12 800	4.3

Reading list

http://relays.tycoelectronics.com/
http://www.gigavac.com/
http://www.jenningstech.com/index.html

blank

Nomenclature and symbols

APR	axial power rating
BJT	bipolar junction transistor
CoP	coefficient of performance
CTE	coefficient of thermal expansion
CVD	chemical vapour deposition
DCB	direct copper bonding
EBL	electron beam lithography
EGS	electronic grade polysilicon
GTO	gate turn-off (thyristor)
IGBT	insulated gate bipolar transistor
LPCVD	low pressure CVD
MBE	molecular beam epitaxy
MFD	magneto-fluid-dynamic
MMF	magnetomotive force
MOSFET	metal oxide semiconductor field effect transistor
mtbf	mean time between failures
mttf	mean time to failure
NTD	neutron transmutation doping
PC	permeance coefficient ($B_d / \mu_o H_d$)
PCM	phase change material
PECVD	plasma enhanced CVD
POH	power on hours
PSG	phospho-silicate glass
PVD	physical vapour deposition
PVT	physical vapour transport
RIE	reactive ion etching
s/b	second breakdown
ScD	skeleton cemented diamond
SCR	silicon controlled rectifier
SOA	safe operating area
TEC	thermoelectric cooler
TIM	thermal interface material

α	gate threshold temperature dependence coefficient
α	temperature coefficient of on-state resistance
α	thermal coefficient of linear expansion, K^{-1}
α	current transfer ratio
α_o	current transfer ratio in mid current region
α_s	characteristic life (hours)
β	base-to-collector current amplification factor
β_f	forward current gain
β_Q	GTO turn-off gain
β_r	reverse current gain
$ß_s$	shape parameter
γ	surface tension, N/m
γ_i	injection efficiency
Γ	gamma function
δ	on-state duty cycle factor
δT, ΔT	temperature difference between regions of heat transfer, T_2-T_1, K
ΔP	system static pressure loss [1Pascal = 1N/m²,]
ΔP_{cmax}	maximum capillary pressure difference between the evaporator and condenser
ΔP_g	hydrostatic pressure drop
ΔP_{liquid}, $\Delta P_{vavpour}$	viscous pressure drops in liquid and vapour phases
P_v	velocity pressure, $½\rho v^2$
ΔT	thermal shock temperature
ΔT	desired air temperature differential (enclosure inlet to discharge ambient air), K
ΔT_{sa}	average temperature difference between heat sink and ambient air

ε	surface property, termed emissivity, $0 \le \varepsilon \le 1$
ε_a	apparent emissivity of a channel
ε_o	free space permittivity, 8.854×10^{-12} F/m
ε_r	relative dielectric constant
ε_s	dielectric permittivity $\varepsilon_s = \varepsilon_r \varepsilon_o$
η_f	fin efficiency
η_v	volumetric heat transfer efficiency
θ	contact angle, rad,
θ_f	volume figure, dimensionless
λ	thermal conductivity, W/cm.K
λ	latent heat, J/kg
λ	wavelength
λ_{eff}	effective thermal conductivity for a heat pipe
μ_n, μ_p	hole/electron mobility, cm^2/V-s
μ_o μ_m μ_{rc}	permeability of vacuum/air, magnet, recoil
v	absolute fluid kinematic viscosity, Ns/m^2, Pa
ξ	electric potential, V/m
ξ_b	breakdown field, V/m
ρ	resistivity, Ω.cm
ρ_m	density of the heatsink material, kg/m^3
ρ_f	density of working fluid (e.g. air, liquid) medium, (= 1/v specific volume), kg/m^3
σ	conductivity, $\Omega^{-1}.cm^{-1}$
σ	Stefen-Boltzmann constant, 5.667×10^{-8} W/m^2K^4
σ	surface tension, N/m
σ_p	symmetrical standard deviation, cm
τ	period of the switching interval (both on and off), s
τ	thermal time constant, s
τ_h, τ_e	minority carrier hole/electron lifetime, s
φ	kT/q, thermal voltage, built in potential, V
Φ	zero external bias, built-in, junction potential or scl potential, V
Φ_b	Schottky barrier height
Φ_f	pressure figure, dimensionless parameter
ω	rotational (angular) velocity at the perimeter
A	total surface area (of die/outside/heatsink fins and base between fins) involved in the heat transfer, cooling, m^2
A_e, A_c	effective evaporator and condenser surface areas
A_g, A_m	cross-sectional area of air gap, magnet
A_x	cross sectional area (fin), m^2
b	thickness of the heat sink, mm
b_t	base transport factor
B	magnetic flux density (induction)
B_d	flux density in magnet at operating point on demagnetization curve
B_g	flux density in air gap
$(BH)_{max}$	maximum energy product
B_i	intrinsic flux density (induction) in a magnet
B_r	residual induction in magnet
B_{sat}	saturation flux density
cd	critical line width
c_p	specific heat capacity of the cooling fluid at constant pressure, $W/m\Delta T$, J/kg.K
C	linear rate constant
C_a	capacitance per unit area of the gate oxide, ε / t_{ox}

C_{ds} non-linear voltage-dependent drain to the source capacitance
C_{gd} non-linear voltage-dependent gate to the drain capacitance
C_{gs} non-linear voltage-dependent gate to the source capacitance
C_f correction factor for position and surface emissivity of heat-sink orientation
C_{in} gate input capacitance, approximately $C_{gd} + C_{gs}$, or C_{iss}
C_{iss} input capacitance
$C_j(v)$ voltage dependant scl capacitance, F
C_{jo} zero bias junction capacitance, F
C_{ob} output capacitance, essentially C_{ds}
C_{oss} common source output capacitance
C_r basic dynamic load capacity, kg
C_R correction factor
C_{rss} reverse transfer capacitance
C'_{sf} fluid surface combination constant
$C_t(v)$ voltage dependant transit capacitance, F

d diameter, m
d_1 n^- drift region width
di_F/dt forward current rate of change, A/s
di_{rr}/dt reverse recovery current rate of change, A/s
dm/dt mass evaporation rate
dv/dt anode impressed dv/dt
D diffusion or diffusivity coefficient, $\mu kT/q = \lambda / \gamma c_p$, m^2/s
D_o outer diameter of the fan impeller, m
D_H diameter (hydraulic/bore), mm
D_n, D_p hole/electron carrier diffusivity
DN_L speed limit, rpm-mm

E emf, circuit applied reverse voltage
E_a activation or threshold energy, eV
E_g band gap, eV
E_o diode model on-state voltage source, V

f friction factor (loss coefficient)
f_s switching frequency, Hz
F(t) cumulative distribution function, a function of age t
$\mathfrak{F}_m$ magnetomotive force

g gap
g_f gravitational acceleration, m/s^2
g_{fs} amplification factor, forward *transconductance n-channel,*
g_d output conductance
G volumetric fluid flow rate, m^3/s

h convection/conduction thermal heat transfer coefficient (of surface material) W/m^2K
h_r radiation heat transfer coefficient, W/m^2K
H capillary or lifting height, height of the fin, length (of heat sink base), m
HP impeller input power to rotate
H(t) hazard rate, failure rate or hazard function
H magnetic field strength, magnetizing force, demagnetizing force
H_c coercive force
H_{ci} intrinsic coercive force
H_d demagnetizing force at operating point of magnet on demagnetization curve
H_g magnetizing force in air gap

i_R leakage current, A
i_F forward current, A
i_{rr} reverse recovery current
I current, A
I_b base current
I_b reverse voltage breakdown diode current, A
I_{bf} forward base current
I_{br} reverse base current
I_c collector current

I_{ceo} collector current when $I_b = 0$ for $V_{(BR)ceo}$
I_{cer} collector current when R_{be} = R, for $V_{(BR)cer}$
I_{ces} collector current when $V_{be} = 0$ for $V_{(BR)ces}$
I_{cex} collector current when V_{eb} = X, for $V_{(BR)cex}$
I_d drain current n-channel
I_{dp} drain current p-channel
I_{DQ} positive - negative temperature coefficient boundary
I_{te} TE current drawn
I_e emitter current, A
I_{GQ} reverse gate current, A
I_G, I_g SCR/triac gate current
I_F diode (maximum) forward current, A
I_{FMG} peak forward gate current, A
I_g gate current, A
I_{GT}, V_{GT} minimum trigger values
I_G, V_G dc gate signal
I_{GQ} GTO minimum negative gate current at anode current I_{GTQ}
I_H holding current, $I_{Latch} > I_H$
I_K I_A SCR cathode/anode/gate current, A
I_L load current, A
I_{Latch} anode latching current, $I_{Latch} > I_H$
I_m maximum current level, A
I_{nom} nominal current, A
I_o reverse (saturation) leakage current, A
I_{RG} negative gate current, A
I_{RM} peak reverse recovery current, A
I_T GTO on-state current , A
I_{tail} storage current level, A
I_{RRM}, I_{DRM} reverse leakage and forward blocking current, A
I_{TSM} peak one cycle surge on-state current, A
I_{rms} rms current, A
I^2t thermal energy crated rating, A^2s

J flux – heat, impurities
J_R reverse recovery W.s/pulse,
J1, J1, J1 SCR junctions

k constant
k Boltzmann's constant, 1.38×10^{-23}, J/K
k_{exp} load factor specific to the system, determined experimentally
k_f leakage coefficient
k_r loss or reluctance factor

k_D characteristics dimension of the geometry
k_T grease temperature factor
K heat transfer coefficient constant units
K thermal resistance pu area, cm^2/W
K_q, K_m, K_p, K_{HP} constants for geometrically and dynamic operation
K_w wire current constant

ℓ vertical height in the direction of the airflow
ℓ thickness of insulation, m
ℓ distance (thickness)
ℓ length
ℓ_g, ℓ_m length of air gap, magnet
ℓ_m line resolution
L heat of vaporization per unit mass
L length (of cold plate), m
L characteristic passage length of the microchannel
L fin depth
L circuit inductance, H
L, W, t length/width/thickness, m
L_c effective channel length
L_{eff} effective length, $\frac{1}{2}(L_{evaporator} + L_{condenser}) + L_{adiabatic}$
L_p minority carrier diffusion length

L_t service life, hours
L_w sound pressure level, dB
L_2 L_{10} life for 98%, 90% survival, second and tenth percentiles

m breakdown multiplication exponent
m mass (weight) of object, kg (density x volume)
m_f mass flow rate of air/fluid through enclosure/heatsink, (equal to $\rho v_f s L$), kg/s
M voltage dependant avalanche multiplication effect
M_ℓ merit number (liquid transport factor), W/m^2

n exponent
n_i intrinsic carrier concentration, 1.4x10^{10} /cc
n_f number of fins
n_q airflow quality constant
$\overline{n}$ index of refraction
N_D, N_A donor/acceptor concentration, cm^{-3}
N speed (fan impeller), rps/rpm
N number of cycles
NA numerical aperture
N_B background doping, cm^{-3}
N_c concentration of the lighter doped region /cc
N_{nom} nominal speed
Nu Nusselt number, non-dimensional heat transfer coefficient, $h\delta/k$, ARe^mPr^n

p, n electron/hole concentration, cm^{-3}
p_o, n_o hole/electron equilibrium carrier concentrations, cm^{-3}
P mean heat added (or being removed - dissipated) from the object, W, watts
P heat transport rate
P equivalent dynamic bearing load, kg
P pressure
P permeance (inverse of reluctance)
P_{cold} amount of heat absorbed at the cold surface of TEC, W
P_c conduction power loss
P_c permeance coefficient
P_d heat load (lost/gained), electrical power dissipated, rate of radiated heat transfer, W
P_D amount/conducted heat dissipated (in enclosure, transferred to cooling system), W
P_{Dtotal} total power to be dissipated
$\hat{P}_d$ maximum allowable power dissipation, W
P_{ref} reference pressure
P_{max} maximum power
P_{hot} minimum total heat to be rejected by the heat exchanger on the hot side
P_G drive input device power loss
P_{GM}, P_G peak and mean gate power
P_f off-state leakage power loss
P_L *load* electrical power dissipation, W
P_{RQ} storage and fall time power loss
P_s switching transition power loss
P_{tec} TEC input dc power
P_{to} initial heat pumping capacity when ΔT is zero
P_{tt} heat pumping capacity at desired ΔT and heat-pumping capacity is decreased
P_T power transported by the heat pipe
$P_\square$ fin perimeter, m
PWL sound power level

q electron charge, 1.602x10^{-19} C
Q pool boiling heat transfer rate
Q_{flow} heat flow
Q_o zero bias scl charge, C
Q_R reverse recovery charge, $Q_1 + Q_2$, C
Q_T total gate charge
Q_Σ total recovery charge, C

r radius, m
r_c effective capillary radius

$r(t_p)$ normalising factor
R resistance, Ω
R_{be} models the lateral p-body resistance
$R_{ds(on)}$ on-resistance n/p-channel
R_d deposition rate
Re Reynolds number, ratio of inertia forces to viscous forces in the fluid, $V\delta/v$
R_g gate resistance
$R_{g\ int}$ internal gate resistance
$R_{g\ ext}$ external gate resistance
R_l resistance modelling linear leakage current, Ω
R_L load resistance, Ω
R_o diode model series resistance, Ω
R_s sheet resistance, Ω/square
R_t thermal resistance of one channel
R_θ thermal resistivity/resistivity
$R_{\theta\ j\text{-}c}$ virtual junction to case thermal resistance, K/W
$R_{\theta\ j\text{-}a}$ total thermal resistance from the virtual junction to the open air (ambient), K/W
$R_{\theta\ c\text{-}s}$ case-to-heat-sink thermal resistance, K/W
$R_{\theta\ s\text{-}a}$ heat-sinking thermal resistance
$\mathfrak{R}_m$ magnet reluctance

s fin spacing
S initial dose per unit area at the surface, cm^{-2}
S_f, S_{fm}, S_{fs} selectivity (mask, substrate)
S_G grease half-life subtraction factor
S_N speed half-life subtraction factor
S_P load half-life subtraction factor
SPL sound pressure level, decibels, dB(A)
S_r snap-off and soft recovery diode properties
$S_{½}$ half-life subtraction factor
$S_{\%s}$ fraction of solids

t fin/plate thickness
t time (required to cool down (or heat up) object), s
t_d delay time
$t_{d\ off}$ turn-off delay time
$t_{d\ on}$ turn-on delay
t_{fi} current fall time
t_{fr} forward recovery characteristics of time
t_{fv} voltage fall time
t_p power pulse width
t_o time to zero current, s
t_{off} turn-off time, $t_s + t_{fi}$
t_{on} turn-on time, $t_d + t_{ri}$
t_{ox} oxide thickness, m
t_{ri} current rise time
t_{rr} reverse recovery time, s
t_{rv} voltage rise, time
t_s storage or saturation time
t_t minority carrier lifetime
t_{tail} current fall time
T cycle or integration period
T absolute temperature, K
T_a ambient temperature
T_A is the ambient temperature
T_{brg} bearing temperature, K
T_c case temperature, K
T_f final temperature, °C
T_{hot}, T_{cold}, (or T_h and T_c)
T_{in}, T_{out} fluid (air, water, etc.) inlet and outlet temperatures, K
$T_{i/s}$ inside temperature, °C
T_j junction temperature, K
$\hat{T}_j$ maximum allowable junction temperature, K
T_{max} maximum operating temperature or desired cold plate surface temperature, K

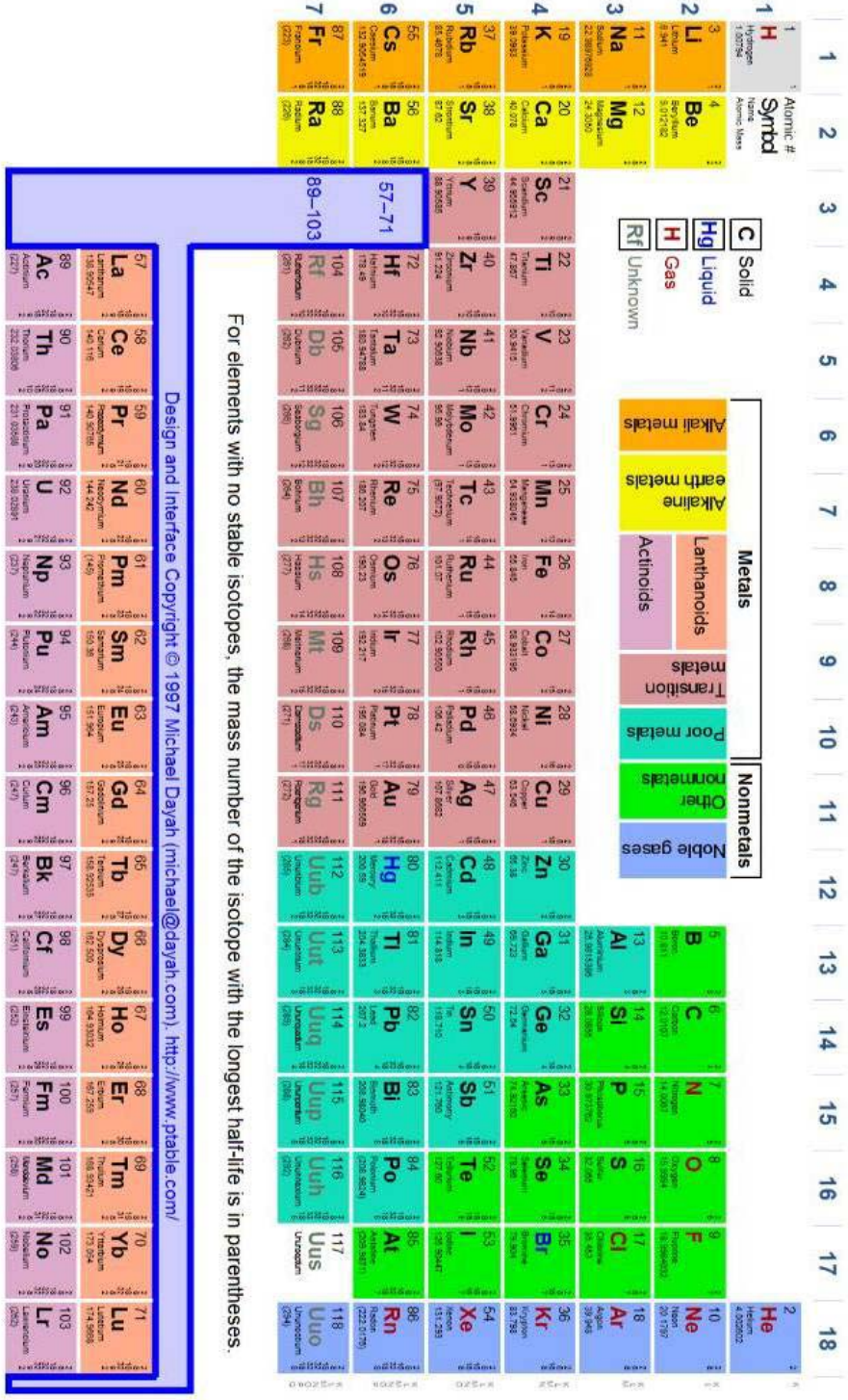
T_{mean} arithmetic mean of T_1 and T_2, specifically ½(T_1 + T_2)
T_{melt} melting temperature
T_o initial/starting temperature, °C
$T_{o/s}$ outside temperature, °C
T_s heated surface temperature
T_{sat} liquid saturation temperature (boiling point)
T_{wall} heated surface temperature

v cooling fluid flow rate, fluid velocity (volumetric flow rate), m^3/s
v velocity of the vertical airflow
v_{ce} collector to emitter voltage
v_{DR} reverse voltage
v_{sat} saturation velocity of electrons in silicon, $9.0x10^6$ cm/s
V, v voltage, V
$V_A(t)$ anode turn-off voltage
V_b avalanche voltage, V
$V_{be(sat)}$ base to emitter saturation voltage
V_{BF} forward anode-cathode breakover voltage
V_{BR} reverse breakdown breakover voltage
$V_{(BR)cbo}$ $V_{(BR)ceo}$ maximum collector-base voltage with the emitter open circuit, base open
$V_{(BR)cex}$ maximum collector-emitter voltage with specific base V-R conditions
$V_{(BR)DSS}$ drain breakdown voltage, V
V_{cbo} collector to base avalanche breakdown voltage
V_{ceo} collector to emitter first breakdown
V_{ceo}, V_{cbo} BJT voltage characteristics
V_{cer}, V_{ces}, V_{cev} BJT voltage characteristics dependent on the external base circuit conditions
$V_{ce(sat)}$ collector emitter saturation voltage
V_{DRM} forward off-state
V_{ds} drain to source voltage
V_{dd} supply voltage
V_{te} TE voltage applied
V_F diode forward voltage, V
V_f velocity between the fins
V_{fp} peak forward voltage
V_g, V_m volume of air gap, magnet
V_{GC} gate junction voltage
V_{gg} gate source voltage maximum
V_{GFM}, V_{GRM} peak forward and reverse gate to cathode voltage
V_{gs} gate voltage
V_{Ls} stray inductance
V_{PT} punch-through voltage, V
V_{RM} diode recovery minimise voltage overshoot, V
V_{RRM} reverse direction
V_s maximum voltage level
V_t tip speed of impeller
V_{Th} gate threshold voltage
V_Z Zener breakdown voltage

w width
W energy, J
W_c width of region/channel, µm
W_{hs} width of heat sink base
W_i intrinsic *i*-layer thickness
W_{n2} *n*-base width
W_o zero bias scl width, m
W_{ref} acoustic reference power
W_{scl} scl width

x_j metallurgical/impurity junction depth, m
x_n, x_p scl penetration into n/p sides, m
X_p penetration depth peak, m

Z altitude, above sea level, m
$Z_{\theta j\text{-}c}$ thermal impedance, K/W



Emissivities of surfaces
(a) Metals

Material	Temperature, K	Emissivity, ε
Aluminum		
Polished	300–900	0.04–0.06
Commercial sheet	400	0.09
Heavily oxidized	400–800	0.20–0.33
Anodized	300	0.8
Bismuth, bright	350	0.34
Brass		
Highly polished	500–650	0.03–0.04
Polished	350	0.09
Dull plate	300–600	0.22
Oxidized	450–800	0.6
Chromium, polished	300–1400	0.08–0.40
Copper		
Highly polished	300	0.02
Polished	300–500	0.04–0.05
Commercial sheet	300	0.15
Oxidized	600–1000	0.5–0.8
Black oxidized	300	0.78
Gold		
Highly polished	300–1000	0.03–0.06
Bright foil	300	0.07
Iron		
Highly polished	300–500	0.05–0.07
Case iron	300	0.44
Wrought iron	300–500	0.28
Rusted	300	0.61
Oxidized	500–900	0.64–0.78
Lead		
Polished	300–500	0.06–0.08
Unoxidized, rough	300	0.43
Oxidized	300	0.63
Magnesium, polished	300–500	0.07–0.13
Mercury	300–400	0.09–0.12
Molybdenum		
Polished	300–2000	0.05–0.21
Oxidized	600–800	0.80–0.82
Nickel		
Polished	500–1200	0.07–0.17
Oxidized	450–1000	0.37–0.57
Platinum, polished	500–1500	0.06–0.18
Silver, polished	300–1000	0.02–0.07
Stainless steel		
Polished	300–1000	0.17–0.30
Lightly oxidized	600–1000	0.30–0.40
Highly oxidized	600–1000	0.70–0.80
Steel		
Polished sheet	300–500	0.08–0.14
Commercial sheet	500–1200	0.20–0.32
Heavily oxidized	300	0.81
Tin, polished	300	0.05
Tungsten		
Polished	300–2500	0.03–0.29
Filament	3500	0.39
Zinc		
Polished	300–800	0.02–0.05
Oxidized	300	0.25

Boiling- and freezing-point properties

	Boiling Data at 1 atm		*Freezing Data*		*Liquid Properties*		
Substance	Normal Boiling Point, °C	Latent Heat of Vaporization, h_{fg} kJ/kg	Freezing Point, °C	Latent Heat of Fusion, h_{if} kJ/kg	Temp., °C	Density, ρ kg/m³	Specific Heat, C_p kJ/kg · °C
Ammonia	−33.3	1357	−77.7	322.4	−33.3	682	4.43
					−20	665	4.51
					0	639	4.62
					25	603	4.78
Argon	−185.9	161.6	−189.3	28	−185.6	1394	1.14
Benzene	80.2	394	5.5	126	20	879	1.72
Brine (20% sodium chloride by mass)	103.9	—	−17.4	—	20	1150	3.11
n-Butane	−0.5	385.2	−138.5	80.3	−0.5	601	2.31
Carbon dioxide	−78.4*	230.5 (at 0°C)	−56.6		0	298	0.59
Ethanol	78.2	838.3	−114.2	109	25	783	2.46
Ethyl alcohol	78.6	855	−156	108	20	789	2.84
Ethylene glycol	198.1	800.1	−10.8	181.1	20	1109	2.84
Glycerine	179.9	974	18.9	200.6	20	1261	2.32
Helium	−268.9	22.8	—	—	−268.9	146.2	22.8
Hydrogen	−252.8	445.7	−259.2	59.5	−252.8	70.7	10.0
Isobutane	−11.7	367.1	−160	105.7	−11.7	593.8	2.28
Kerosene	204–293	251	−24.9	—	20	820	2.00
Mercury	356.7	294.7	−38.9	11.4	25	13560	0.139
Methane	−161.5	510.4	−182.2	58.4	−161.5	423	3.49
					−100	301	5.78
Methanol	64.5	1100	−97.7	99.2	25	787	2.55
Nitrogen	−195.8	198.6	−210	25.3	−195.8	809	2.06
					−160	596	2.97
Octane	124.8	306.3	−57.5	180.7	20	703	2.10
Oil (light)					25	910	1.80
Oxygen	−183	212.7	−218.8	13.7	−183	1141	1.71
Petroleum	—	230–384			20	640	2.0
Propane	−42.1	427.8	−187.7	80.0	−42.1	581	2.25
					0	529	2.51
					50	449	3.12
Refrigerant-134a	−26.1	216.8	−96.6	—	−50	1443	1.23
					−26.1	1374	1.27
					0	1295	1.34
					25	1207	1.43
Water	100	2257	0.0	333.7	0	1000	4.22
					25	997	4.18
					50	988	4.18
					75	975	4.19
					100	958	4.22

Properties of solid nonmetals

Composition	Melting Point, K	Properties at 300 K: ρ kg/m³	C_p J/kg · K	k W/m · K	α × 10⁶ m²/s	Properties at Various Temperatures (K), k(W/m · K)/C_p(J/kg · K): 100	200	400	600	800	1000
Aluminum oxide, sapphire	2323	3970	765	46	15.1	450 —	82 —	32.4 940	18.9 1110	13.0 1180	10.5 1225
Aluminum oxide, polycrystalline	2323	3970	765	36.0	11.9	133 —	55 —	26.4 940	15.8 1110	10.4 1180	7.85 1225
Beryllium oxide	2725	3000	1030	272	88.0			196 1350	111 1690	70 1865	47 1975
Boron	2573	2500	1105	27.6	9.99	190 —	52.5 —	18.7 1490	11.3 1880	8.1 2135	6.3 2350
Boron fiber epoxy (30% vol.) composite	590	2080									
k, ∥ to fibers				2.29		2.10	2.23	2.28			
k, ⊥ to fibers				0.59		0.37	0.49	0.60			
C_p			1122			364	757	1431			
Carbon											
Amorphous	1500	1950	—	1.60	—	0.67 —	1.18 —	1.89 —	21.9 —	2.37 —	2.53 —
Diamond, type IIa insulator	—	3500	509	2300		10,000 21	4000 194	1540 853			
Graphite, pyrolytic	2273	2210									
k, ∥ to layers				1950		4970	3230	1390	892	667	534
k, ⊥ to layers				5.70		16.8	9.23	4.09	2.68	2.01	1.60
C_p			709			136	411	992	1406	1650	1793
Graphite fiber epoxy (25% vol.) composite	450	1400									
k, heat flow ∥ to fibers				11.1		5.7	8.7	13.0			
k, heat flow ⊥ to fibers				0.87		0.46	0.68	1.1			
C_p			935			337	642	1216			
Pyroceram, Corning 9606	1623	2600	808	3.98	1.89	5.25 —	4.78 —	3.64 908	3.28 1038	3.08 1122	2.96 1197
Silicon carbide	3100	3160	675	490	230			— 880	— 1050	— 1135	87 1195
Silicon dioxide, crystalline (quartz)	1883	2650									
k, ∥ to c-axis				10.4		39	16.4	7.6	5.0	4.2	
k, ⊥ to c-axis				6.21		20.8	9.5	4.70	3.4	3.1	
C_p			745			—	—	885	1075	1250	
Silicon dioxide, polycrystalline (fused silica)	1883	2220	745	1.38	0.834	0.69 —	1.14 —	1.51 905	1.75 1040	2.17 1105	2.87 1155
Silicon nitride	2173	2400	691	16.0	9.65	— —	— 578	13.9 778	11.3 937	9.88 1063	8.76 1155
Sulfur	392	2070	708	0.206	0.141	0.165 403	0.185 606				
Thorium dioxide	3573	9110	235	13	6.1			10.2 255	6.6 274	4.7 285	3.68 295
Titanium dioxide, polycrystalline	2133	4157	710	8.4	2.8			7.01 805	5.02 880	3.94 910	3.46 930

Properties of solid metals

Composition	Melting Point, K	Properties at 300 K: ρ kg/m³	C_p J/kg · K	k W/m · K	α × 10⁶ m²/s	Properties at Various Temperatures (K), k(W/m · K)/C_p(J/kg · K): 100	200	400	600	800	1000
Aluminum:											
Pure	933	2702	903	237	97.1	302 482	237 798	240 949	231 1033	218 1146	
Alloy 2024-T6 (4.5% Cu, 1.5% Mg, 0.6% Mn)	775	2770	875	177	73.0	65 473	163 787	186 925	186 1042		
Alloy 195, Cast (4.5% Cu)		2790	883	168	68.2			174	185		
Beryllium	1550	1850	1825	200	59.2	990 203	301 1114	161 2191	126 2604	106 2823	90.8 3018
Bismuth	545	9780	122	7.86	6.59	16.5 112	9.69 120	7.04 127			
Boron	2573	2500	1107	27.0	9.76	190 128	55.5 600	16.8 1463	10.6 1892	9.60 2160	9.85 2338
Cadmium	594	8650	231	96.8	48.4	203 198	99.3 222	94.7 242			
Chromium	2118	7160	449	93.7	29.1	159 192	111 384	90.9 484	80.7 542	71.3 581	65.4 616
Cobalt	1769	8862	421	99.2	26.6	167 236	122 379	85.4 450	67.4 503	58.2 550	52.1 628
Copper:											
Pure	1358	8933	385	401	117	482 252	413 356	393 397	379 417	366 433	352 451
Commercial bronze (90% Cu, 10% Al)	1293	8800	420	52	14		42 785	52 160	59 545		
Phosphor gear bronze (89% Cu, 11% Sn)	1104	8780	355	54	17		41 —	65 —	74 —		
Cartridge brass (70% Cu, 30% Zn)	1188	8530	380	110	33.9	75	95 360	137 395	149 425		
Constantan (55% Cu, 45% Ni)	1493	8920	384	23	6.71	17 237	19 362				
Germanium	1211	5360	322	59.9	34.7	232 190	96.8 290	43.2 337	27.3 348	19.8 357	17.4 375
Gold	1336	19,300	129	317	127	327 109	323 124	311 131	298 135	284 140	270 145
Iridium	2720	22,500	130	147	50.3	172 90	153 122	144 133	138 138	132 144	126 153
Iron:											
Pure	1810	7870	447	80.2	23.1	134 216	94.0 384	69.5 490	54.7 574	43.3 680	32.8 975
Armco (99.75% pure)		7870	447	72.7	20.7	95.6 215	80.6 384	65.7 490	53.1 574	42.2 680	32.3 975
Carbon steels:											
Plain carbon (Mn ≤ 1%, Si ≤ 0.1%)		7854	434	60.5	17.7			56.7 487	48.0 559	39.2 685	30.0 1169

Properties of solid metals *(Continued)*

Composition	Melting Point, K	Properties at 300 K: ρ kg/m³	C_p J/kg · K	k W/m · K	$\alpha \times 10^6$ m²/s	Properties at Various Temperatures (K), k(W/m · K)/C_p(J/kg · K): 100	200	400	600	800	1000
AISI 1010		7832	434	63.9	18.8			58.7	48.8	39.2	31.3
								487	559	685	1168
Carbon–silicon (Mn ≤ 1%, 0.1% < Si ≤ 0.6%)		7817	446	51.9	14.9			49.8	44.0	37.4	29.3
								501	582	699	971
Carbon–manganese–silicon (1% < Mn ≤ 1.65% 0.1% < Si ≤ 0.6%)		8131	434	41.0	11.6			42.2	39.7	35.0	27.6
								487	559	685	1090
Chromium (low) steels:											
$\frac{1}{2}$ Cr–$\frac{1}{2}$ Mo–Si (0.18% C, 0.65% Cr, 0.23% Mo, 0.6% Si)		7822	444	37.7	10.9			38.2	36.7	33.3	26.9
								492	575	688	969
1Cr–$\frac{1}{2}$ Mo (0.16% C, 1% Cr, 0.54% Mo, 0.39% Si)		7858	442	42.3	12.2			42.0	39.1	34.5	27.4
								492	575	688	969
1Cr–V (0.2% C, 1.02% Cr, 0.15% V)		7836	443	48.9	14.1			46.8	42.1	36.3	28.2
								492	575	688	969
Stainless steels:											
AISI 302		8055	480	15.1	3.91			17.3	20.0	22.8	25.4
								512	559	585	606
AISI 304	1670	7900	477	14.9	3.95	9.2	12.6	16.6	19.8	22.6	25.4
						272	402	515	557	582	611
AISI 316		8238	468	13.4	3.48			15.2	18.3	21.3	24.2
								504	550	576	602
AISI 347		7978	480	14.2	3.71			15.8	18.9	21.9	24.7
								513	559	585	606
Lead	601	11,340	129	35.3	24.1	39.7	36.7	34.0	31.4		
						118	125	132	142		
Magnesium	923	1740	1024	156	87.6	169	159	153	149	146	
						649	934	1074	1170	1267	
Molybdenum	2894	10,240	251	138	53.7	179	143	134	126	118	112
						141	224	261	275	285	295
Nickel:											
Pure	1728	8900	444	90.7	23.0	164	107	80.2	65.6	67.6	71.8
						232	383	485	592	530	562
Nichrome (80% Ni, 20% Cr)	1672	8400	420	12	3.4			14	16	21	
								480	525	545	
Inconel X-750 (73% Ni, 15% Cr, 6.7% Fe)	1665	8510	439	11.7	3.1	8.7	10.3	13.5	17.0	20.5	24.0
						—	372	473	510	546	626
Niobium	2741	8570	265	53.7	23.6	55.2	52.6	55.2	58.2	61.3	64.4
						188	249	274	283	292	301
Palladium	1827	12,020	244	71.8	24.5	76.5	71.6	73.6	79.7	86.9	94.2
						168	227	251	261	271	281
Platinum:											
Pure	2045	21,450	133	71.6	25.1	77.5	72.6	71.8	73.2	75.6	78.7
						100	125	136	141	146	152
Alloy 60Pt–40Rh (60% Pt, 40% Rh)	1800	16,630	162	47	17.4			52	59	65	69
								—	—	—	—

Properties of solid metals *(Concluded)*

Composition	Melting Point, K	Properties at 300 K: ρ kg/m³	C_p J/kg · K	k W/m · K	$\alpha \times 10^6$ m²/s	Properties at Various Temperatures (K), k(W/m · K)/C_p(J/kg · K): 100	200	400	600	800	1000
Rhenium	3453	21,100	136	47.9	16.7	58.9	51.0	46.1	44.2	44.1	44.6
						97	127	139	145	151	156
Rhodium	2236	12,450	243	150	49.6	186	154	146	136	127	121
						147	220	253	274	293	311
Silicon	1685	2330	712	148	89.2	884	264	98.9	61.9	42.4	31
						259	556	790	867	913	946
Silver	1235	10,500	235	429	174	444	430	425	412	396	379
						187	225	239	250	262	277
Tantalum	3269	16,600	140	57.5	24.7	59.2	57.5	57.8	58.6	59.4	60.2
						110	133	144	146	149	152
Thorium	2023	11,700	118	54.0	39.1	59.8	54.6	54.5	55.8	56.9	56.9
						99	112	124	134	145	156
Tin	505	7310	227	66.6	40.1	85.2	73.3	62.2			
						188	215	243			
Titanium	1953	4500	522	21.9	9.32	30.5	24.5	20.4	19.4	19.7	20.7
						300	465	551	591	633	675
Tungsten	3660	19,300	132	174	68.3	208	186	159	137	125	118
						87	122	137	142	146	148
Uranium	1406	19,070	116	27.6	12.5	21.7	25.1	29.6	34.0	38.8	43.9
						94	108	125	146	176	180
Vanadium	2192	6100	489	30.7	10.3	35.8	31.3	31.3	33.3	35.7	38.2
						258	430	515	540	563	597
Zinc	693	7140	389	116	41.8	117	118	111	103		
						297	367	402	436		
Zirconium	2125	6570	278	22.7	12.4	33.2	25.2	21.6	20.7	21.6	23.7
						206	264	300	332	342	362

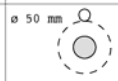
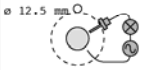
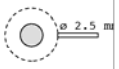
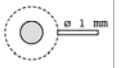
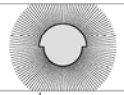
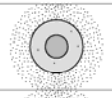
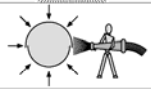
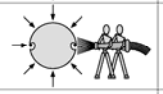
Boiling- and freezing-point properties

Substance	Boiling Data at 1 atm: Normal Boiling Point, °C	Boiling Data at 1 atm: Latent Heat of Vaporization, h_{fg} kJ/kg	Freezing Data: Freezing Point, °C	Freezing Data: Latent Heat of Fusion, h_{if} kJ/kg	Liquid Properties: Temp., °C	Liquid Properties: Density, ρ kg/m³	Liquid Properties: Specific Heat, C_p kJ/kg · °C
Ammonia	−33.3	1357	−77.7	322.4	−33.3	682	4.43
					−20	665	4.51
					0	639	4.62
					25	603	4.78
Argon	−185.9	161.6	−189.3	28	−185.6	1394	1.14
Benzene	80.2	394	5.5	126	20	879	1.72
Brine (20% sodium chloride by mass)	103.9	—	−17.4	—	20	1150	3.11
n-Butane	−0.5	385.2	−138.5	80.3	−0.5	601	2.31
Carbon dioxide	−78.4*	230.5 (at 0°C)	−56.6		0	298	0.59
Ethanol	78.2	838.3	−114.2	109	25	783	2.46
Ethyl alcohol	78.6	855	−156	108	20	789	2.84
Ethylene glycol	198.1	800.1	−10.8	181.1	20	1109	2.84
Glycerine	179.9	974	18.9	200.6	20	1261	2.32
Helium	−268.9	22.8	—	—	−268.9	146.2	22.8
Hydrogen	−252.8	445.7	−259.2	59.5	−252.8	70.7	10.0
Isobutane	−11.7	367.1	−160	105.7	−11.7	593.8	2.28
Kerosene	204–293	251	−24.9	—	20	820	2.00
Mercury	356.7	294.7	−38.9	11.4	25	13560	0.139
Methane	−161.5	510.4	−182.2	58.4	−161.5	423	3.49
					−100	301	5.78
Methanol	64.5	1100	−97.7	99.2	25	787	2.55
Nitrogen	−195.8	198.6	−210	25.3	−195.8	809	2.06
					−160	596	2.97
Octane	124.8	306.3	−57.5	180.7	20	703	2.10
Oil (light)					25	910	1.80
Oxygen	−183	212.7	−218.8	13.7	−183	1141	1.71
Petroleum	—	230–384			20	640	2.0
Propane	−42.1	427.8	−187.7	80.0	−42.1	581	2.25
					0	529	2.51
					50	449	3.12
Refrigerant-134a	−26.1	216.8	−96.6	—	−50	1443	1.23
					−26.1	1374	1.27
					0	1295	1.34
					25	1207	1.43
Water	100	2257	0.0	333.7	0	1000	4.22
					25	997	4.18
					50	988	4.18
					75	975	4.19
					100	958	4.22

Degrees of protection

IP codes according to IEC 60529 standard

Degrees of protection are identified by IP followed by two numbers followed by an optional letter, as defined in the table to follow.

FIRST NUMBER PROTECTION AGAINST SOLID BODY PENETRATION			SECOND NUMBER PROTECTION AGAINST LIQUID PENETRATION			ADDITIONAL LETTER[1]	DEGREE OF PROTECTION
IP	Tests		IP	Tests			BRIEF DESCRIPTION
0		No protection	0		No protection		
1		Protected against solid bodies greater than 50 mm	1		Protected against water drops falling vertically (condensation)	A	Protected against access with back of hand
2[1]		Protected against solid bodies greater than 12 mm	2		Protected against water drops falling up to 15° from the vertical	B	Protected against access with finger
3		Protected against solid bodies greater than 2.5 mm	3		Protected against water showers up to 60° from the vertical	C	Protected against access with tool
4		Protected against solid bodies greater than 1 mm	4		Protected against water splashes from any direction	D	Protected against access with wire
5		Protected against dust (excluding damaging deposits)	5		Protected against water jets from any hosed direction		
6		Total protection against dust	6		Protected against water splashes comparable to heavy seas		
The first two numbers are defined by NF EN 60 529, IEC 529 and DIN 40 050			7		Protected against total immersion		

IEC 947 and IEC 947-3 Standards

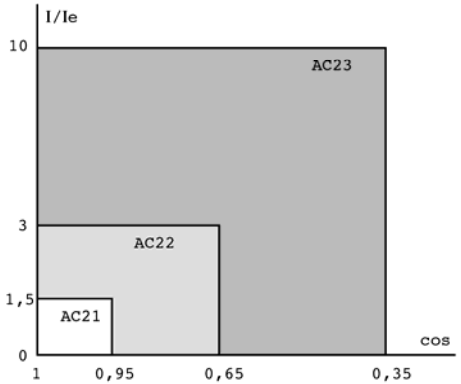
Selecting contactors according to IEC 947-3 standard

UTILIZATION	CATEGORY	USE	APPLICATION
AC AC20	DC DC20	No-load making and breaking	Disconnector (device without on-load making and breaking capacity)
AC21	DC21	Resistive loads including moderate overloads.	Switches at installation head or for resistive circuits (heating, lighting, except discharge lamps, etc.).
AC22	DC22	Inductive and resistive mixed loads including moderate overloads.	Switches in secondary circuits or reactive circuits (capacitor banks, discharge lamps, shunt motors, etc.).
AC23	DC23	Loads made of motors or other highly inductive loads.	Switches feeding one or several motors or inductive circuits (electric carriers, brake magnet, series motor, etc.).

Breaking and making capacities

Unlike circuit breakers, where these criteria indicate tripping or short-circuit making characteristics and perhaps requiring device replacement, switch making and breaking capacities correspond to utilization category maximum performance values.
In these uses, the switch must still maintain its characteristics, in particular its resistance to leakage current and temperature rise.

	MAKING		BREAKING		N° OF OPERATING CYCLES
	I/I_e	cos φ	I/I_e	cos φ	
AC 21	1.5	0.95	1.5	0.95	5
AC 22	3	0.65	3	0.65	5
AC 23 $I_e \leq 100$ A	10	0.45	8	0.45	5
$I_e > 100$ A	10	0.35	8	0.35	3
		L/R (ms)		L/R (ms)	
DC 21	1.5	1	1.5	1	5
DC 22	4	2.5	4	2.5	5
DC 23	4	15	4	15	5



Making and breaking capacities

Electrical and mechanical endurance

This standard establishes the minimum number of electrical (full load) and mechanical (no-load) operating cycles that must be performed by devices. These characteristics also specify the device's theoretical lifespan during which it must maintain its characteristics, particularly resistance to leakage current and temperature rise.
This performance is linked to the device's use and rating. According to anticipated use, two additional application categories are offered:

- category A: frequent operations (in close proximity to the load)
- category B: infrequent operations (at installation head or wiring system).

I_e (A)	≤ 100	≤ 315	≤ 630	≤ 2500	> 2500
N° CYCLES/HOUR	120	120	60	20	10
N° OF OPERATIONS IN CAT. A					
without current	8500	7000	4000	2500	1500
with current	1500	1000	1000	500	500
Total	10000	8000	5000	3000	2000
N° OF OPERATIONS IN CAT. B					
without current	1700	1400	800	500	300
with current	300	200	200	100	100
Total	2000	1600	1000	600	400

Definitions

Conventional thermal current (I_{th}):
Value of the current the disconnect switch can withstand with poles in closed position, in free air for an eight hour duty, without the temperature rise of its various parts exceeding the limits specified by the standards.

Rated insulation voltage (U_i):
Voltage value which designates the unit and to which dielectric tests, clearance and creepage distances are referred.

Rated impulse withstand voltage (U_{imp}):
Peak value of an impulse voltage of prescribed form and polarity which the equipment is capable of withstanding without failure under specified conditions of test and to which the values of the clearances are referred.

Rated operating current (I_e):
Current value determined by endurance tests (both mechanical and electrical) and by making and breaking capacity tests.

Short circuit characteristics

- short-time withstand current (I_{cw}): Allowable rms current for 1 second.
- short circuit making capacity (I_{cm}): peak current value which the device can withstand when closed on a short-circuit.
- conditional short circuit current: the rms current the switch can withstand when associated with a protection device limiting both the current and short circuit duration.
- dynamic withstand: peak current the device can withstand in a closed position.

The characteristic established by this standard is the short-time withstand current (I_{cw}) from which minimal dynamic withstand is deduced. This essential withstand value corresponds to what the switch can stand without welding.

Glossary of terms

Glossary of Wafer Processing terminology

Alloying:- The process of forming a low-resistance contact between the aluminium metal and silicon substrate on a metallised semiconductor wafer.
Amorphous Si, a-Si:- non-crystalline thin-film silicon; features no long-range crystallographic order; inferior electrical characteristics as compared to single-crystal and poly Si but cheaper and easier to manufacture; used primarily to fabricate solar cells.
Angstrom, Å:- unit of length commonly used in semiconductor industry, though not recognised as a standard international unit; 1 Å = 10^{-10} m = 10^{-4} micrometer = 0.1 nm.
Annealing:- The process of combining hydrogen with uncommitted atoms at or near the silicon-silicon dioxide interface on a metallised semiconductor wafer.
Ashing:- The process of removal (by volatilization) of organic materials (e.g. photoresist) from the wafer surface using strongly oxidizing ambient; e.g. oxygen plasma ashing.
Backlapping:- The process of mechanically thinning the backside of a finished semiconductor wafer.
Backside metallisation:- The process of depositing a metal layer on the backside of a finished wafer.
Bandgap, energy gap E_g:- forbidden energy levels separating the valence and conduction bands. Electrons are allowed to have energies at these levels.
Barrier metal:- thin layer of metal, e.g. TiN, sandwiched between other metal and semiconductor (or insulator) to prevent potentially harmful interactions between these two, e.g. spiking.
Boat:- 1. a device made of high purity temperature resistant materials such as fused silica, quartz, poly Si, or SiC. designed to hold many semiconductor wafers during thermal or other processes;
2. device designed to simultaneously contain source material during evaporation while at the same time heating the source to its melting point; made of highly conductive, temperature resistant material through which current is passed.
Chip:- The final integrated semiconductor circuit.
Conduction band:- the upper energy band in a semiconductor separated from the valence band by the energy gap; The conduction band is not completely filled with electrons.
Constant-source diffusion:- also know as unlimited-source diffusion or predeposition; concentration of diffusant (dopant) on the surface of the wafer remains constant during the diffusion process, i.e. while some dopant atoms diffuse into the substrate additional dopant atoms are continuously supplied to the surface of the wafer.
Crystal pulling:- The process of forming a crystal ingot; a seed crystal of silicon is attached to a rod and "pulled" out of a silicon melt to form an ingot.
Czochralski Crystal Growth, CZ:- process utilizing crystal pulling to obtain single-crystal solids; the most common method for obtaining large diameter semiconductor wafers (300mm Si wafers); desired conductivity type and doping level is accomplished by adding dopants to molten material. Wafers used in high-end Si microelectronics are almost uniquely CZ grown.
Czochralski method:- The crystal pulling method used to form crystal ingots.
Chemical vapour deposition:- The process of applying a thin film to a substrate using a controlled chemical reaction.
Deposition:- A general term used to describe the addition of material layers on a semiconductor wafer.
Die:- An individual device or chip cut from a semiconductor wafer.
Diffusion:- A doping process; a high-temperature furnace is used to diffuse an applied layer of dopant into the wafer surface.
Diffusion coefficient, D:- determines rate with which element moves in a given solid by diffusion; depends strongly on temperature; expressed in cm^2/s; varies between elements by orders of magnitude, e.g. in the case of diffusion in silicon diffusion coefficient for gold, Au, is in the range of 10^{-3} cm^2/s (fast diffusant) while for Sb is in the range of 10^{-17} cm^2/s.
Dopant:- element introduced intentionally into a semiconductor to establish either p-type or n-type conductivity; Common dopants in silicon are: Boron (p-type) and phosphorous, arsenic, and antimony (n-type).
Doping:- The process of introducing impurity elements (dopants) into a semiconductor wafer to form regions of differing electrical conductivity. The two most common doping processes are diffusion and ion implantation.
Drive in:- high temperature (>800°C) operation performed on semiconductor wafer in an inert ambient; causes motion of dopant atoms in semiconductor in the direction of concentration gradient (diffusion); used to drive dopant atoms deeper into semiconductor.

Electron beam (e-beam) *evaporation*:- source material is evaporated as a result of highly localized heating by bombardment with high energy electrons; the electron beam is spatially confined and accelerated by electrostatic interactions. The direction and cross-section of the beam can be precisely controlled and rapidly altered to scan the target; evaporated material is very pure; bombardment of metal with electrons is accompanied by generation of low intensity X-rays which may create defects in the oxide present on the surface of the substrate; typically, an anneal is needed to eliminate those defects.

Epi Layer:- The term epitaxial comes from the Greek word meaning 'arranged upon.' In semiconductor technology, it refers to the single crystalline structure of the film. The structure comes about when silicon atoms are deposited on a bare silicon wafer in a CVD reactor. When the chemical reactants are controlled and the system parameters are set correctly, the depositing atoms arrive at the wafer surface with sufficient energy to move around on the surface and orient themselves to the crystal arrangement of the wafer atoms. Thus an epitaxial film deposited on a <111>-oriented wafer will take on a <111> orientation.

Epitaxial layer:- layer grown in the course of epitaxy.

Epitaxy:- process by which a thin *epitaxial* layer of single-crystal material is deposited on single-crystal substrate; epitaxial growth occurs in such way that the crystallographic structure of the substrate is reproduced in the growing material; also crystalline defects of the substrate are reproduced in the growing material. Although crystallographic structure of the substrate is reproduced, doping levels and the conductivity type of a epitaxial layer is controlled independently of the substrate; e.g. the epitaxial layer can be made more pure chemically than the substrate.

Etching:- The process of removing silicon dioxide layers, accomplished by "wet etching" with chemicals or by "dry etching" with ionized gases.

Evaporation:- common method used to deposit thin film materials; material to be deposited is heated in vacuum (10^{-6} - 10^{-7} Torr range) until it melts and starts evaporating; this vapour condenses on a cooler substrate inside the evaporation chamber forming smooth and uniform thin films; not suitable for high melting point materials; PVD method of thin film formation.

External, extrinsic gettering:- process in which gettering of contaminants and defects in a semiconductor wafer is accomplished by stressing its back surface (by inducing damage or depositing material featuring different than semiconductor thermal expansion coefficient) and then thermally treating the wafer; contaminants and/or defects are relocated toward back surface and away from the front surface where semiconductor devices can be formed.

Fick's law:- describe diffusion in solids; 1st and 2nd Fick's law; 1st Fick's law describes motion by diffusion of an element in the solid in the direction of the concentration gradient; 2nd Fick's law determines changes of concentration gradient in the course of diffusion (function of time and diffusion coefficient).

Filament evaporation:- thermal evaporation; source material is contacted to the filament (a refractory metal) and melted by high current flowing through the filament; alternatively, a "boat" which contains material to be evaporated may be made out of refractory metal;

Float-zone Crystal Growth, FZ:- method used to form single crystal semiconductor substrates (alternative to CZ); polycrystalline material is converted into single-crystal by locally melting the plane where a single crystal seed is contacting the polycrystalline material; used to make very pure, high resistance Si wafers; does not allow as large wafers (< 200mm) as CZ does; radial distribution of dopant in FZ wafer is not as uniform as in CZ wafer.

Gettering:- process which moves contaminants and/or defects in a semiconductor away from its top surface into its bulk and traps them there, creating a denuded zone.

HMDS:- Hexamethyldisilizane; improves adhesion of photoresist to the surface of a wafer; especially designed for adhesion of photoresist to SiO_2; deposited on wafer surface immediately prior to deposition of resist.

Hydrogenated a-Si:- amorphous silicon (a-Si) containing substantial quantities of hydrogen; hydrogen passivated Si dangling bonds and results in substantially improved electrical properties of A-Si

Ingot:- circular piece of single-crystal semiconductor material resulting from a crystal growth process; an ingot is ready to be shaped and sliced into wafers used to manufacture semiconductor devices.

Intrinsic gettering:- process in which gettering of contaminants and/or defects in a semiconductor is accomplished (without any physical interactions with the wafer) by a series of heat treatments.

Ion implantation:- A doping process; the dopant material is ionized and magnetically accelerated to strike the wafer surface, thereby embedding the dopant into the substrate.

Lapping:- The process of mechanically grinding the surface of a sliced wafer.

Lead frame:- The die attachment surface and lead attachment points that a die or chip is attached to prior to wire bonding and packaging.

limited-source diffusion:- also known as drive-in; concentration of diffusant (dopant) on the surface decreases during the diffusion process, i.e. while some dopant atoms diffuse into the substrate no new dopant atoms are supplied to the surface of the wafer.

Metallization:- formation of metal contacts and interconnects in the manufacturing of semiconductor devices.

Metal-semiconductor contact:- key component of any semiconductor device; depending on materials involved in the contact its properties can differ drastically; ohmic contact (linear, symmetric current-voltage characteristic)in the case when work function of metal matches work function of semiconductor (no potential barrier at the interface); rectifying contact(non-linear, highly asymmetric, diode-like current-voltage characteristic) in the case when work function of metal differs from the work function of semiconductor (potential barrier at the interface)- commonly referred to as a Schottky diode.

Minority carriers:- one of two carrier types (electrons of holes)whose equilibrium concentration is lower than that of the other type; holes in n-type semiconductors, electrons in p-type semiconductors.

N-type semiconductor:- semiconductor in which the concentration of electrons is much higher than the concentration of holes (p>>n); electrons are majority carriers and dominate conductivity.

Ohmic contact:- metal-semiconductor contact with very low resistance independent of applied voltage (may be represented by constant resistance); to form an "ohmic" contact metal and semiconductor must be selected such that there is no potential barrier formed at the interface (or potential barrier is so thin that charge carriers can readily tunnel through it).

Oxidation:- The process of oxidizing the wafer surface to form a thin layer of silicon dioxide.

Passivation:- The process of applying a final passivating or protective layer of either silicon nitride or silicon dioxide to a wafer.

Photolithography:- The process of creating patterns on a silicon substrate. The main steps of the process include photoresist application, mask alignment, photoexposure, developing, and etching the portions of the substrate that are unprotected by the resist.

Photomask:- A mask that delineates the pattern applied to a substrate during photolithography.

Photoresist:- A photo-sensitive material used in photolithography to transfer pattern from the mask onto the wafer; a liquid deposited on the surface of the wafer as a thin film then solidified by low temperature anneal; in the areas in which photoresist can be reached by UV radiation photochemical reactions change its properties, specifically, solubility in the developer; two types of photoresist:- positive and negative.

Polishing:- process applied to either reduce roughness of the wafer surface or to remove excess material from the surface; typically polishing is a mechanical-chemical process using a chemically reactive slurry.

Polycrystalline silicon:- An amorphous form of silicon with randomly oriented crystals, used to produce silicon ingots.

Polycrystalline material, poly:- many (often) small single-crystal regions are randomly connected to form a solid; size of regions varies depending on the material and the method of its formation. Heavily doped poly Si is commonly used as a gate contact in silicon MOS and CMOS devices.

Physical Vapour Deposition, PVD:- deposition of thin film occurs through physical transfer of material (e.g. thermal evaporation and sputtering)from the source to the substrate; chemical composition of deposited material is not altered in the process.

P-type semiconductor:- semiconductor in which the concentration of holes is much higher than the concentration of electrons (n>>p); holes are majority carriers and dominate conductivity.

Quartz:- single-crystal SiO_2.

Quartzite:- Silica sand used as a raw material to produce metallurgical grade silicon.

Reactive ion etching RIE:-, RIE variation of plasma etching that uses physical sputtering and chemically reactive species in which during etching semiconductor wafer is placed on the RF powered electrode; wafer takes on potential which accelerates etching species extracted from plasma toward the etched surface; chemical etching reaction is preferentially taking place in the direction normal to the surface, i.e. etching is more anisotropic than in plasma etching but is less selective; leaves etched surface damaged; the most common etching mode in semiconductor manufacturing, also used to remove metal layers.

Rectifying contact:- metal-semiconductor contact displaying asymmetric current-voltage characteristics, i.e. allowing high current to flow across under the forward bias condition and blocking current off under the reverse bias; this behaviour is controlled by the bias voltage dependent changes of the potential barrier height in the contact region.

Seed crystal:- single crystal material used in crystal growing to set a pattern for the growth of material in which this pattern is reproduced.

Semiconductor:- solid-state material in which (unlike in metals and insulators) (1) large changes in electrical conductivity can be effected by adding very small amounts of impurity elements known as dopants, (2) electrical conductivity can be controlled by both negatively charged electrons and positively charged holes and (3) electrical conductivity is sensitive to temperature, illumination, and magnetic field.

Silicon:- A semi-metallic element used to create a wafer.

Silicon dioxide, SiO_2:- silica; native oxide of silicon; the most common insulator in semiconductor device technology; high quality films are obtained by thermal oxidation of silicon; thermal SiO_2 forms smooth, low-defect interface with Si; can be also readily deposited by CVD; Key parameters: energy gap E_g ~ 8eV; dielectric strength 5-15 x 10^6 V/cm; dielectric constant k = 3.9; density 2.3

g/cm^3; refractive index n =1.46; melting point ~ 1700°C; prone to contamination with alkali ions and sensitive to high energy radiation (i.e. X-rays); single crystal SiO_2 is known as quartz.

Silicon Nitride, Si_3N_4:- dielectric material with energy gap = 5 eV and density ~3.0 g/cm^3; excellent mask against oxidation of Si and KOH; properties depend on deposition method: dielectric strength ~10^7 V/cm, dielectric constant k ~6-7, bulk resistivity 10^{15}-10^{17} ohm-cm; deposited by CDV.

Silyation:- The process of introducing silicon atoms into the surface of an organic photoresist in order to harden the photoresist.

Single-crystal:- crystalline solid in which atoms are arranged following specific pattern throughout the entire piece of material; in general, single crystal material features superior electronic and photonic properties as compared to polycrystalline and amorphous materials, but is more difficult to fabricate; all high-end semiconductor electronic and photonic materials are fabricated using single-crystal substrates.

Slice orientation: - the angle between the surface of a slice and the growth plane of the crystal. The most common slice orientations are (100), (111) and (110).

Slicing:- term refers to the process of cutting of the single-crystal ingot into wafers; high precision diamond blades are used.

Slurry:- a liquid containing suspended abrasive component; used for lapping, polishing and grinding of solid surfaces; can be chemically active; key element of CMP processes.

Spiking:- uncontrolled penetration of semiconductor substrate by contact metal; problem with Al in contact with silicon; may short ultra-shallow p-n junction underneath the contact.

Sputtering, sputter deposition:- bombardment of a solid (target) by high energy chemically inert ions (e.g. Ar+); causes ejection of atoms from the target which are then re-deposited on the surface of a substrate purposely located in the vicinity of the target; common method of Physical Vapour Deposition of metals and oxides.

Sputtering target:- source material during sputter deposition processes; typically a disc inside the vacuum chamber which is exposed to bombarding ions, knocking source atoms loose and onto samples.

Sputter yield:- efficiency of the sputtering process (differs for different materials).

Surface damage:- process related disruption of the crystallographic order at the surface of single-crystal semiconductor substrates; typically caused by surface interactions with high energy ions during dry etching and ion implantation.

Staebler - Wronski effect:- degradation of electrical output of hydrogenated amorphous silicon solar cells as a result of prolonged illumination.

Stripping:- process of material removal from the wafer surface; typically implies that removal is not carried out for the pattering purpose, e.g. resist stripping in which case entire resist is removed following lithography and etching.

Target:- source material used during evaporation or deposition; In sputtering, typically in the form of high purity disc. In e-Beam evaporation, typically in the form of a crucible. In thermal evaporation, the source material is typically held in a boat which is heated resistively.

Thermal oxidation, thermal oxide:- growth of oxide on the substrate through oxidation of the surface at elevated temperature; thermal oxidation of silicon results in a very high quality oxide, SiO2; most other semiconductors do not form device quality thermal oxide, hence, "thermal oxidation" is almost synonymous with "thermal oxidation of silicon".

Valence band:- the lower energy band in a semiconductor that is completely filled with electrons at 0 K; electrons cannot conduct in valence band.

Volume defect:- voids and/or local regions featuring different phase (e.g. precipitates or amorphous phase) in crystalline materials.

Wafer:- thin (thickness depends on wafer diameter, but is typically less than 1 mm), circular slice of single-crystal semiconductor material cut from the ingot of single crystal semiconductor; used in manufacturing of semiconductor devices and integrated circuits; wafer diameter may range from 25 mm to 300 mm.

Wafer bonding:- process in which two semiconductor wafers are bonded to form a single substrate; commonly applied to form SOI substrates; bonding of wafers of different materials, e.g. GaAs on Si, or SiC on Si; is more difficult than bonding of similar materials.

Wafer fabrication:- process in which single crystal semiconductor ingot is fabricated and transformed by cutting, grinding, polishing, and cleaning into a circular wafer with desired diameter and physical properties.

Wafer flat:- flat area on the perimeter of the wafer; location and number of wafer flats contains information on crystal orientation of the wafer and the dopant type (n-type or p-type).

Work function difference:- defines characteristics of contact between two materials featuring different work function; for conductor-semiconductor contact w.f.d. determines height of potential barrier in the contact plane, and hence, determines whether contact is ohmic or rectifying.

Glossary of Fuselink terminology (Fuseology)

'A' Fuselink (formerly Back-Up Fuselink):- A current limiting fuselink capable of breaking under specified conditions all currents between the lowest current indicated on its operating time-current characteristic and its rated breaking capacity.

Ambient Temperature:- The temperature of the surrounding medium which comes in contact with the fuse. The medium is usually air. Fuse current carrying capacity tests are performed at 25°C and are affected by changes in ambient temperature. A fuse runs hotter as the normal operating current approaches or exceeds its current rating. At room temperature, 25°C, a fuse should last indefinitely if operated at no more than 75% of fuse ampere rating. The fuse ambient temperature may be significantly higher because it is enclosed or mounted near other heat producing components, such as resistors, transformers, etc.

Ampacity:- The current a conductor can carry continuously without exceeding its temperature rating. Ampacity is a function of cable size, insulation type and the conditions of use.

Ampere Rating:- Same as Current Rating or the current carrying capacity of a fuse. The continuous current carrying capability of a fuse under defined conditions. When a fuse is subjected to a current above its ampere rating, it will open the circuit after a predetermined period of time. Continuous load current should not exceed 75% of fuse ampere rating (at 25°C ambient) except fuses that may be specifically loaded to 100% of their ampere rating.

Ampere Squared Seconds, I^2t:- A measure of thermal (heat) energy associated with current flow during fuse clearing. I^2t is equal to I^2_{RMS} x t, where t is the duration of current flow in seconds. It can be expressed as melting I^2t, arcing I^2t or their sum as Clearing I^2t. Clearing I^2t is the total I^2t passed by a fuse as the fuse clears a fault, with t being equal to the time elapsed from the initiation of the fault to the instant the fault has been cleared. Melting I^2t is the minimum I^2t required to melt the fuse element. 'I' is the effective let-through RMS current, which is squared, and 't' is the time of opening, in seconds.

Arc Quenching Time:- As part of the Operating Time it is the time between the arc starting and the final current zero. Depending on the Melting Time the Arc Quenching Time is typically just a few ms up to a couple of 100 ms.

Arcing Time:- The arcing time is the interval of time between the instant of the initiation of the arc and the instant of final arc extinction. That is the time from when the fuselink has melted until the over current is interrupted, or cleared.

Arc Voltage:- The highest fuse voltage during the Operating Time of the fuse.

Arcing withstand Time:- Longest time between separation of the melting element and the faultless interruption of the current through the fuse-switch. Typical values are above 100 ms.

Breaking Capacity:- The breaking capacity is the highest value (for ac the rms. value of the ac component) of prospective current that a fuselink is capable of breaking at a stated voltage under specified conditions of use and behaviour. The rating which defines the fuses ability to safely interrupt and clear short circuits. This rating is much greater than the ampere rating of a fuse. The highest current at rated voltage that an over current protective device is intended to interrupt under specified conditions. During a fault or short circuit condition, a fuse may receive an instantaneous overload current many times greater than its normal operating current. Safe operation requires that the fuse remain intact (no explosion or body rupture) and clear the circuit. Also known as interrupting rating or short circuit rating.

Breaking Range:- Breaking range is a range of prospective currents within which the breaking capacity of a fuselink is assured.

Clearing Time:- The total time from the beginning of the over current to the final opening of the circuit at rated voltage by an over current protective device. Clearing time is the total of the melting time and the arcing time.

Conventional Non-Fusing Current I_{nf}:- A value of current specified as that which the fuselink is capable of carrying for a specified time (conventional time) without melting. The conventional time relates to the thermal time constant of the fuselink and varies between one and four hours depending on the current rating.

Conventional Fusing Current I_f:- Current specified as that which causes operation of the fuselink within a specified time (conventional time). The conventional time relates to the thermal time constant of the fuselink and varies between one and four hours depending on the current rating.

Coordination:- The use of over-current protective devices which will isolate only that portion of an electrical system which has been overloaded or faulted.

Current-Limiting Fuselink:- A current-limiting fuse link limits the current to a substantially lower value than the peak value of the prospective current during and by its operation in a specified current range.

Current Limitation:- Fuse operation relating to short circuits only. When a fuse operates in its current limiting range, it will clear a short circuit in less than ½ cycle. Also, it will limit the instantaneous peak let-thru current to a value substantially less than that obtainable in the same circuit if that fuse were replaced with a solid conductor of the same impedance.

Current Rating:- The nominal amperage value of the fuse. It is established as a value of current which the fuse can carry, based on a controlled set of test conditions

Cut-Off Current:- The cut-off or let-through current is the maximum instantaneous value reached by the current during the breaking operation of a fuselink when it operates in such a manner as to prevent the current from reaching the otherwise attainable maximum. In case of a short-circuit, the maximum value of the short circuit current. This value is required for the analysis of the dynamic impact of the short-circuit current on the protected equipment.

Cut-off (current) characteristic:- The cut-off (current) characteristic or let-through current characteristic is a curve giving the cut-off current as a function of the prospective current, under specified operating conditions.

Derating:- Term for reducing influences on the Rated Breaking Current of the fuse. The Derating value is multiplied by the Rated Current then divided by the loading current. Typical influencing factors include high surrounding temperature, terminal cross section, installation volume, pulse load, shock load, and over-waves.

Discrimination:- Classification of relevant parameters (Time/Current-Characteristic; Integrals; Operating Times etc.) of two or more overload protection devices to each other. In the case of overloads, only the protection device should react. Sequential fuses with the same characteristic, are selected in the proportion 1:1.6. A fuse with a rated current of 100 A should be downstream of a fuse rated 160 A. For the short-circuit range the comparison of the melting integrals versus the Operating integral of the downstream fuse is important.

Dissipated Power:- When a current passes through a fuse link, a small amount of energy is dissipated due to the fuse links resistance.

Dual Element Fuse:- Often confused with time delay, dual element is a term describing fuse element construction. A fuse having two current responsive elements in series.

Element:- A calibrated conductor inside a fuse which melts when subjected to excessive current. The element is enclosed by the fuse body and may be surrounded by an arc-quenching medium such as silica sand. The element is sometimes referred to as a link.

Fast-Acting Fuse:- Fast-acting fuses have no intentional built in slow-blow and are used in circuits without transient inrush currents. Fast-acting fuses open quickly on overload and short-circuits. This type of fuse is not designed to withstand temporary overload currents.

Fault current:- A current resulting from a fault, a circuit condition in which the current flows through an abnormal, unintended path.

Fusing factor:- The fusing factor is the ratio, greater than unity, of the minimum fusing current to the fuse current rating.

Fuse:- A fuse is a device that by the fusing of one or more of its specially designed and proportioned components, opens the circuit in which it is inserted by breaking the current when this exceeds a given value for a sufficient time. An over-current protective device containing a calibrated current carrying member which melts and opens a circuit under specified over-current conditions. It is common practice to refer to a 'fuselink' as a 'fuse'.

Fuse Element:- Part of the Fuse-Link, which melts when the fuse operates. It consists of perforated metal stripes. The dimension of the perforation reflects the Characteristic and the Rated Current of the Fuse-Link. Depending on the Rated Current the Fuse-Links contain several paralleled Fuse Elements. Typical materials are copper and pure silver.

Fuse initiated opening time:- Time between separating of the melting elements and the faultless interruption of the failure current through the fuse. Typically between 30 and 100 ms.

Fuse Selection Guide:- The fuse must carry the normal circuit load current without nuisance openings. However, when an over-current occurs the fuse must interrupt the over-current, limit the energy let-through, and withstand the voltage across the fuse during arcing. To select a fuse the following must be considered:

Normal operating current (The current rating of a fuse is typically derated 25% for operation at 25°C to avoid nuisance blowing. For example, a fuse with a 10A current rating is not usually recommended for operation at more than 7.5A in a 25°C ambient.)
Overload current and time interval in which the fuse must open.
Application voltage (AC or DC Voltage).
Inrush currents, surge currents, pulses, start-up currents characteristics.
Ambient temperature.
Applicable standards agency requirements, such as UL, CSA, VDE.

Other considerations include: Reduce installation cost, ease of removal, mounting type/form factor, etc.

Fuse Type:- There are three basic types of fuses:

1. Slow Blow/Time Lag/Time Delay fuses
2. Fast acting fuses
3. Very fast acting fuses

A major type of Time Delay fuse is the dual-element fuse. This fuse consists of a short circuit strip, soldered joint and spring connection. During overload conditions, the soldered joint gets hot enough to melt and the spring shears the junction loose. Under short circuit conditions, the short circuit element operates to open the circuit. Slow-blow fuse allows temporary and harmless inrush currents to pass without opening, but is so designed to open on sustained overloads and short circuits. Slow-blow fuses are ideal for circuits with a transient surge or power-on inrush. These circuits include: motors, transformers, incandescent lamps and capacitate loads. This inrush may be many times the circuit's full load amperes. Slow-blow fuses allow close rating of the fuse without nuisance opening. Typically, Slow Blow fuses are rated between 125% to 150% of the circuit's full load amperes.

Fusing Current:- Value of fuse current which will be interrupted within a given time. Valid for general purpose fuse-links. Normally the testing current is about 1.6 times the Rated Current.

Gate:- Limiting values within which the characteristics, for example time-current characteristics, shall be contained.

High Speed Fuses:- Fuses with no intentional time-delay in the overload range and designed to open as quickly as possible in the short circuit range. Often used to protect solid-state devices.

Homogeneous Series of Fuselinks:- A series of fuselinks, within a given size.

I^2t (Joule Integral) :- See Joule integral.

I^2t (Ampere Squared Seconds):- A measure of the thermal energy associated with current flow. I^2t is equal to $I_{RMS}{}^2$ x t , where t is the duration of current flow in seconds.

Clearing I^2t is the total I^2t passed by a fuse as the fuse clears a fault, with t being equal to the time elapsed from the initiation of the fault to the instant the fault has been cleared.
Melting I^2t is the minimum I^2t required to melt the fuse element.

I^2t *Characteristic*:- A curve giving I^2t values (pre-arcing I^2t and/or operating I^2t) as a function of prospective current under specific operating conditions.

Interrupting Rating (Abbreviated IR):- Same as breaking capacity or short circuit rating. The maximum current a fuse can safely interrupt at rated voltage. Some special purpose fuses may also have a *Minimum Interrupting Rating*. This defines the minimum current that a fuse can safely interrupt. Safe operation requires that the fuse remain intact. Interrupting ratings may vary with fuse design and range from 35A AC for some 250V metric size (5 x 20mm) fuses up to 200kA AC for the 600V industrial fuses.

Joule integral:- The I^2t or Joule integral is a measure of the thermal stress or thermal energy let through by the fuse during short circuit interruption. It is the integral of the square of the current over a given time and is expressed in ampere square seconds.

Two values of I^2t are provided for MV-fuse links:

- Pre arcing or melting I^2t - for high short circuit currents - this is practically a constant.
- Operation I^2t – this varies with circuit conditions.

Let-through current:- The cut-off or let-through current is the maximum instantaneous value of current attained during the breaking operation of a MV-fuse link. This important when the MV-fuse link operates in such that the circuit prospective peak current is not reached.

Let-through current characteristic:- The cut-off (current) characteristic or let-through current characteristic is a curve giving the cut-off current as a function of the prospective current, under specific operating conditions.

Melting Current:- Current during an increase in prospective Short-Circuit Current, at which the Fuse Element melts. This current is usually lower than the Cut-off Current, because this normally increases during the Quenching Time.

Melting Integral:- Current Integral for the Melting time of the fuse. The Melting Integral depends on the size of the Melting Elements and is therefore independent of voltage. The minimum value is normally given, for analysing discrimination.

Melting time:- The amount of time required to melt the fuselink during a specified over current. The pre-arcing time or melting time is the interval of time between the beginning of a current large enough to cause a break in the fuse element and the instant when an arc is initiated. The Time/Current-Characteristic provides the virtual Melting Time for different current closing angles. Virtual Melting Time = Melting Integral / failure current.

Minimum Breaking Current:- Smallest failure current at which a back-up fuse can operate at its rated voltage. Values are often between 3 to 4 times Rated Current. The minimum breaking current is a minimum value of prospective current that a link is capable of breaking at a stated voltage under specified conditions.

Non fusing Current:- Defined value of current, at which (under certain circumstances) a fuse-link must not operate within a given time, Conventional Time. For a General Purpose Fuse, this value is normally 1.25 times Rated Current.

Operating time:- The operating time or total clearing time is the sum of the pre-arcing time and the arcing time. Also the summation of Melting Time and Arc Quenching Time of the Fuse. Over a Melting Time of 100ms the Operating Time can generally be equated with the Melting Time. For shorter Melting Times, the Operating Time can be more than double of the Melting Time. Below 5ms, the Operating Time should be calculated via the Operating Integral.

Operating Integral:- Current integral over the operating time of the fuse. Information is particularly valid for melting times less than 5ms, whence the fuse has operated with current limitation. Usually the datasheet value is the highest expect for the given reference voltage. Values at lower service voltage are calculated through the conversion diagram.

Overcurrent:- An over-current is a current exceeding the rated current, normal load current, conductor ampacity or equipment continuous current rating. An over-current can be an overload current, fault current or short circuit current.

Overcurrent Discrimination:- Co-ordination of the relevant characteristics of two or more over-current protective devices such that, on the occurrence of over-currents within specific limits, the device intended to operate within these limits does so, while the others do not.

Overload:- Classified as an overcurrent which exceeds the circuit normal full load current. The operation of conductors or equipment at a current level that will cause damage if allowed to persist. The current does not leave the normal current carrying path of the circuit, that is, it flows from the source, through the conductors, through the load, back through the conductors to the source.

Overload current:- A current resulting from an overload occurring in a normally working electrically circuit, for example an overloaded motor. If there is no protective device operating in a limited time of several seconds, the electrical system would overheat and cable isolation, etc. would melt and cause damage.

Overload Curve of an Fuselink:- A curve showing the time for which a fuselink shall be able to carry the current without deterioration.

Peak Let-Thru Current, I_P:- The instantaneous value of peak current let-thru by a current limiting fuse, when clearing a fault current of specified magnitude in its current limiting range.

Power Dissipation:- Power dissipation is the power released in a fuse link carrying a stated current under specified conditions of use and behaviour, usually including a constant rms. current until steady temperature conditions are reached.

Pre-Arcing Time:- The pre-arcing time or melting time is the interval between the beginning of a current large enough to cause a break in the fuse element and the instant when an arc is initiated.

Prospective Current of a Circuit (with respect to the fuse):- The prospective current is the current that would flow in a circuit if a fuse situated therein were replaced by a link of negligible impedance. The prospective current is the quantity to which the breaking capacity and characteristics of the fuse are normally referred, for example, I^2t and cut-off current characteristic.

Prospective Short Circuit Current:- The prospective short circuit current is the value of the current that would flow if there was no protection in the circuit. The lower the power factor of the installation, the higher the peak value of this destructive current.

Rated Breaking Capacity (Low/High Voltage Fuses):- Capacity of a fuse to operate between the lowest and the Rated Breaking Current, which is a certified, effective value. Normally fuses can operate at higher currents. Typical values for Low-Voltage fuses are: 100, 120, 200 or 300 kA and for High-Voltage fuses 20kA to 63 kA. For miniature fuses, it is the current at which a fuse can operate normally under specified conditions at a fixed Voltage.

Rated Current of a Fuselink I_n:- A value of current that the fuselink can carry continuously without deteriorating or without operating under specified standardised conditions, including in free air with a defined cable cross-sections. Often the Rated Current has to be reduced by the Derating-value.

Rated Frequency:- The rated frequency is the frequency for which the fuse link has been designed and to which the values of the other characteristics correspond. Standard values of rated frequency are 50 Hz and 60 Hz.

Rated Insulation Level:- The rated insulation level (of a MV-fuse base) is the voltage values (both power-frequency and impulse) which characterize the insulation of the fuse base with regards to its capability of withstanding the dielectric stresses.

Rated Values:- Rated values, usually stated for HV-fuse links, are

- voltage
- current
- breaking capacity
- frequency

All given for specified operating conditions.

Rated Voltage:- The Rated voltage, V_n, is the maximum value of voltage at which an fuse link can be used, and safely interrupt an over-current. This rated voltage must be higher or equal to the highest voltage of the system in which the fuse link is installed. Effective value of the Operating Voltage of a fuse; normally an alternating voltage, at a frequency between 42 to 62 Hz.

Recovery Voltage:- The recovery voltage is the voltage which appears across the terminals of a fuse after the breaking of the current. This voltage is considered in two successive intervals of time, one during which a transient voltage exists, followed by a second during which the power frequency or the steady-state recovery voltage alone exists.

Selectivity:- A main fuse and a branch fuse are said to be selective if the branch fuse will clear all over-current conditions before the main fuse opens. Selectivity is desirable because it limits outage to that portion of the circuit which has been overloaded or faulted. Also called *selective coordination*.

Short Circuit:- A high value of over-current resulting from a fault of negligible impedance between conductors with difference potential and under normal operating conditions. A short circuit current can be many hundreds or even thousands of times larger than the normal load current.

Striker:- A striker is a mechanical device forming part of a fuselink which, when the fuse operates, releases the energy required to cause operation of other apparatus or indicators or to provide interlocking.

Switching voltage:- The switching voltage is the maximum instantaneous value of voltage, which appears across the terminals of a fuse during its operating time. Under short circuit conditions this will often exceed the peak system voltage for a period of time. It is typically two to three times the Rated Voltage.

Time-current characteristic:- The time-current characteristic is a curve giving the time, for example pre-arcing time (or operating time), as a function of the prospective current and respectively short-circuit currents. under specified operating conditions. The time-current curve is used to achieve co-ordination with the other fuses or devices in the same installation.

Time/Current-Curve:- Curve for calculating the Melting Time of the fuse at designed overload and respectively short-circuit current. The opening time is considered nominal. Time/Current-Curves refer to a temperature between 20°C and 30°C, are given for times between 4ms and 10000s, and are drawn as a family of curves on a double logarithmic grid (opening time in seconds for the fuse for a range of over-currents).

Time Delay Fuse:- A fuse which will carry an over-current of a specified magnitude for a minimum specified (in standards) time without opening.

Take-Over Current:- at operating the Striker Pin: Value of the symmetrical three phase current at which the breaking varies between the fuse and the switch. Below this value the current will be interrupted in the first quenching pole through a fuse and the current in both other poles through the switch. Above the value, the current is interrupted in all 3 poles only through the fuses. Depending on the Rated Voltage of the switch, values are between 600A and 3000A.

Threshold Current:- The minimum available fault current at which a fuse is current limiting.

Total clearing time:- The operating time or total clearing time is the sum of the pre arcing time and the arcing time.

Very Fast-Acting Fuses:- Very fast-acting (Current-Limiting) fuses will limit both the magnitude and duration of current flow under short circuit conditions. Because of their high current limiting ability, these fuses are frequently used to protect semiconductor circuits.

Virtual time:- The virtual time is the value of Joule integral divided by the square of the prospective current value. Usually stated for a MV-fuse link, are the values of pre-arcing time and of operating time.

Virtual Melting time:- Standardised value of melting time, which considers currents of types AC or DC and the different current curves and switching angles. The Melting Time in the Time/Current-Characteristics is generally given by the Virtual Melting Time. The value is calculated by the Melting integral of the Rated Current.

Voltage Rating:- The maximum voltage at which a fuse is designed to operate. The maximum open circuit voltage in which a fuse can be used, yet safely interrupt an overcurrent. Exceeding the voltage rating of a fuse impairs its ability to clear an overload or short circuit safely. Voltage ratings are assumed to be for AC, unless specifically labelled as DC.

Glossary of Relay terminology

Arc:- An electric discharge between mating relay contacts when an energized circuit is interrupted. Plasma current flow between opening relay contacts. An arc is enabled by the electric power of the load circuit (turn off spark) ionizing the gas between the contacts. The stability of the arc depends on various parameters such as contact material, air pressure, contact gap, etc. An arc locally produces high temperature causing contact erosion. In cases of strong erosion, spark suppression becomes necessary.

Arc suppression:- An arc will form as contacts come together and currents flow, and when they break apart. With ac current the condition is seldom a problem in relays, but with high dc loads the arc can be substantial causing contact damage. Arc suppression can be achieved using a blow out magnet.

Bounce:- Occurs as a moving contact strikes a fixed contact and 'bounces' before remaining full at rest. This has to be minimised, as creates signal noise and contact wear.

Bounce, armature:- See *rebound, armature*.

Bounce Time:- The time from the first to the last closing or opening of a relay contact.

Break:- The opening of closed contacts to interrupt an electrical circuit.

Break-Before-Make:- Disconnecting the present circuit before connecting a new circuit. Also known as Break/Make.

Break Contact:- NC contact. The break contact is closed in the release (rest) state of a monostable relay and opens (breaks) when the armature moves to the core (operate state).

Bridging:- (1) Normal bridging: The normal make-before-break action of a make-break or D contact combination. In a stepping switch, the coming together momentarily of two adjacent contacts by a wiper shaped for that purpose in the process of moving from one contact to another. (2) Abnormal bridging: The undesired closing of open contacts caused by a metallic bridge or protrusion developed by arcing.

Bunching, contact:- The undesired, simultaneous closure of make-and-break contacts during vibration, shock, or acceleration. Also, the simultaneous closure of the contacts of a continuity transfer or bridging contact combination.

Changeover Contact:- Contact configuration with make and break contact. Changing the switch position opens the closed contact first and then closes the formerly open contact.

Chatter, armature:- The undesired vibration of the armature due to inadequate ac performance or external shock and vibration.

Chatter, Contact:- Externally caused, undesired vibration of mating contacts during which there may or may not be actual physical contact opening. If there is no actual opening but only a change in resistance, it is referred to as dynamic resistance.

Closing Time:- Time between energization of the coil until the moment the contacts of the first current path to be closed actually close.

Coil:- That part of a relay which is energised to create a magnetic field that attracts a lever that in turn carries out the switching function. An assembly consisting of one or more windings, usually wound over an un-insulated iron core on a bobbin or spool. May be self-supporting, with terminals and any other required parts such as a sleeve or slugs.

1. Concentrically Wound-: A coil with two or more insulated windings wound one over the other.
2. Double Wound-: A coil consisting of two windings wound on the same core.
3. Parallel Wound:- A coil having multiple windings wound simultaneously, with the turns of each winding being contiguous, termed bifilar wound.
4. Sandwich Wound:- A coil consisting of three concentric windings in which the first and third windings are connected series aiding to match the impedance of the second winding. The combination is used to maintain transmission balance.
5. Tandem Wound:- A coil having tow or more windings, one behind the other, along the longitudinal axis. Also referred to as a two, three, or four-section coil, etc.

Coil Hi-Pot:- The minimum voltage (potential) which the relay coil terminals will isolate when the relay is properly mounted.

Coil Operating Range:- Expressed as a multiple of the rated control circuit voltage V_c for the lower and upper limits.

Coil Resistance:- The DC resistance of the energised relay coil measured at 25°C, not including a parallel device for coil suppression. *Shock* - The number of gravities (G's) a relay can sustain when tested by a ½ sine pulse (calibrated impact) for 11 milliseconds without the closed contacts opening or the open contacts closing. *Vibration* - The simple harmonic motion at rated gravities and frequency (G/Hz) that a relay can sustain without uncontrolled opening of closed contacts or closing of open contacts.

Coil Suppression Circuit:- Circuit to reduce the inductive switch off voltage peak of the relay coil (EMC protection, switch off voltage peak). Note that most of the circuits reduce the armature release speed, which can decrease the relay lifetime, especially valid for diodes in parallel to the coil. From the various solutions, the use of a Zener diode is particularly suitable.

Cold:- An unenergized electrical circuit.

Cold Switching:- Closing the relay contacts before applying voltage and current, plus removing voltage and current before opening the contacts. (Contacts do not make or break current.) Also termed Dry Circuit Switching. Larger currents may be carried through the contacts without damage to the contact area since contacts will not arc when closed or opened. Maximises contact life.

Contact:- Made out of contact material and part of the contact set where the electrical load circuit is opened or closed.

Contact Arrangement:- Relays are typically one of the following arrangements and contact forms:

single pole single throw (SPST) - Normally Open, NO, NO-double make
Normally Closed, NC, NC-double break
latching
single pole double throw (SPDT) - latching
double pole double throw (DPDT)
four pole double throw (4PDT)

Contact, Auxiliary:- A contact combination used to operate a visual or audible signal to indicate the position of the main contacts, establish interlocking circuits, or hold a relay operated when the original operating circuit is opened.

Contact Bounce:- The intermittent undesirable opening of closed mechanical contacts or closing of open contacts. Internally caused intermittent and undesired opening of closed contacts, or closing of open contacts, of a relay, caused by one or more of the following:

(1) Impingement of mating contacts;
(2) Impact of the armature against the coil core on pickup or against the backstop on dropout;
(3) Momentary hesitation or reversal of the armature motion during the pickup or dropout stroke.

Contact bounce period depends upon the type of relay and varies from ½ms for small reed relays to 10-20ms for larger solenoid types. Solid-state or mercury wetted contacts (Hg) do not have a contact bounce characteristic.

Contact, Break:- See contact, normally closed.

Contact, break-before-make:- A contact combination in which one contact opens its connection to another contact and then closes its connection to a third contact.

Contact, break-make:- See contact, break-before-make.

Contact Capacitance:- The capacitance of the relay measured (a) between the open contact, or (b) between contact terminals and ground. Measured at 1 kHz.

Contact Configuration:- Relay switch configuration (make, break or changeover contact). According to the application, various contact configurations are used. Contacts which are moved by the armature system are called → movable contacts, and non moving contacts stationary contacts.

Contact, Double Break:- A contact combination in which contact on a single conductive support simultaneously open electrical circuits connected to two independent contacts. This provides two contact air gaps in series when the contact is open Note: In B combination is terminal is brought out form the movable contact. In the Y combination, it is not.

Contact, Double Make:- A contact combination in which contacts on a single conductive support simultaneously close electrical circuits connected to the contact of two independent contacts, and provides two contact air gaps in series when the contact is open. (Sometimes called normally open, double-make contact.) Note: In U combination a terminal is brought out from the movable arm. In the X combination it is not.

Contact, Double Throw:- A contact combination having two positions as in break-make, make-break, and the like.

Contact Erosion:- Material loss at the contact surfaces, for example due to material evaporation by an arc.

Contact Force:- The force which two contact tips (points) exert against each other in the closed position under specified conditions.

Contact Gap:- The gap between the contact tips (points) under specified conditions, when the contact circuit is open.

Contact Interrupter:- On a stepping relay or switch, a contact combination operated directly by the armature that opens and closes the winding circuit, permitting the device to step itself.

Contact Life:- The maximum number of expected closures before failure. Life is dependent on the switched voltage, current, and power. Failure is usually when the contact resistance exceeds an end of life value. Typical failure mode is non-closure of the contact as opposed to a contact sticking closed.

Contact, Low Level:- Contact that control only the flow of relatively small currents in relatively low-voltage circuits; e.g., alternating currents and voltages encountered in voice or tone circuits, direct currents in the order of microamperes, and voltages below the softening voltages of record for various contact materials (that is, 0.080 volt for gold, 0.25 volt for platinum, etc.) Also defined as contacts switching loads where there is no electrical arc transfer of detectable thermal effect and where only mechanical forces can change the conditions of the contact interface.

Contact, Main:- The primary set of contacts of a relay, usually defined as those having the highest current rating.

Contact, Make:- See contact, normally open.

Contact, make-before-break:-See contact, continuity transfer.

Contact, make-break:- See contact, continuity transfer.

Contact Material:- For relays a variety of contact materials are in use. They operate under a wide range of loads in terms of voltage and current. Inductive loads can cause high switch off voltages and strong arcs, capacitors create inrush current peaks. Arcs and improper coil suppression can reduce the lifetime of a contact. So far, no universal contact material is known, that can be used on all load types with optimum performance. Contact manufacturers, relay developers, and users have established the following criteria to describe a contact:

- Electrical resistance
- Resistance to contact erosion
- Resistance to material transfer
- Resistance to welding

Contact, Normally Closed:- A contact combination which is closed when the armature is in its unoperated position. A pair of contacts are together at rest making an electrical circuit.

Contact, Normally Open:- A contact combination that is open when the armature is in its unoperated position. A pair of contacts are separated at rest with no electrical connection. (Generally applies to monostable relays.)

Contact, Off:- normal-A form C contact combination on a stepping switch that is in one condition when the relay or stepping switch is in its normal position and in the opposite condition for any other position of the relay or stepping switch; i.e., when not in its reset or home position.

Contact, Operate Time:- Time from initial energization to the first opening of closed contact or first closing of open contact, prior to bounce.

Contact Potential:- A voltage produced between contact terminals due to the temperature gradient across the relay contacts, and the reed-to-terminal junctions of dissimilar metals. (The temperature gradient is typically caused by the power dissipated by the energized coil.) Also known as contact offset voltage, thermal EMF, and thermal offset. This is a major consideration when measuring voltages in the microvolt range. There are special low thermal relay contacts available to address this need. Special contacts are not required if the relay is closed for a short period of time where the coil has no time to vary the temperature of the contact or connecting materials (welds or leads).

Contact Rating:- The voltage, current, and power capacities of relay contacts under specified environmental conditions.

Contact, Reed:-

1. A glass-enclosed, magnetically operated contact using thin, flexible, magnetic conducting strips as the contacting members.
2. Contact assembly, the contact members of which are blades either fully or partly of magnetic material and which are moved directly by a magnetic force.

Contact Release Time:- Time form initial de-energization of the relay coil to the first opening of a closed contact prior to bounce.

Contact Resistance:- The resistance between closed load contacts. In vacuum relays, this measurement is typically made at 6V dc with a 1A rms load. In gas-filled relays, 1A at 28V dc is used to measure contact resistance. 'Kelvin' connections should be used to obtain accurate readings. The resistance can be obtained from the ratio of the voltage drop across the relay and the load current (Ohm's law). Surface layers (fritting) can result in non-linear contact resistances and increased voltage.

Contact Transfer Time:- Time during which the moving contact first opens from a closed position and first makes with the opposite throw of the contact. It is floating in a non-contacting position prior to bounce and after energizing or de-energizing the coil.

Contact Weld:- A contact failure due to fusing of contacting surfaces to the extent that the contacts fail to separate when intended.

Continuous Current, Carry:- The maximum current that can be carried by the closed contacts of the relay for a sustained time period. This current rating is determined by the relay envelope temperature rise and must be derated at RF frequencies. A glass relay is allowed a 62°C rise, and a ceramic relay a 100°C temperature rise. Current ratings can be increased by external cooling, such as by forced air or heat sinks.

Crosstalk:- The electrical coupling between a closed contact circuit and other open or closed contact on the same relay or switch, expressed in decibels down form the signal level.

Current, maximum rate of rise on state (*di/dt*):- The maximum non-repetitive rate of current rise the output can withstand without being damaged.

1. With the relay output(s) turned on by the application or removal of the control voltage and/or current.
2. With the relay output(s) driven into break-over with the input at non-operate level.

Current, minimum load, $I_{Tmin(rms)}$:-The minimum current required to maintain the relay in the on-state (nominal load voltage applies). Applies mainly to solid-state relays.

Current, non-repetitive surge, I_{TSM}:- The maximum allowable, non-repetitive, peak, sinusoidal current that may be applied to the output for one full cycle at nominal line frequency. Relay control may be lost during and following the surge until the junction temperature falls below the maximum rated temperature.

Current rated contact:- The current which the contacts are designed to handle for their rated life.

Current, repetitive overload, $I_{TO(rms)}$ -The maximum allowable repetitive rms overload current that may be applied to the output for a specific duration and duty cycle while still maintaining output control. Applies mainly to solid state relays.

De-energize:- To remove power from a relay coil.

Dielectric:- An insulating medium capable of recovering, as electrical energy, all or part of the energy required to establish an electrical field (voltage stress). The field, or voltage stress, is accompanied by displacement or charging currents. A vacuum is the only perfect dielectric.

Dielectric strength, V_{ISO}:- The maximum allowable ac rms voltage (50/60Hz) which may be applied between two specified test points such as input-output, input-case, output-case in solid state relays, and between current-carrying and non-current-carrying metal members in electromechanical relays, without a leakage current in excess of 1mA.

Dropout, to drop out:- A monostable relay drops out when it changes from an energized to an un-energized condition. Not applicable latching relays.

Dropout, time:- See time, release.

Dropout Voltage:- The maximum coil voltage at which an operating relay releases and all normally closed contacts close. The voltage at which a relay (coil) de-energises sufficiently for the operating lever to move back to its rest position. It is normally expressed as a % of the nominal coil voltage.

Dry Circuit Switching:- Switching below specified levels of voltage and current to minimize any physical and electrical changes in the contact junction. Also see Cold Switching.

Dry reed relay:- See relay, reed.

Dynamic contact resistance:- A change in contact electrical resistance due to a variation in contact pressure on a contacts mechanically closed; occurrence is during non-bounce condition.

Electrical Endurance:- Number of on-load operating cycles (i.e. with current on the main contacts) a contactor can achieve, without failure, varies depending on the utilization category. The lifetime varies with the load. If not stated otherwise, the reference values apply for resistive or inductive loads with suitable spark suppression.

Electrostatic screening:- Screening plate between coil and contact to provide electrostatic screening in reed relays.

Energization:- The application of power to a coil winding of a relay to generate a magnetic field to move the armature. With respect to an operating coil winding, use of the word commonly assumes enough power to operate the fully. The energizing value is the product of the coil current and the number of wire turns of the coil.

Expected Mechanical Life:- The minimum number of operations for which a relay can be expected to operate reliably. "Cold" switching applications approach this figure.

Form:- A: Configuration which has one single-pole single-throw normally open (SPST no) contact.
B: Configuration which has one single-pole single-throw normally closed (SPST nc) contact.
C: Contact configuration which has one single pole-double throw (SPDT) contact. (One common point connected to one normally open and one normally closed contact.) Sometimes referred to as a transfer contact.

Freezing, magnetic:- Sticking of the relay armature to the core due to residual magnetism.

Fritting:- Electrical breakdown which can occur under special conditions (voltage, current) whenever thin contact films prevent electrical conductivity between closed contacts. Fritting is a process which generates (A-fritting) and/or widens (B-fritting) a conducting current path through such a semi-conducting film on a contact surface. During A-fritting, electrons are injected into the undamaged film. The electron current alters the condition of the film producing a 'conductive channel'. During the following B-fritting, the current widens the channel increasing the conductivity.

Gaging, relay contact:- The setting of relay contact spacing to determine the point in the armature's stoke at which specified contacts function.

Gap, contact:- The distance between a pair of mating relay contacts when the contacts are open.

Gap, heel:- A gap or nonmagnetic separation in the magnetic circuit other than between the armature and pole face. Generally, located between the heel piece and pole piece of an ac relay.

Gap, residual:- The thickness of nonmagnetic material in the magnetic circuit between the pole face centre and the nearest point on the armature when the armature is in the fully seated position.

Grass:- See *dynamic contact resistance*.

Hard failure:- Permanent failure of the contact being tested.

Hermetic seal:- An enclosure that is sealed by fusion to ensure a low rate of gas leakage. In a reed switch, a glass-to-metal seal is employed.

Hesitation, armature:- Delay or momentary reversal of armature motion in either the pickup or dropout stroke.

Hold value specified:- As the current or voltage on an operated relay is decreased, the value at or above which all relay contacts must restore to their un-operated positions.

Hold Voltage:- The lowest voltage that can be applied without any change in state of the contacts from their energized position. This is just above the maximum drop-out voltage.

Hot:- An energized electrical circuit.

Hot switching:- A circuit design that applies the switched load to the switch contacts at the time of opening and closure.

Inrush:- Inrush current is the peak current passing across the contacts of a relay when the contact is first made and is dependent on the load being switched. A relay which has contacts rated for a continuous current, the nominal contact current, may be capable of withstanding much higher currents for short periods. Inrush current can form a surge flowing through a relay switching a

low impedance source load - typically a highly reactive circuit, or one with a non-linear load characteristic such as a tungsten lamp load. Such abusive load surges are sometimes encountered when reed relays are inadvertently connected to test loads containing undischarged capacitors, or to long transmission lines with appreciable amounts of stored capacitive energy. Excessive inrush currents can cause switch contact welding or premature contact failure.

Insulation resistance, R_{ISO}:- The minimum allowable dc resistance between input and output of solid state relays and between contacts and coil for electromechanical and reed relays, at a specified voltage, usually 500V dc.

Isolation:- The value of insulation resistance, dielectric strength, and capacitance measured between the input and outputs, input to case, output to case, and output to output when applicable.

Latching:- In relay or switching technology, this refers to the ability to keep the contact status in place even if power is removed from the equipment.

Latching relay:- In a latching relay, after the coil input voltage is disconnected, the contacts remain in the last reached switching position. Normally latching relays are reset contact position. Latching relays only require a short set respectively reset impulse. A permanent coil power supply after setting/resetting the relay is neither necessary nor allowed: maximal pulse durations depend on the relay family. Hence the distinguishing characteristic of monostable relays in respect to a fail safe behaviour is the fact that the predefined contact rest position will be reached at break down of the power supply. This behaviour cannot be shown by latching relays due to the bistable working principle they are based on.

Leakage Current:- The rms current conducted by the output circuit of the relay at maximum rated voltage with the contacts open.

Limiting continuous current:- The highest current (effective value for AC loads) a relay can carry under specified conditions without exceeding its specified upper limit temperature. This is not the current that can be switched with any load over the specified lifetime.

Load:- The electrical circuit which is being switched is measured and defined by

1. current in amperes, A
2. voltage in volts, V: dc or ac, and
3. load type (Inductive or resistive current flow when the contact is first made).

A relay is generally limited by the amount of heat that occurs when an electrical current passes across its contacts. This represents the 'load' that a relay can switch and is normally presented as an electrical value This is usually stated as a contact current in A then a voltage often standardised at 250Vac/dc followed by a maximum capacity at a resistive load. This is the result of multiplying current by voltage, expressed as VA. It is usually the maximum permissible load at any time including starting and stopping.

Load, curve:- The static force/displacement characteristics of the total spring-load of the relay.

Load Life:- The minimum number of cycles the relay will make, carry, and break the specified load without contact sticking or welding, and without exceeding the electrical specifications of the device. Load life is established using various methods including Weibull probability methods.

Magnet, blowout:- A device that establishes a magnetic field in the contact gap to help extinguish the arc by displacing it.

Magnetic interaction:- Mainly relevant to reed relays. The tendency of a relay to be influenced by an external magnetic field. This influence can result in depression or elevation of the pull-in and drop out voltage of the affected relay, possibly causing operation outside its specification. Magnetic interaction can be minimized by alternating the polarity of adjacent relay coils, by magnetic shielding, or by placing two relays at right angles to each other.

Magnetic shield:- Mainly relevant to reed relays. A ferromagnetic material used to minimize magnetic coupling between the relay and external magnetic fields.

Make:- The closure of open contacts to complete an electric circuit.

Maximum operate voltage (or must operate voltage):- Voltage at room ambient temperature (RT) a relay must operate at. To guarantee proper function of all relays, the applied coil voltage in the application must be above this specified operating voltage. The actual operate voltage of an individual relay, the maximum operate voltage and the application system value are sometimes all called operate voltage.

Maximum voltage U_{max} or V_{max}:- Maximum coil voltage at RT, at which the coil reaches the specified upper limit temperature without contact load (maximum continuous thermal load at 23°C).

Maximum switching power:- Maximum permissible power switched by the relay contacts, i.e. the product of the switching current and switching voltage.

Mechanical Endurance:- Number of off-load operating cycles (i.e. without current on the main contacts) a contactor can achieve.

Mechanical Life:- This is the number of operations which a relay can be expected to perform while maintaining mechanical integrity. Mechanical life is normally tested with no load or voltage applied to the power contacts and is established using various methods including Weibull analysis.

Mechanical shock, non-operating:- The mechanical shock level (amplitude, duration and wave shape) to which the relay may be subjected without permanent electrical or mechanical damage (during storage or transportation).

Mechanical shock, operating:- That mechanical shock level (amplitude, duration and wave shape) to which the relay may be subjected without permanent electrical or mechanical damage during its operating mode.

Mercury wetted (contact) relay:- A form of reed relay in which the contacts are wetted by a film of mercury (Hg) obtained by a capillary action from a mercury pool encapsulated within the reed switch. Usually has a required operating position (usually vertical) to avoid liquid mercury from shorting the contacts; other types are position insensitive. This type of relay is usually higher power and longer life, but at a higher dollar cost. Another benefit of this type of contact is the repeatability of contact resistance and virtually no contact bounce.

Minimum recommended voltage:- Minimum load voltage to ensure an adequate contact cleaning (see also 'fritting').

Minimum voltage U_{min} or V_{min}:- Minimum coil voltage at RT where a relay is still able to operate.

Minimum release voltage (must release voltage):- Voltage at RT a relay must release at. To guarantee proper function of all relays, the limit in the application must be below this specified release voltage .The release voltage of an individual relay, the guaranteed minimum (must) release voltage and the system value are sometimes all called release voltage.

Minimal operation time:- Shortest control duration to ensure complete closing or opening of a contactor.

NC contact (normally closed):- Same as break contact. The break contact is closed in the release (rest) state of a monostable relay and opens (breaks) when the armature moves to the core (operate state).

NO contact (normally open):- Same as make contact. Contact is open in the release (rest) state of a monostable relay and closes (makes) when the relay coil is energized (operate state).

Nonpickup value, specified:- As the current or voltage on an unoperated relay is increased, the value which must be reached before any contact change occurs.

Nonrelease, specified:- See operating characteristics, hold value.

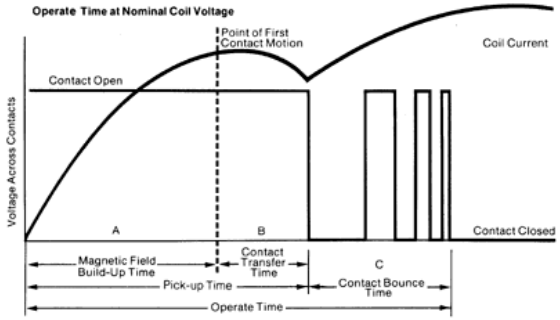
Offstate dv/dt:- The application of both position and negative voltages with maximum specified rate of rise to the output terminals.

Operate:- A relay operates when sequentially it starts, it passes from an initial condition towards the prescribed operated condition, and it switches.

Operating characteristics:- Pickup, non-pickup, hold and dropout, voltage and current.

Operating temperature range:- The ambient temperature range over which an un-mounted relay is specified to operate.

Opening time:- Time from the beginning of state causing breaking until the moment when the contacts of the last current path to be opened are open.



Operate time:- The time in milliseconds between voltage being first applied to the relay coil and final closure of all normally open contacts or the time from energizing the relay coil till the first break of the NC contact. This includes time for the coil to build up its magnetic field (a significant limiting factor) and transfer time of the moveable contact between stationary contacts, and bounce time after the initial contact make. As the coil resistance depends on the ambient temperature, the operate time varies with the operate voltage and the ambient temperature.

Overdrive:- A term used to indicate use of greater than normal coil current (applied voltage), and usually employed in obtaining well-controlled bounce and fast operate time or pulse response.

Overload current:- Test done to make sure that relays withstand overload conditions, e.g. withstand short circuit conditions until the fuse opens. Relay will carry the specified currents at 23°C (I_{rated} = rated current as given in contact data section for each relay).

Overtravel armature dropout:- The portion of the armature travel that occurs between closure of the normally closed contact(s) and the fully released static position of the armature.

Overtravel armature pickup:- The portion of the armature travel occurring between closure of the normally open contact(s) and the fully operated static position of the armature.

Paschen test:- Test to detect sealing damage to a hermetically seal capsule. In the case of a cracked switch capsule or damaged switch seal, atmospheric oxygen can leak into the switch and eventually oxidize the switch contacts, causing increased contact resistance and possible contact failure. The presence of oxygen causes the breakdown avalanche voltage to increase, due to the ability of the electronegative oxygen to scavenge free electrons. The Paschen test observes the variation and magnitude of the breakdown voltage as a switch is opened, hence used to diagnose the presence of oxygen.

Peak Test Voltage:- The peak AC voltage (at 60 Hz) which can be applied between external high voltage terminals or between the open terminals and ground for up to one minute with no evidence of failure. Peak test voltages must not be exceeded, even for very short pulses.

Pole, double:- A term applied to a contact arrangement to denote that it includes two separate contact combinations, that is, two single-pole contact assemblies.

Pole piece:- The end of an electromagnet, sometimes separable from the main section, and usually shaped so as to distribute the magnetic field in a pattern best suited to the application.

Pole, single:- A term applied to a contact arrangement to denote that all contacts in the arrangement connect in one position or another to a common state.

Pull-in Voltage:- The minimum coil voltage required to operate a relay for all normally open contacts to close. The voltage at which a relay (coil) operates and switches. It is normally expressed as a % of the nominal coil voltage. Sometimes known as threshold voltage. It is affected by temperature.

Race, relay:- A deficient circuit condition wherein successful operation depends upon a sequence of two or more independent contacts and in which the sequence is not insured by electrical or mechanical interlocking restraints. Ratchet relay. See relay, stepping.

Relay:- An electrically controlled mechanical device that opens and closes electrical contacts when a voltage (or current) is applied to a coil. A relay provides isolation of control signals from switched signals.

Rated breaking capacity; Rated making capacity:- Value of rms current a contactor can break or make at a fixed voltage value, within the conditions specified by the standards, depending on the utilization category.

Rated impulse withstand voltage, V_{imp}:- The highest peak value of an impulse voltage of prescribed form 1.2/50, which does not cause breakdown under specified conditions of test.

Rated insulation voltage, V_i:- Voltage value which designates the unit and to which dielectric tests, clearance and creepage distances are referred.

Rated operating current, I_e:- Current value stated by the manufacturer and taking into account the rated operating voltage, V_e, the rated frequency, the rated duty, the utilization category, the electrical contact life and the type of the protective Hammond Enclosure.

Rated operating voltage, V_e:- Voltage value to which utilization characteristics of the contactor are referred, i.e. phase to phase voltage in 3 phase circuits. The voltage which can safely be applied to the relay for sustained periods of time without failure. This voltage rating decreases as AC frequency increases. Rated operating voltages approach peak test voltage only at lower frequencies.

Rating, contact:- The electrical load-handling capability of relay contacts under specified conditions and for a prescribed number of operations.

Rating, short time:- The value of current or voltage that the relay can stand, without injury, for specified short time intervals. (For ac circuits, the rms total value, including the dc component, should be used). The rating recognized the limitations imposed by both thermal and electromagnetic effects.

Rebound, armature:- (1) The return motion or bounce-back toward the unoperated position after the armature strikes the pole face during pickup, referred to as armature pickup rebound; (2) The forward motion or bounce in the direction of the operated position when the armature strikes its backstop on dropout, referred to as armature dropout rebound.

Relay:- An electric device that is designed to interpret input conditions in a prescribed manner and after specified conditions are met to respond to cause contact operation or similar abrupt change in an associated electric control circuit. Notes: (a) Inputs are usually electric, but may be mechanical, thermal or other quantities. (b) A relay may consist of several units, when responsive to specified inputs, the combination providing the desired performance characteristic.

Relay, alternating current (ac):- A relay designed for operation from an alternating-current source.

Relay, direct current (dc):- A relay designed for operation from a direct-current source.

Relay, electrical:- A device designed to produce sudden, predetermined changes in one or more electrical output circuits, when certain conditions are fulfilled in the electrical input circuits controlling the device.

1. The term relay shall be restricted to a relay unit having a single relaying function between its input circuits and its output circuits.
2. The term relay includes all the components which are necessary for its specified operation.
3. The adjective ‘electrical’ can be deleted when no ambiguity may occur.

Relay, electromechanical:- An electrical relay in which the designed response is developed by the relative movement of mechanical elements under the action of a current in the input circuits.

Relay, latching:- A relay that maintains its contacts in the last position assumed without the need of maintaining coil energization.

1. Magnetic latching- A relay that remains operated, held either by remanent magnetism in the structure or by the influence of a permanent magnet, until reset.
2. Mechanical latching- A relay in which the armature or contacts may be latched mechanically in the operated or unoperated position until reset manually or electrically.

Relay, mercury contact:-

1. Mercury-wetted contact-A form of reed relay in which the reeds and contacts are glass enclosed and are wetted by a film of mercury obtained by capillary action from a mercury pool in the base of a capsule vertically mounted.
2. Mercury contact-A relay mechanism in which mercury establishes contact between electrodes in a sealed capsule.

Relay, over current:- A relay that is specifically designed to operate when its coil voltage reaches or exceeds a predetermined value.

Relay, polarized:- A relay whose operation is dependent upon the polarity of the energizing current.

1. Bistable. A tow-position relay that will remain in its last operated position keeping the operated contacts closed after the operating winding is de-energized.
2. Centre-stable. A polarized relay that is operated in one of two energized positions, depending on the polarity of the energizing current, and that returns to a third, off position, when the operating winding is de-energized.
3. Double-biased. See bistable.
4. Magnetic latching. See bistable.
5. Monostable. A monostable polarized relay is a two-position relay that requires current of a pre-determined polarity for operation and returns to the off position when the operating winding is de-energized or is energized with reversed polarity.
6. Single-biased. See monostable.
7. Single-side-stable. See centre-stable.
8. Three-position centre-off. See centre-stable.
9. Un-biased. See centre-stable.

Relay, reed:- A relay using glass-enclosed, hermetically sealed, magnetically actuated reeds as the contact members. No mercury or other wetting material is used. Typical atmosphere inside the glass enclosure is nitrogen.

Relay, RF switching:- A relay designed to switch electrical ac energy frequencies >20kHz.

Relay, solid state (SSR):- A relay with isolated input and output whose functions are achieved by means of electronic components and without moving parts.

Relay, undercurrent:- A relay specifically designed to function when its energizing current falls below a predetermined value. (See relay, current sensing.)

Relay, undervoltage:- A relay specifically designed to function when its energizing voltage falls below a predetermined value.

Relay, vacuum:- A relay whose contacts are sealed in a low pressure environment.

Release Time:- The time in milliseconds between removal of power from the relay coil and final closure of all normally closed contacts. This time includes contact bounce.

Resistance, contact:- The electrical resistance of closed contacts measured at the associated terminals.

Resistance, dynamic contact:- Variation in contact resistance due to changes in contact pressure during the period in which contacts are motion, before opening or after closing.

Resistance to shock:- Requirements applicable for instance to vehicles, crane operation or switchgear slide-in module systems. At the quoted permissible ‘g’ values, contactors must not undergo a change in switching state and O/L relays must not trip.

Resistance to vibration:- Requirements applicable to all the vehicles, vessels and other similar transport systems. At the quoted amplitude and vibration frequency values, the unit must be capable to achieve the required duty.

Resistance, winding:- The total terminal-to-terminal resistance of a winding at a specified temperature.

Self de-energize:- The removal of power from a relay coil by an auxiliary switch or contact within the relay itself. Usually applies to latching relays only.

Self de-energizing switch:- A secondary relay or auxiliary contact usually enclosed within the primary relay which removes power from the primary relay coil after it has transferred position. Usually applies to latching relays only.

Shield, electrostatic:- A conductive metallic sheath surrounding the relay's reed switch, connected to at least one external relay pin, and designed to minimize capacitive coupling between the switch and other relay components, thus reducing high frequency noise pickup. Similar to a coaxial shield, but not necessarily designed to maintain a 50 Ohm RF impedance environment.

Shield, magnetic:- An optional plate or shell constructed of magnetically permeable material such as nickel-iron or mu-metal, fitted external to the relay's coil. Its function is to reduce the effects of magnetic interaction between adjacent relays, and to improve the efficiency of the relay coil. A magnetic shell also reduces the influence of external magnetic fields, which is useful in security applications. Magnetic shields can be fitted externally, or may be buried inside the relay housing.

Soft failure:- Intermittent, self-recovering failure of a contact.

Static contact resistance:- The DC resistance of closed contacts as measured at their associated contact terminals. Measurement is made after stable contact closure is achieved.

Sticking (contacts):- A reed switch failure mechanism, whereby a closed contact fails to open by a specified time after relay de-energization. Can be sub-classified as hard or soft failures.

Switch, dry reed:- See contact, reed.

Switch, stepping:- A class of electromagnetically operated, multiposition switching devices. Their wipers are rotated in steps so that contact is successively made between the wiper tips and contacts that are separated electrically and mounted in a circular arc called a bank.

Switching Capacity:- Switching capacity is the product of switching voltage and switching current. The current which a relay will switch will vary according to the voltage being used. Note that the maximum often includes the value occurring at peaks (see 'inrush'). A minimum also applies, because contact materials that can withstand high current loads may be poor at making contact at low current loads.

Switching frequency:- Number of operating cycles per hour.

Short time current permissible:- Value of current which the contactor can withstand in closed position for a short time period and within specified conditions.

Time, actuation:- The time interval from coil energization or de-energization to the functioning of a specified contact; same as time, contact actuation, subdivided as follows:

1. Time, final actuation-The sum of the initial actuation time and the contact bounce intervals following such actuation.
2. Time, initial actuation-The time from coil energization or de- energization to the first closing of a previously open contact or the first opening of a previously closed contact.

Time, bridging:- The time in which all contacts of a continuity transfer combination are electrically connected during the transfer.

Time constant:- Ratio of inductance to the resistance : *L/R* = mH/Ohm, ms.

Time, contact bounce:- The time interval from initial actuation of a contact to the end of bounce.

Time, contact stagger:- The time interval between the functioning of contacts on the same relay.(For example, the time difference between the opening of two normally closed contacts on pickup.)

Time, operate:- (1) The time interval from coil energization to the functioning of the last contact to function. Where not otherwise stated, the functioning time of the contact in question is taken as its initial actuation time (that is, it does not include contact bounce time).
(2) For a solid state or hybrid relay in a non-operated state, the time from the application of the pickup voltage to the change of state of the output.

Time, release:- (1) The time interval from coil de-energization to the functioning of the last contact to function. Where not otherwise stated, the functioning time of the contact in question is taken as its initial actuation time (that is, it does not include contact bounce time).
(2) For a solid state or hybrid relay in an operated state, the time from the application of the dropout voltage to the change of state of the output.

Time, seating:- The time interval from coil energization to the seating of the armature.

Time transfer:- The time interval between opening the closed contact and closing the open contact of a break-before-make contact combination.

Type 1 co-ordination:- There has been no discharge of parts beyond the enclosure. Damage to the contactor and the overload relay is acceptable.

Type 2 co-ordination:- No damage to the overload relay or other parts has occurred, except that welding of contactor or starter contacts is permitted, if they are easily separated.

Voltage Breakdown:- An undesirable condition of arcing within a relay due to over-voltage.

Voltage, off state:- In solid state relay, the following determine whether the relay will stay off under each load voltage condition:

1. Critical rate of rise of commutation voltage, *dv/dt*. The maximum value of the rate of rise of principal voltage which will cause switching from the off state to the on state.
2. Maximum off state voltage, $V_{D\ max\ rms}$. The maximum effective steady state voltage that the output is capable of withstanding when in off state.
3. Maximum rate of rise of off state voltage, dv/dt. The rate of rise of the off-state voltage which the output can withstand without false operation.

4. Minimum off state voltage, $V_{D\ min\ rms}$. The minimum effective voltage which the relay will switch.
5. Non-repetitive peak voltage, V_{DSM}. The maximum off-state voltage that the output terminals are capable of withstanding without breakover or damage.

Voltage, on state:- In solid state relays, the output terminal wave form at rated current consists of repetitive half-cycles (+and-) of distinctive voltage drops. Each voltage state is necessary for load current conduction and may be specified for specific applications, as follows:

1. Instantaneous on state voltage, V_T. The instantaneous voltage across the output when in the on condition.
2. Maximum RMS on state voltage, $V_{T\ RMS}$. Maximum RMS voltage drop across the relay output at maximum load current $I_{T\ RMS}$.
3. Minimum power factor load, PF_{MIN}. The minimum power factor load the relay will switch and still meet all of its electrical specifications.
4. Peak on state voltage, V_{TM}. The maximum value of V_T excluding ± 20° of zero crossing of the voltage waveform.

Voltage, rated coil:- The coil voltage at which the relay is intended to operate for the prescribed duty cycle. Note: The use of any coil voltage less than rated may compromise the performance of the relay.

Voltage, reverse polarity:- The maximum allowable reverse voltage which may be applied to the input of a solid state relay without permanent damage.

Winding, non-inductive:- A winding in which the magnetic fields produced by two parts of the winding cancel each other and provide non-inductive resistance.

Glossary of Varistor terminology

AC Standby Power (Varistor), P_D:- Varistor AC power dissipation measured at rated rms voltage $V_{M(ac)}$.

Capacitance (Varistor), *C*:- Capacitance between the two terminals of the varistor measured at *C* specified frequency and bias.

Clamping Voltage, V_C:- Peak voltage across the varistor measured under conditions of a specified peak VC pulse current and specified waveform. Peak voltage and peak currents are not necessarily coincidental in time.

Dynamic Impedance (Varistor), Z_X:- measure of small signal impedance at a given operating point as defined by: $Z_X = dV_X / dI_X$

Lifetime Rated Pulse Currents (Varistor):- Derated values of I_{TM} for impulse durations exceeding that of an 8/20µs wave-shape, and for multiple pulses which may be applied over device rated lifetime.

Nominal Varistor Voltage, $V_{N(dc)}$:- Voltage across the varistor measured at a specified pulsed DC current, $I_{N(dc)}$, of specific duration, $I_{N(dc)}$ of specific duration. $I_{N(dc)}$ is specified by the varistor manufacturer.

Nonlinear Exponent, α:- A measure of varistor nonlinearity between two given operating currents, I_1 and I_2, as described by $I = kV^{\alpha}$
where k is a device constant, $I_1 \le I \le I_2$, and $\alpha_{12} = \log I_2 / I_1 / \log V_2 / V_1$

Overshoot Duration (Varistor):- The time between the point voltage level (V_C) and the point at which the voltage overshoot has decayed to 50% of its peak. For the purpose of this definition, clamping voltage is defined with an 8/20µs current waveform of the same peak current amplitude as the waveform used for this overshoot duration.

Peak Nominal Varistor Voltage, $V_{N(ac)}$ Voltage across the varistor measured at a specified peak AC current, $I_{N(ac)}$, of specific duration. $I_{N(ac)}$ is specified by the varistor manufacturer.

Rated DC Voltage (Varistor), $V_{M(dc)}$:- Maximum continuous DC voltage which may be applied.

DC Standby Current (Varistor), I_D:- Varistor current measured at rated voltage, $V_{M(dc)}$.

Rated Peak Single Pulse Transient Currents (Varistor), I_{TM}:- Maximum peak current applied for a single 8/20µs impulse, with rated line voltage also applied, without causing device failure.

Rated Recurrent Peak Voltage (Varistor), V_{PM}:- Maximum recurrent peak voltage which may be applied for a specified duty cycle and waveform.

Rated RMS Voltage (Varistor), $V_{M(ac)}$:- Maximum continuous sinusoidal RMS voltage which may be applied.

Rated Single Pulse Transient Energy (Varistor), W_{TM}:- Energy which may be dissipated for a single impulse of maximum rated current at a specified wave-shape, with rated RMS voltage or rated DC voltage also applied, without causing device failure.

Rated Transient Average Power Dissipation (Varistor), $P_{T(av)M}$:- Maximum average power which may be dissipated due to a group of pulses occurring within a specified isolated time period, without causing device failure.

Resistance (Varistor), R_X:- Static resistance of the varistor at a given operating point as defined by:
$R_X = V_X / I_X$

Response Time (Varistor):- The time between the point at which the wave exceeds the clamping voltage level (V_C) and the peak of the voltage overshoot. For the purpose of this definition, clamping voltage as defined with an 8/20µs current waveform of the same peak current amplitude as the waveform used for this response time.

Varistor Voltage, V_X:- Voltage across the varistor measured at a given current, I_X.

Voltage Clamping Ratio (Varistor), V_C / V_P:- A figure of merit measure of the varistor clamping effectiveness as defined by the symbols V_C / $V_{M(ac)}$, V_C / $V_{M(dc)}$.

Voltage Overshoot (Varistor), V_{ost}:- The excess voltage above the clamping voltage of the device for a given current that occurs when current waves of less than 8µs virtual front duration are applied. This value may be expressed as a % of the clamping voltage (V_C) for an 8/20µs current wave.

Glossary of PTC and NTC Thermistor terminology

Amorphous:- Without crystallization in the ultimate texture of a solid substance. Used to describe the device material structure in the tripped state.

Breakdown voltage:- The maximum voltage that a PTC thermistor can support under stipulated time and temperature conditions. The PTC thermistor will breakdown when exceeding this voltage.

Carbon Black:- A conductive material used in PTC devices to provide a path for current flow under normal operating conditions.

Conductive Plastic:- A plastic material, such as a polymer, containing conductive particles, such as carbon black, that provide a path for current flow.

Current-time characteristic:- The current-time characteristic is the relationship at a specified ambient temperature between the current through a thermistor and time, upon application or interruption of voltage to it.

Current, Hold, I_{hold}:- The maximum current a PTC device can pass without interruption.

Current, Maximum, I_{max}:- The maximum fault current a PTC device can withstand without damage at the rated voltage.

Current Rating:- The nominal amperage value marked on the fuse. It is established by the manufacturer as a value of current which the fuse can be loaded to, based on a controlled set of test conditions (see Rerating).

Current, Trip, I_{trip}:- The minimum current that will switch a device from the low resistance to the high resistance state.

Curie point temperature (Resistance - temperature characteristics):- A PTC fuse maintains almost the same resistance, until certain temperature. After this temperature is exceeded, the resistance rises up sharply. This transition point is called the Curie point. The critical temperature is defined to be the Curie point temperature, where the actual resistance value is twice the reference value measured at 25°C.

Derating:- Fuses are essentially temperature-sensitive devices. Even small variations from the controlled test conditions can greatly affect the predicted life of a fuse when it is loaded to its nominal value, usually expressed as 100% of rating. The fuse temperature generated by the current passing through the fuse increases or decreases with ambient temperature change.

Dissipation constant:- The dissipation constant is the ratio, (W/°C) at a specified ambient temperature, of a change in power dissipation in a thermistor to the resultant body temperature change.

Electrode:- A device or material that emits or controls the flow of electricity. Nickel and Copper elements are used in PTC devices to aid even distribution of current across the surface of the device.

Fault Current:- The peak current that flows through a device or wire during a short circuit or arc back.

Form Factor:- The package that holds the chemical make-up of polymer and carbon. PTCs are packaged in the following forms; radial, axial, surface mount chips, disks, and washers.

Fuse:- A current limiting device used for protection of equipment. Typically a wire or chemical compound which breaks a circuit when the current exceeds a rated value.

Fuse Resistance:- The resistance of a fuse is usually an insignificant part of the total circuit resistance. Since the resistance of fractional amperage fuses can be several ohms, this fact should be considered when using them in low-voltage circuits. Most fuses are manufactured from materials which have positive temperature coefficients, and therefore, it is common to refer to cold resistance and hot resistance (voltage drop at rated current), with actual operation being somewhere in between. The factory should be consulted if this parameter is critical to the design analysis. Resistance data on all of our fuses is available on request.

Heat capacity, *H*:- The heat capacity of a thermistor is the amount of heat required to increase the body temperature of it by one degree centigrade, 1°C. Heat capacity is a common rating of standard PTC thermistors and is expressed in Joules per cubic centimetre per degree C (J/cm^3/°C). The heat capacity per unit volume relationship of standard PTC thermistors is approximately 5 J/cm^3/°C.

Hysteresis:- The period between the actual beginning of the signalling of the device to trip and the actual tripping of the device.

Initial current (I_{in}):- the current that results instantaneously in the circuit switch when starting to closing.

Initial resistance ($R_{25°C}$):- This is the part's resistance value at 25°C which is measured under conditions of 1.0V dc or less, and 10mA or less without self-heating.

Inrush current:- Inrush current is the initial surge of current that results when power is first applied to a load having a low starting impedance, such as a discharged capacitor, a cold lamp filament, or a stopped motor's winding.

Inrush current limiter:- Specially designed and constructed NTC thermistors may be used as inrush current limiters. Available in a wide range of current handling and zero-power resistance value combinations.

Insulation thermistor:- thermistor stipulated insulation resistance and voltage test requirement.

Interrupting Rating:- Also known as breaking capacity or short circuit rating, the interrupting rating is the maximum approved current which the fuse can safely interrupt at rated voltage. During a fault or short circuit condition, a fuse may receive an instantaneous overload current many times greater than its normal operating current. Safe operation requires that the fuse remain intact (no explosion or body rupture) and clear the circuit.

Leakage Current:- An undesirable small value of stray current that flows through a device after the device has changed state to a high resistance mode.

Let through Current:- The amount of current though a circuit after a device is signalled to trip and the device is at full operation limiting current.

Low category temperature:- Minimum ambient temperature at which a PTC thermistor can operate continuously.

Material constant (Beta , ß in K):- The material constant of a NTC thermistor is a measure of its resistance at one temperature compared to its resistance at a different temperature. Its value may be calculated by the formula shown below and is expressed in degrees Kelvin (K). The reference temperatures used in this formula for determining material constant ratings of thermistors are 298.15°K and 348.15°K.

Maximum Fault Current:- The Interrupting Rating of a fuse must meet or exceed the maximum fault current of the circuit.

Maximum Inrush Current:- The maximum current (effective value) through the PTC thermistor under maximum rated voltage. Exceeding this current may result in PTC device damage.

Maximum operating temperature:- The maximum operating temperature is the maximum body temperature at which the thermistor will operate for an extended period of time with acceptable stability of its characteristics. This temperature is the result of internal or external heating, or both, and should not exceed the maximum value specified.

Maximum operating voltage, V_{max}:- The maximum operating voltage is the maximum rated voltage, either direct current or 50/60 Hz rms alternating current, expressed in volts (Vdc or Vac), that a standard PTC thermistor will continuous withstand for an extended period without affecting its normal characteristics.

Maximum power rating:- The maximum power rating of a thermistor is the maximum power which a thermistor will dissipate for an extended period of time with acceptable stability of its characteristics.

Maximum steady-state current (I_{max}):- The maximum steady-state current is the rating of the maximum current, normally expressed in amperes (A), allowable to be conducted by an inrush limiting NTC thermistor for an extended period of time.

Maximum surge current:- The maximum surge current is the maximum permissible surge current in a circuit and, in conjunction with the maximum peak voltage, determines the minimum required zero-power resistance of the thermistor required to limit it adequately.

Minimum switching current (I_s):- The minimum switching current is the minimum amount of current, normally expressed in amperes (A), that, when conducted by a standard PTC thermistor, is required to cause it to switch to its high resistance state.

Negative temperature coefficient (NTC):- A NTC thermistor is one in which the zero-power resistance decreases with an increase in temperature.

Non-insulation thermistor:- thermistors that do not require an insulation voltage and insulation resistance test.

Non-trip Current:- Also called rated current or holding current, or non-operating current, means the current at which PTC thermistor resistance does not exceed the specified value for designated time and temperature conditions.

Overload Current Condition:- The current level for which protection is required. Fault conditions may be specified, either in terms of current or, in terms of both current and maximum time the fault can be tolerated before damage occurs. Time-current curves are used to match the fuse characteristic to the circuit needs, noting that the curves are based on average data.

Peak current ($I_{in\ p\text{-}p}$):- Peak-peak value of initial current.

Polymer:- A synthetic plastic material consisting of large molecules made up of a linked series of repeated simple monomers. The insulating medium used in PTC devices which maintains the carbon chains in suspension during over-current while permitting the carbon chains to form during normal operation.

Polymeric Positive Temperature Coefficient (PPTC):- A characteristic of PTC devices that describes a large increase in resistance as the device reaches its trip temperature.

Positive temperature coefficient (PTC):- A PTC thermistor is one in which the zero-power resistance increases with an increase in temperature.

Pulses:- The general term 'pulses' is used in this context to describe the broad category of wave shapes referred to as surge currents, start-up currents, inrush currents, and transients, Electrical pulse conditions can vary considerably from one application to another. Different fuse constructions may not all react the same to a given pulse condition. Electrical pulses produce thermal cycling and possible mechanical fatigue that could affect the life of the fuse. The start-up pulse should be defined and then compared to the time-current curve and I^2t rating for the fuse. Nominal melting I^2t is a measure of the energy required to melt the fusing element and is pressed as Ampere squared seconds, (A^2s).

Recovery time:- The recovery time of a thermistor is the approximate time required for it to cool sufficiently after power is removed and allow it to provide the characteristics required when power is reapplied.

Resistance at maximum current, R_{Imax}:- The resistance at maximum current is the approximate resistance of an inrush current limiting thermistor, expressed in ohms, when it is conducting its rated maximum steady-state current.

Resistance ratio characteristic:- The resistance ratio characteristic identifies the ratio of the zero-power resistance of a thermistor measured at 25°C to that resistance measured at 125°C.

Resistance-temperature characteristic:- The resistance-temperature characteristic is the relationship between the zero-power resistance of a thermistor and its body temperature.

Resistance, Initial (R_{min} - R_{max}):- The resistance range of the PTC devices, before circuit insertion.

Resistance, Post Trip (R_{1max}):- The maximum post-trip resistance one hour after a PTC device has been tripped and power has been removed.

Resistance, Post Reflow (R_{1max}):- The maximum resistance one hour after a PTC surface mount device has been reflow soldered.

Restore time:- Time to restore PTC thermistor resistance to twice the zero-power resistance after the power is removed.

Silicon PTC thermistor:- A silicon PTC thermistor is a type PTC thermistor that has an approximately linear resistance-temperature characteristic and a temperature coefficient of resistance of approximately +0.7%/°C. Silicon PTC thermistors are distinguished from standard PTC thermistors.

Stability:- Stability of a thermistor is the ability of a thermistor to retain specified characteristics after being subjected to designated environmental or electrical test conditions.

Standard PTC thermistor:- A standard PTC thermistor is a type of PTC thermistor that has a switch temperature. Standard PTC thermistors are distinguished from silicon PTC thermistors.

Standard Reference Temperature:- The standard reference temperature is the thermistor body temperature at which nominal zero-power resistance is specified, 25°C.

Switch Temperature:- The temperature at which the resistance value of the PTC thermistor increases to twice the zero-power resistance, also called Curie temperature, or reference temperature or transition temperature.

Switching time, t_s:- If V_{max} and I_{max} are known, the PTC thermistor's switch-off behaviour can be described in terms of switching time t_s. This is the time it takes at applied voltage for the current passing through the PTC to be reduced to half of its initial value, at T_A = 25 °C.

Temperature - wattage characteristics:- The temperature-wattage characteristic of a thermistor is the relationship at a specified ambient temperature between the thermistor temperature and the applied steady state wattage.

Temperature at minimum resistance (T_{rmin}):- Temperature corresponding to minimum resistance.

Temperature coefficient of resistance, α:- The temperature coefficient of resistance is the ratio at a specified temperature, T, of the rate of change of zero-power resistance with temperature to the zero-power resistance of the thermistor. The temperature coefficient is commonly expressed in percent per degree C (%/°C).

Temperature range under maximum voltage:- Operating ambient temperature range that the PTC thermistor can continuously operate under maximum voltage.

Thermal cooling time constant τ_{th}:- The thermal cooling time constant refers to the time necessary for an unloaded (zero power conditions) thermistor to vary its temperature by 63.2% of the difference between its mean temperature and the ambient temperature.
Equation for temperature change: $T(t_2) = T(t_1) \pm 0.632x(T(t_1) - T_A)$ with $t_2 - t_1 = \tau_{th}$

Thermistor:- A thermistor is a thermally sensitive resistor whose primary function is to exhibit a change in electrical resistance with a change in body temperature.

Trip Current:- Initial current which causes PTC thermistor resistance to leap, also called operating current.

Trip Endurance:- A test used to determine the duration of time a PTC device will sustain its maximum rated voltage in the tripped state without failure.

Trip Cycle Life:- A test used to determine the number of trip cycles (at V_{max} and I_{max}) a PTC device will sustain without failure.

Upper category temperature:- Maximum ambient temperature at which a PTC thermistor can operate continuously.

Zero-power resistance, R_T:- The zero-power resistance is the dc resistance value of a thermistor measured at a specified temperature with a power dissipation by the thermistor low enough that any further decrease in power will result in not more than 0.1% (or 1/10 of the specified measurement tolerance, whichever is smaller) change in resistance.

Zero-power temperature coefficient of resistance, α_T:- The Zero-power temperature coefficient of resistance is the ratio at a specified temperature, *T*, of the rate of change of zero-power resistance with temperature to the zero-power resistance of the thermistor.

Glossary of Electrochemical Battery terminology

Absorption:- The retention of Hydrogen by the Misch Metal (Hydrogen-absorbing) alloys of the negative electrode.

Active Material:- Chemicals that give rise to electro-chemical reactions, and which generate electrical energy in the battery.

Ageing:- Permanent loss of capacity with frequent use or the passage of time due to unwanted irreversible chemical reactions in the cell.

AGM (Absorbed Glass Mat) battery:- A lead acid battery using a micro-glass mat (which also act as a separator) to promote recombination of the gases produced by the charging process.

AGM (Absorbed Glass Mat):- Micro-glass material used to contain the electrolyte and also function as a separator in a valve-regulated lead acid battery.

Alkaline Electrolyte:- An aqueous alkaline solution (such as potassium hydroxide) which provides a medium for the ionic conduction between the positive and negative electrodes of a cell.

Ampere (A):- A unit of electrical current or rate of flow of electrons. One volt across one ohm of resistance causes a current flow of one ampere. One ampere is equal to $6.235x10^{18}$ electrons per second passing a given point in a circuit.

Ampere hours (Ah):- The unit of measure used for comparing the capacity or energy content of a batteries with the same output voltage. For automotive (Lead Acid) batteries the SAE defines the Amp-hour capacity as the current delivered for a period of 20 hours until the cell voltage drops to 1.75 V.
Strictly - One Ampere hour is the charge transferred by one amp flowing for one hour. 1Ah = 3600 Coulombs. One C, 1C, means Ah current for 1 hour, ½C means current of half Ah for 2 hours, etc.

Ampere-Hour Capacity:- The number of ampere-hours that can be delivered by a storage battery under specified conditions as to temperature, rate of discharge and final voltage.

Ampere-Hour Efficiency:- The electrochemical efficiency of a storage battery expressed as the ratio of ampere-hours output to the ampere-hours input required for recharge.

Anode:- An electrode through which current enters any non-metallic conductor. The electrode in an electrochemical cell where oxidation takes place, releasing electrons. During discharge the negative electrode of the cell is the anode. During charge the situation reverses and the positive electrode of the cell is the anode.

Battery:- Two or more electrochemical cells enclosed in a container and electrically inter-connected in an appropriate series/parallel arrangement to provide the required operating voltage and current levels. Under common usage, the term battery also applies to a single cell if it constitutes the entire electrochemical storage system.

Battery Life:- End of Life. The period during which a cell or battery is capable of operating above a specified capacity or efficiency performance level. For example, with lead-acid batteries, end-of-life is generally taken as the point in time when a fully charged cell can deliver only 80% of its rated capacity. Beyond this state of aging, deterioration and loss of capacity begins to accelerate rapidly. Life may be measured in cycles and/or years, depending on the type of service for which the cell or battery is intended.

Burning Centre:- The centre-to-centre distance between adjacent plates of the same polarity.

C Rate:- The discharge or charge current in amperes, expressed in multiples of the rated capacity. For example, the C5 rate is the capacity in ampere hours available at the 5-hour discharge rate to a specified end voltage. A discharge of 0.5C5 is a discharge at 50% of the C5 rate.

Cadmium Electrode:- A third electrode in lead-acid battery for separate measurements of the electrode potential of positive and negative plate groups.

Capacity:- The amount of electrical energy that can be supplied by a cell/battery - expressed in Ah, and in specified discharge conditions.
Capacity Test:- A test that discharges the battery at constant current at room temperature to a cutoff voltage of usually 1.75V/cell in the case of a lead-acid battery.
Cathode:- An electrode through which current leaves any non-metallic conductor. The electrode in an electrochemical cell where reduction takes place, gaining electrons. During discharge the positive electrode of the cell is the cathode. During charge the situation reverses and the negative electrode of the cell is the cathode.
Cell (Primary):- A cell designed to produce electric current through an electrochemical reaction that is not efficiently reversible and hence the cell, when discharged, cannot be efficiently recharged by an electric current.
Cell (Storage):- An electrolytic cell for generation of electric energy, in which the cell after discharge may be restored to a charged condition by an electric current flowing in a direction opposite to the flow of current when the cell discharges.
Cell reversal:- A condition which may occur multi cell series chains in which an over discharge of the battery can cause one or more cells to become completely discharged. The subsequent volt drop across the discharged cell effectively reverses its normal polarity.
Charge acceptance:- quantifies the amount of electric charge which accumulates in a battery.
Charge Efficiency:- The ratio of the output of a cell during discharge to the input of a cell during charge. This ratio can be expressed in Efficiency of Capacity, Nominal Voltage, or Power.
Charge:- The operation which inputs electrical energy to a cell/battery.
Charge equalization:- brings all of the cells in a battery or string to the same state of charge.
Charge Rate:- The current applied to a cell to restore its capacity. The charge rate is usually expressed in terms of the cells C Rate.
Charge retention:- refers to a battery's ability to hold a charge. It diminishes during storage.
Charged and Dry:- A battery assembled with dry, charged plates and no electrolyte.
Charged and Wet:- A fully charged battery containing electrolyte and ready to deliver current.
Charge, state of:- Available or remaining capacity of a battery expressed as a % of the rated capacity.
Cold Cranking Amps:-A performance rating for automobile starting batteries. It is defined as the current that the battery can deliver for 30 seconds and maintain a terminal voltage greater than or equal to 1.20 volts per cell, at -18°C, when the battery is new and fully charged. Starting batteries may also be rated for Cranking Amps, which is the same thing but at a temperature of 0°C.
Constant Current Charge:- A charge that maintains the current at a constant value. For some types of batteries this may involve two rates, called a starting and a finishing rate. This procedure may damage the battery if performed on a repetitive basis.
Constant Potential Charge or Constant Voltage Charge:- A charge that holds the voltage at the terminals at a constant value and the current is limited only by the resistance of the battery and/or the capacity of the charge source.
Copper Contamination:- The formation of copper sulphate on the negative plates, usually caused by unintentional exposure of terminal posts' copper inserts to the electrolyte.
Coulombic Efficiency:- The ratio (expressed as a percentage) between the energy removed from a battery during discharge compared with the energy used during charging to restore the original capacity. Also called Charge Efficiency or Charge Acceptance.
Coup-de-Fouet:- The voltage dip followed by a subsequent voltage recovery that occurs when initially discharging a battery that has been on long-term float operation.
Cut-off Voltage:- A set voltage that determines when the discharging of a cell/battery should end.
Cycle:- A discharge and its subsequent recharge.
Cycle Life:- The total number of charge/discharge cycles before the battery reaches end of life (generally 80% of rated capacity).
Deep cycle battery - A battery designed to be discharged to below 80% Depth of Discharge. Used in marine, traction and EV applications.
Deep discharge - Discharge of at least 80% of the rated capacity of a battery.
Dendritic growth:- The formation from small crystals in the electrolyte of tree like structures which degrade the performance of the cell.
Depth of discharge DOD:- The ratio of the quantity of electricity or charge removed from a cell on discharge to its rated capacity discharge, expressed as a percent of rated capacity. For example, the removal of 25 ampere-hours from a fully charged 100 ampere-hours rated cell results in a 25% depth of discharge. Under certain conditions, such as discharge rates lower than that used to rate the cell, depth of discharge can exceed 100%.
Discharged:- A storage cell when, as a result of delivering current, in the case of the lead-acid cell, the plates are sulphated, the electrolyte is exhausted, and there is little or no potential difference between the terminals.
Discharge Factor:- A number equivalent to the time in hours during which a battery is discharged at constant current usually expressed as a percentage of the total battery capacity, i.e., C/5 indicates a discharge factor of 5 hours. Related to discharge rate.

Discharge Rate:- Any specified amperage rate at which a battery is discharged.
Efficiency:- The ratio of the output of a rechargeable cell or battery on discharge to the input required to restore it to the initial state of charge.
Electrochemical Cell:- A device containing two conducting electrodes, one positive and the other negative, made of dissimilar materials (usually metals) that are immersed in a chemical solution (electrolyte) that transmits positive ions from the negative to the positive electrode and thus forms an electrical charge. One or more cells constitute a battery.
Electrode:- Positive or negative plate containing materials capable of reacting with electrolyte to produce or accept current.
Electrode (Electrolyte) potential:- The voltage developed by a single electrode, determined by its propensity to gain or lose electrons. The difference in potential between the electrode and the immediately adjacent electrolyte, expressed in terms of a standard electrode potential difference.
Electrolysis:- Electrochemical reaction that causes the decomposition of a compound.
Electrolyte:- A substance which dissociates into ions (charged particles) when in aqueous solution or molten form and is thus able to conduct electricity. It is the medium which transports the ions carrying the charge between the electrodes during the electrochemical reaction in a battery.
End Gravity:- The specific gravity of a lead-acid cell at the end of a prescribed discharge.
Energy density:- The amount of energy stored in a battery. It is expressed as the amount of energy stored per unit volume or per unit weight (Wh/L or Wh/kg).
Equalisation:- The process of bringing every cell in a battery chain to the same state of charge (SOC)
Equalizing charge:- Charge applied to a battery which is greater than the normal float charge and is used to completely restore the active materials in the cell, bringing the cell float voltage and the specific gravity of the individual cells back to 'equal' values.
Fauré Plate:- see Pasted Plate.
Final Voltage:- The cut-off voltage of a battery. The prescribed voltage reached when the discharge is considered complete. Also known as end point voltage or EPV. This voltage is almost equivalent to limit of practical use. Typical values:

1.0 V per cell for NiCd and NiMH
1.75 V per cell for sealed lead acid
2.75 V per cell for lithium ion and lithium polymer
2.0 V per cell for primary lithium
0.9 V per cell for alkaline and carbon zinc

Finishing Rate:- The rate of charge, in amperes, to which charging current is reduced near the end of the charge for some types of batteries to prevent gassing and temperature rise.
Float Plate:- A pasted plate.
Float Charging:- A recharge at a very low rate, accomplished by connection to a bus whose voltage is slightly higher than the open circuit voltage of the battery. A method of maintaining a battery in a charged condition by continuous, long term, constant voltage charging at level sufficient to balance self-discharge.
Flooded Lead Acid cell:- In 'flooded' batteries, the oxygen created at the positive electrode is released from the cell and vented into the atmosphere. Similarly, the hydrogen created at the negative electrode is also vented into the atmosphere. This can cause an explosive atmosphere in an unventilated battery room. Furthermore the venting of the gasses causes a net loss of water from the cell. This lost water needs to be periodically replaced. Flooded batteries must be vented to prevent excess pressure from the build up of these gasses. *Sealed Lead Acid* (SLA) and *Valve Regulated Lead Acid* (VRLA) Cells overcome these problems.
Fuel Cell:- An electrochemical generator in which the reactants are stored externally and may be supplied continuously to a cell.
Gas Recombination:- The process by which oxygen gas generated from the positive plate during the final stage of charge is absorbed into the negative plate, preventing loss of water.
Gassing:- The generation or evolution of a gaseous product at one or both electrodes as a result of the electrochemical action. Gassing commonly results from local action (self discharge) or from the electrolysis of water in the electrolyte during charging. In lead acid batteries *gassing* produces hydrogen and oxygen. Significant gassing occurs when the battery is nearing the fully charged state while recharging or when the battery is on equalizing charge.
Gel cell:- An SLA battery which uses gelled electrolyte, an aqueous electrolyte that has been immobilised by the addition of a gelling agent.
Grid:- A metallic framework used in a battery for conducting electric current and supporting the active material.
Half Cell Reaction:- The electrochemical reaction between the electrode and the electrolyte.
Hydration (Lead):- Reaction between water and lead or lead compounds. Gravities lower than those found in discharged cells are apt to produce hydration, which appears as a white coating on plate groups and separators in a cell. A condition whereby lead dissolves into the electrolyte in a discharged cell and plates out onto the separator during recharge, resulting in numerous short circuit paths between the positive and negative plates.

Hydrometer:- A tool for testing the specific-gravity of a fluid, such as the electrolyte in a flooded battery. Typically a squeeze-bulb is used to suck up a sample of the fluid, and a float indicates the specific gravity.

Immobilized Electrolyte:- A lead-acid batteries technique where the electrolyte (the acid) is held in place against the plates instead of being a free-flowing liquid. The two most common techniques are gel and glass mat.

Impedance:- The resistive value of a battery to an AC current expressed in ohms (W). Generally measured when fully charged, at 1000 Hz.

Intercalation:- This insertion of ions into the crystalline lattice of a host electrode without changing its crystal structure. A reaction where lithium Ions are reversibly removed or inserted into a host without a significant structural change to the host.

Internal Pressure:- The pressure within a sealed cell caused by oxygen or hydrogen evolution.

Internal Resistance:- The opposition or resistance to the flow of a direct electric current within a cell or battery; the sum of the ionic and electronic resistance of the cell components. Its value varies with the current, state of charge, temperature, and age. With an extremely heavy load, such as an engine starter, the cell voltage may drop significantly. This voltage drop is due to the internal resistance of the cell. A cell that is partly discharged has a higher internal resistance than a fully charged cell, hence it will have a greater voltage drop under the same load. This change in internal resistance is due to the accumulation of lead sulphate in the plates.

Interstitial:- A space between things closely set, or between the parts, which compose a body; a narrow chink; a crack; a crevice; a hole.

Lithium Cobaltite:- ($LiCoO_2$) Dark blue powder; insoluble in water. The compound exhibits both the fluxing property of lithium oxide and the adherence-promoting property of cobalt oxide. Intercalates lithium ions in battery applications.

Manchex:- A type of Planté cell in which the positive plate is cast with openings provided for the active material, which are buttons of soft-lead ribbon. The active material is corrugated and rolled into spirals, which are forced into the grids by hydraulic pressure.

Memory effect (Voltage Depression):- Reversible, progressive capacity loss in nickel based batteries found in NiCad and to a lesser extent in NiMH batteries. It is caused by a change in crystalline formation from the desirable small size to a large size which occurs when the cell is repeatedly recharged before it is fully discharged.

Metal Hydride (MH):- The negative electrode composed of Misch metal (Hydrogen-storing) alloys.

MF (Maintenance Free Battery):- A VRLA sealed absorbed glass mat (AGM) battery.

Microporous Separator:- A veneer or grooved-type separator made of any material that has many microscopically small pores.

Migration:- The movement of charged ions under the influence of a potential gradient.

Misch Metal (M):- The matrix of the negative electrode composed of Hydrogen-storing alloys.

Nickel Metal Hydride (NiMH): A cell or battery system composed of a Nickel (Ni) positive electrode and a metal hydride (MH) negative electrode.

Negative Plate:- The grid and active material that current flows to from the external circuit when a battery is discharging.

Negative Terminal:- The terminal from which current flows through the external circuit to the positive terminal when the cell discharges.

Nernst equation:- Used by cell designers to calculate the voltage of a chemical cell from the standard electrode potentials, the temperature and to the concentrations of the reactants and products.

Nominal Voltage:- A general value to indicate the voltage of a battery in application.

Open Circuit Voltage:- The voltage of a battery when it is not delivering or receiving power, and has been at rest long enough to reach a steady state (normally, at least 4 hours).

Overcharge:- The forcing of current through a cell after all the active material has been converted to the charged state. In other words, charging continued after 100% state of charge is achieved. The result will be the decomposition of water in the electrolyte into hydrogen and oxygen gas, heat generation, and corrosion of the positive electrode.

Oxygen Recombination:- The process in which oxygen generated at the positive electrode during over-charge reacts with hydrogen at the negative electrode to produce water.

Pasted (Fauré) Plate:- A plate consisting of a grid filled with active material applied as a paste.

Peukert's equation:- A formula that shows how the available capacity of a lead-acid battery changes according to the rate of discharge. The capacity of a battery is expressed in Amp-Hours, but it turns out that the simple formula of current times hours does not accurately represent the situation. Peukert found that the equation:

$$C = I^n \times t$$

fits the observed behaviour of batteries. '*C*' is the theoretical capacity of the battery, I is the current, t is time, and n is the Peukert number, a constant for the given battery. The equation captures the fact that at higher currents, there is less available energy in the battery.

Peukert number:- A value that indicates how well a lead-acid battery performs under heavy currents. The Peukert number is the exponent in Peukert's equation. A value close to 1 indicates that the battery performs well; the higher the number, the more capacity is lost when the battery is discharged at high currents. The Peukert number of a battery is determined empirically.

Planté Plate:- A formed lead plate of large area, the active material of which is formed directly from a lead substrate.

Polarization:- Change in voltage at terminals when a specified current is flowing; equal to the difference between the actual and the equilibrium (constant open circuit condition) potentials of the plates, exclusive of the internal resistance drop.

Positive Plates:- The grid and active materials of a storage battery from which current flows to the external circuit when the battery is discharging.

Positive Terminal:- The terminal that current flows toward in the external circuit from the negative terminal.

Potassium Hydroxide (KOH):- The electrolyte provides the ion transport mechanism between the electrodes, used in NiMH cells.

Primary cell:- A cell that is non-rechargeable. A cell or battery that can be discharged only once.

Prismatic cell:- A slim rectangular sealed cell in a metal case. The positive and negative plates are stacked usually in a rectangular shape rather than rolled in a spiral as done in a cylindrical cell.

Rapid Charge:- A rate of charging a cell or battery that results in fully charging a battery to full capacity between 2½ to 6 hours.

Rated Capacity:- Ampere hours of discharge that can be removed from a fully charged cell or battery, at a specific constant discharge rate at a specified temperature and at a specified cut-off voltage.

Recombinant system:- Sealed secondary cells in which gaseous products of the electrochemical charging cycle are made to recombine to recover the active chemicals. A closed cycle system preventing loss of active chemicals. Used in NiCd and *SLA* batteries.

Resealable Safety Vent:- The resealable vent built into cylindrical and prismatic cells to prevent the build up of high internal pressures.

Reversal:- A change in the normal polarity of a cell or battery.

Safety Vent:- This is a device to release the gas when the internal pressure of the battery exceeds the pre-set value.

Sealed cells:- A cell which remains closed and does not release gas or liquid when operated within the limits of charge and temperature specified by the manufacturer. An essential component in *recombinant* cells.

Secondary cell:— the process is reversible so that charging and discharging may be repeated over and over.

Sediment:- The sludge or active material shed from plates that drops to the bottom of cells.

Sediment Space:- The portion of a container beneath the element; sediment from the wearing of the plates collects here without short-circuiting.

Self-discharge:- Loss of charge due to local action, without external current flow. The decrease in the state of charge of a cell or a battery, over a period of time during storage or not in use, due to internal electrochemical losses. Typical values:

1% per day for NiCd
2% per day for NiMH
~0% per day for Lithium Ion and Lithium Polymer

Self Discharge Rate:- the percent of capacity lost on open circuit over a specified period of time.

Separator:- A device in a storage battery that prevents metallic contact between plates of opposite polarity in a cell. In sealed lead acid batteries it normally is absorbent glass fibre to hold the electrolyte in suspension.

Shelf Life:- The duration under specified conditions that at the end a cell or battery can be stored and retain its performance.

SLA Battery:- Sealed Lead Acid battery. In sealed batteries the generated oxygen combines chemically with the lead and then the hydrogen at the negative electrode, and then again with reactive agents in the electrolyte, to recreate water. A *recombinant system.* The net result is no significant loss of water from the cell.

Spalling:- Shedding of active material, usually from positives, during formation due to incomplete or improper plate curing.

Sponge Lead (Pb):- A porous mass of lead crystals and the chief material of a full-charged negative plate.

Standby Service:- An application in which the battery is maintained in a fully charged condition by trickle of float charging.

State of Charge:- The amount of electrochemical energy left in a cell or battery. The available ampere hours in a battery at any given time relative to its full charge capacity.

Starved Electrolyte:- A term occasionally applied to a VRLA cell, meaning that the cell contains little or no free electrolyte.

Sulphation:- Growth of lead sulphate crystals in Lead-Acid batteries which inhibits current flow. Refers to the formation of hard lead sulphate crystals in the plates that are difficult, if not impossible, to reconvert to active material. Sulphation is caused by storage at low state of charge.

Stratification:- Layering of high specific gravity electrolyte in lower portions of a cell, where it does not circulate normally and is of no use.
Temperature Correction:- In storage cells, specific gravity and charging voltage vary inversely with temperature, while the open circuit voltage varies directly though slightly with temperature.
Thermal Runaway:- A condition in which a cell or battery (especially valve-regulated types) on constant potential charge can destroy itself through internal heat generation being greater than that which can be externally dissipated. Can cause failure through cell dry-out, shortened life, and/or melting of the battery.
Treeing:- Growth of a lead dendrite or filament through a crack or hole of a separator, short-circuiting the cell.
Trickle Charge:- A low-rate continuous charge approximately equal to a battery's internal losses and capable of maintaining the battery in a fully-charged state. Method of charging in which the battery is either continuously or intermittently connected to a constant current charging source to maintain the battery in a fully charged condition. Not recommended for use with AGM batteries.
Tubular Plate:- A plate in which the active material is contained in porous tubes, each tube having a centrally located grid.
Vent:- An opening that permits the escape of gas from a cell or mould.
Venting:- A release of gas either controlled (through a vent) or accidental from a battery cell.
Vent Valve:- A normally closed check valve located in a cell which allows the controlled escape of gases when the internal pressure exceeds its rated value.
Volt Efficiency:- The ratio of the average voltage of a cell or battery during discharge to the average voltage during subsequent recharge.
VRLA (Valve Regulated Lead Acid):- Sealed batteries which feature a safety valve venting system designed to release excessive internal pressure, while maintaining sufficient pressure for recombination of oxygen and hydrogen into water.
Watthour:- A unit of electrical energy or work, equal to one watt acting for one hour.
Watthour Capacity:- The number of watthours a storage battery can deliver under specific conditions of temperature, rate of discharge and final voltage.
Watthour Efficiency:- A storage battery's energy efficiency expressed as ratio of watthour output to the watthours of the recharge.
Wet Shelf Life:- The time a wet secondary cell can be stored before its capacity falls to the point that the cell cannot be easily recharged.

Glossary of Fuel Cell terminology

Alkali:- A chemical base produces negative ions, the opposite of an acid. Certain types of alkalis (especially potassium hydroxide) are used as fuel cell electrolytes.
Alkaline Fuel Cell (AFC):- A type of hydrogen/oxygen fuel cell in which the electrolyte is concentrated KOH (varies between 35 to 85 wt% depending on the intended operating temperature) and hydroxide ions(OH^-) are transported from the cathode to the anode. Temperature of operation can vary from <120°C to approximately 250°C depending upon electrolyte concentration.

Anode reaction: $2H_2 + 4OH^- \rightarrow 4H_2O + 4e^-$
Cathode reaction: $O_2 + 2H_2O + 2e^- \rightarrow 4OH^-$
Overall reaction: $2H_2 + O_2 \rightarrow 2H_2O$

Anion:- A negative ion. Alkali, molten carbonate and solid oxide fuel cells are anion-mobile cells - anions migrate through the electrolyte toward the anode.
Anode:- One of two electrodes in a fuel cell or battery. In a fuel cell it is where the fuel reacts or oxidizes, and releases electrons, that is, where the chemical reaction produces positive ions. For cells that create potential, it is also the electrode towards which the negative ion flows.
Biomass:- All organic substances: plants, wood chips, bales of straw, liquid manure, organic wastes, etc.
Bipolar plates:- Electrical conductive plate in a fuel cell stack that acts as an anode for one cell and a cathode for the adjacent cell. The plate may be made of metal or a conductive polymer (which may be a carbon-filled composite). The plate usually incorporates flow channels for the fluid feeds and may also contain conduits for heat transfer.
Catalyst:- A substance that causes or speeds a chemical reaction, by lowering the amount of energy needed to cause the reaction, without itself being affected. The catalyst lowers the activation energy required, allowing the reaction to proceed more quickly or at a lower temperature. In a fuel cell, there will typically be a catalyst used for the electrodes (to break down hydrogen into electrons and protons). Catalysts are also often used in reforming fuel.
Catalyst loading:- This is related to the amount of catalyst used in a fuel cell or fuel cell system. It often refers specifically to the mass of catalyst per unit area of an electrode.
Cathode:- One of two electrodes in a fuel cell or battery. In a fuel cell, it is where oxygen (usually taken from the air) reduction occurs – electrode where negative ions are produced.

Cation:- A positive ion. Phosphoric acid and PEM fuel cells are cation-mobile cells – the cation migrates through the electrolyte toward the cathode.
CHP:- Combined Heat and Power. This is the additional production of electricity from processes which otherwise produce only space heat or domestic hot water (DHW); also known as cogeneration.
Cogeneration:- The simultaneous use of waste heat from industrial processing, a steam turbine, or a fuel cell to generate electricity. Harnessing otherwise wasted heat boosts the efficiency of power-generating systems.
Cryoadsorption storage:- (Greek krýos: cold, frost.) special type of graphite storage. Carbon is able to adsorb hydrogen. Different qualities of carbon can adsorb higher quantities of hydrogen under certain temperature and pressure conditions than could be stored without the carbon under the same conditions. Temperatures are below 0°C (cryogenic) and above boiling temperature of hydrogen (20 K). The pressure levels are above 5 MPa.
Desulphuriser:- Some fuels contain sulphur which can be damaging to fuel cell performance. A desulphuriser may therefore be used to remove sulphur from the gas stream entering the fuel cell stack and maintain peak electrical output.
Diffusion:- Diffusion is the movement of a chemical, usually under a pressure differential. In fuel cells, diffusion may happen through a palladium membrane to purify hydrogen or through the fuel cell electrodes before splitting of a hydrogen molecule.
Direct Fuel Cell:- A type of fuel cell in which a hydrocarbon fuel is fed directly to the fuel cell stack, without requiring an external reformer to generate hydrogen.
Direct Methanol Fuel Cell (DMFC):- A type of fuel cell in which the fuel is methanol (CH_3OH), in gaseous or liquid form. The methanol is oxidized directly at the anode with no reformation to hydrogen. The electrolyte is typically a PEM.

Anode reaction: $2CH_3OH + 2H_2O \rightarrow 2CO_2 + 12H^+ + 12e^-$
Cathode reaction: $12H^+ + 3O_2 + 12e^- \rightarrow 6H_2O$
Overall reaction: $2CH_3OH + 2H_2O + 3O_2 \rightarrow 2CO_2 + 6H_2O$

Distributed generation:- Distributed generation involves the small-scale production of electrical power much closer to the end user than conventional power supply does. Distributed generation often requires lower power units.
Electrode:- An electrical terminal that conducts an electric current into or out of a fuel cell. The electrode is where reaction of a chemical species occurs and electrons are either released or accepted.
Electrolyte:- A chemical compound that conducts ions from one electrode to the other inside a fuel cell. The electrolyte does not react with the ions and does not conduct free electrons.
Electrolyser:- In an electrolyser, an electric current splits water into hydrogen and oxygen. Reverse process of the fuel cell.
External reforming:- External reforming occurs where a fuel is reformed to hydrogen - hydrocarbon fuel (methanol, gasoline, natural gas, propane, etc.) prior to entering a fuel cell stack.
Fuel:- A fuel is a chemical which can be used in a fuel cell system to produce electricity. The fuel is typically either hydrogen or something which can produce hydrogen when reformed.
Fuel Cell:- A electrochemical device for generating continuous electricity by the chemical combination a fuel and oxygen or oxidant, without combustion. A fuel cell will continuously produce electricity as long as fuel is supplied to it. Reverse process of electrolyser.
Fuel processor:- A fuel processor is a device that is capable of reforming a fuel to produce a gas stream containing hydrogen and then clean this up to produce a gas flow of sufficiently high quality to be used as the input for a fuel cell stack.
Graphite:- A soft form of the element carbon. It is used as a lubricant, as a moderator in nuclear reactors, and for other products. It does not burn easily or fuse at high temperatures, and is an important material in the construction of phosphoric acid fuel cells. Carbon is able to adsorb hydrogen. The amount of adsorbed hydrogen depends on temperature, pressure and the quality/ structure of the carbon used. Carbon structures in the nanometre range (one nanometre corresponds to 10^{-9} meters), e.g. balls, tubes or fibres
Grid-connected:- A grid-connected fuel cell is designed to function when connected to the electrical grid.
Hydrocarbon:- A chemical compound consisting of hydrogen and carbon formed in a variety of bond structures, such as oil, methane, propane, butane, etc. These are often used as fuels.
Hydrogen:- H_2. A chemical element consisting of one proton and one electron. Two hydrogen atoms combine with one oxygen atom to form a molecule of water. Hydrogen serves as the fuel for most fuel cells.
Internal reforming:- Some fuel cells operate at sufficiently high temperatures to be able to internal convert a hydrocarbon fuel to hydrogen within the fuel cell stack.
Ion:- An atom that carries a positive or negative charge due to the loss or gain of an electron.
IR Loss (Ohmic Polarization):- Losses created by the resistance to the flow of ions in the electrolyte and resistance to flow of electrons through the electrode and bipolar plate materials. Because both the electrolyte and fuel cell electrodes obey Ohm's law, the ohmic losses can be expressed by the equation *V=IR*
I^2R *Loss*:- Power loss due to the current *I* flow through the resistance R of a conductor.

Islanding:- Operation of a separate non-utility power source with or without a portion of an electric utility system- isolated from the remainder of the utility system. When a fuel cell is grid-connected, islanding of the fuel cell is required to allow safe work on the grid.

kWh:- Kilowatt-hour (1,000 watts for one hour). A measure of electric power consumption.

Matrix:- A framework within a fuel cell that supports an electrolyte.

Membrane:- The separating layer in a fuel cell that acts as electrolyte (a ion-exchanger) as well as a barrier film separating the gases in the anode and cathode compartments of the fuel cell.

Metal hydride storage:- Device that can store hydrogen by use of a metal alloy. The hydrogen is soaked into the alloy like into a sponge and fills the spaces in the crystal lattice of the alloy. The storage is filled applying a modest over-pressure and is usually operated in the temperature range of 20-80°C.

MPa:- mega Pascals (SI pressure unit); one MPa corresponds to a pressure of 10 atmospheres (10 bars).

Molten Carbonate:- A type of fuel cell electrolyte that contains carbon, oxygen and another element. Solid at room temperature, it must be melted in order to function.

Molten Carbonate Fuel Cell (MCFC):- A type of fuel cell consisting of a molten electrolyte of Li_2CO_3/ Na_2CO_3 in which the species CO_3^{2-} is transported from the cathode to the anode. Operating temperatures are typically near 650°C.

Anode reaction: $2H_2 + 2CO_3^{2-} \rightarrow 2H_2O + 2CO_2 + 4e^-$
Cathode reaction: $O_2 + 2CO_2 + 4e^- \rightarrow 2CO_3^{2-}$
Overall reaction: $2H_2 + O_2 \rightarrow 2H_2O$

Nafion:- A sulphuric acid in a solid polymer form. It is usually the electrolyte of PEM fuel cells.

Outage:- An outage occurs when a fuel cell or other power source which is producing electricity fails.

Oxygen:- O_2. A chemical diatomic element consisting of eight protons, eight neutrons and eight electrons. Two hydrogen atoms combine with one oxygen atom to form a molecule of water.

Phosphoric Acid:- A solution of the elements phosphorus, hydrogen, and oxygen that serves as the electrolyte for one type of fuel cell. Chemically- H_3PO_4.

Phosphoric Acid Fuel Cell (PAFC):- A type of fuel cell in which the electrolyte consists of concentrated phosphoric acid (H_3PO_4) and protons (H^+) are transported from the anode to the cathode. The operating temperature range is generally 160 - 220°C.

Anode reaction: $2H_2 \rightarrow 4H^+ + 4e^-$
Cathode reaction: $O_2 + 4H^+ + 4e^- \rightarrow 2H_2O$
Overall reaction: $2H_2 + O_2 \rightarrow 2H_2O$

Photobiological water splitting:- A biological processes that liberates hydrogen or where hydrogen is produced as an intermediate product. For example, photosynthesis use the solar radiation as source of energy, while fermentation processes that take place in the absence of light take advantage of the energy stored in the feedstock (e.g. glucose).

Polymer:- A natural or synthetic compound composed of repeated links of simple molecules.

Potassium Hydroxide:- A solution of the elements potassium, hydrogen, and oxygen that serves as the electrolyte for one type of fuel cell. Chemically:- KOH.

Power density:- The power density of an individual fuel cell is the power produced related to the active area or volume of the cell.

Proton Exchange Membrane (PEM):- A polymer sheet that serves as the electrolyte in PEM fuel cell. The film prevents hydrogen and oxygen meeting and also carries protons across to complete the electrical circuit.

Proton Exchange Membrane Fuel Cell (PEMFC):- A type of acid based fuel cell in which the exchange of protons (H^+) from the anode to the cathode is achieved by a solid, aqueous membrane impregnated with an appropriate acid. The electrolyte is a called a proton-exchange membrane (PEM). The fuel cells typically run at low temperatures (<100°C) and pressures (< 5 atm).

Anode reaction: $2H_2 \rightarrow 4H^+ + 4e^-$
Cathode reaction: $O_2 + 4H^+ + 4e^- \rightarrow 2H_2O$
Overall reaction: $2H_2 + O_2 \rightarrow 2H_2O$

Reformate:- Fuel processor output gas stream containing hydrogen, carbon monoxide and carbon dioxide. This reformate gas stream will eventually pass to the fuel cell stack, possibly after purification.

Reformer:- A device that extracts pure hydrogen from hydrocarbons which have reacted with water vapour and heat in the presence of a catalyst.

Reforming:- The process of producing a hydrogen-rich gas stream for eventual use in a fuel cell from a feedstock. The thermal or catalytic conversion of a hydrocarbon fuel into more volatile products with higher calorific ratings.

Regenerative Fuel Cells:- A regenerative (or reversible) fuel cell is able to react a fuel and an oxidant to produce electricity and other chemical species or operate in reverse. This allows ready production of power when it is economically viable. Several fuel cell types in which fuel and, in some types, the oxidant are regenerated from the oxidation product.

Renewable energy:- Energy sources that do not require the use of limited fossil fuel resources. They include wind power, hydroelectric or geothermal power and photovoltaics. They can often be used to produce hydrogen for use in fuel cells.

Reversible fuel cell:- A reversible, or regenerative, fuel cell is able to react a fuel and an oxidant to produce electricity and other chemical species or operate in reverse, such that the cell may be recharged with a separate power source if desired. Where hydrogen and water are the fuels, water and electricity are produced. When required, water can be electrolysed, and hydrogen and oxygen produced, upon the input of electricity. For example, the hydrogen/oxygen fuel cell may be recharged by providing power for water electrolysis with hydrogen storage. Also called a Regenerative Fuel Cell.

Shift conversion:- The reaction of carbon monoxide CO and water to give hydrogen and carbon dioxide. This provides more hydrogen to create power in the fuel cell and removes carbon monoxide which is detrimental to some types of fuel cell. This process is performed immediately after the reformer and before the preferential oxidizer to reduce CO from approximately 10% down to 0.5% to 0.1% usually through a water gas shift reaction.

Solid Oxide:- A solid combination of oxygen and another element (often zirconium) that serves as the electrolyte for a particular fuel cell.

Solid Oxide Fuel Cell (SOFC):- A type of fuel cell in which the electrolyte is a solid, nonporous metal oxide, typically ZrO_2 doped with Y_2O_3, and O^{2-} is transported from the cathode to the anode. Any carbon monoxide (CO) in the reformate gas is oxidized to carbon dioxide (CO_2) at the anode. Temperatures of operation are typically 800 to 1000°C.

Anode reactions: $H_{2(g)} + O^{2-} \rightarrow H_2O_{(g)} + 2e^-$
$2CO_{(g)} + O^{2-} \rightarrow CO_{2(g)} + 2e^-$
Cathode reaction: $O_2 + 4e^- \rightarrow 2O^{2-}$
Overall reaction: $O_2 + H_2 + CO \rightarrow H_2O + CO_2$

Specific power:- The specific power of a system is the power produced divided by the weight of the system.

Stack:- arrangement of individual fuel cells connected in series within a generating assembly.

Standard Conditions:- The performance of most fuel cells will be quoted under standard conditions in order to allow easy comparison.

Steam reforming:- The catalytic reaction of a hydrocarbon fuel with water to produce hydrogen. More hydrogen is produced for the same amount of fuel than by auto thermal reforming or partial oxidation but heat input is required to maintain the reaction.

Tubular cells:- The two common designs of solid oxide fuel cells are tubular and planar. A tubular system separates the fuel or reformate from the oxidant, inside and outside the tube. Fuel Cells that are formed in cylindrical fashion and allow fuel and oxidant to flow on the inner or outer surfaces of the pipe.

Water-gas shift reaction:- This reaction between carbon monoxide and water produces hydrogen and carbon dioxide. It is therefore used after the fuel has been reformed to provide more hydrogen to power a fuel cell and to remove carbon monoxide which may poison performance.

Glossary of Solar Electric terminology

Absorbers:- Dark-coloured objects that soak up heat in thermal solar collectors. In a photovoltaic device, the material that readily absorbs photons to generate charge carriers (free electrons or holes).

Amorphous semiconductor:- A non-crystalline semiconductor material. Easier and cheaper

Amorphous Silicon:- A thin-film, silicon photovoltaic cell having no crystalline structure. Manufactured by depositing layers of doped silicon on a substrate. See also single-crystal silicon an polycrystalline silicon.

Angle of incidence:- Angle between the normal to a surface and the direction of incident radiation; applies to the aperture plane of a solar collector. Most modern solar panels have only minor reductions in power output within plus/minus 15°.

Antireflection coating:- A thin coating of a material, which reduces the light reflection and increases light transmission, applied to a photovoltaic cell surface.

Array:- Any number of photovoltaic modules connected together to provide a single electrical output. Arrays are often designed to produce significant amounts of electricity.

Autonomous system:- A stand-alone PV system that has no back-up generating source. May or may not include storage batteries. Most battery systems are designed for a certain minimum "days of autonomy" - which means that the batteries can supply sufficient power with no sunlight to charge the batteries. This varies from 3 to 5 days in the sunbelt, to 5 to 10 days elsewhere.

Azimuth:- Angle between the north direction and the projection of the surface normal into the horizontal plane; measured clockwise from north. As applied to the PV array, 180° azimuth.

Band Gap:- In a semiconductor, the energy difference between the highest valence band and the lowest conduction band.

Band Gap Energy (E_g):- The amount of energy (in electron volts) required to free an outer shell electron from its orbit about the nucleus to a free state, and thus promote it from the valence to the conduction level.

Barrier Energy:- The energy given up by an electron in penetrating the cell barrier; a measure of the electrostatic potential of the barrier.

Baseline performance value:- Initial values of I_{sc}, V_{oc}, P_{mp}, I_{mp} measured by the accredited laboratory and corrected to Standard Test Conditions.

Cathodic protection:- A method of preventing oxidation (rusting) of exposed metal structures, such as bridges and pipelines, by imposing between the structure and the ground a small electrical voltage that opposes the flow of electrons and that is greater than the voltage present during oxidation.

Cell:- The basic unit of a photovoltaic panel or battery.

Cell barrier:- A very thin region of static electric charge along the interface of the positive and negative layers in a photovoltaic cell. The barrier inhibits the movement of electrons from one layer to the other, so that higher-energy electrons from one side diffuse preferentially through it in one direction, creating a current and thus a voltage across the cell. Also called depletion zone, cell junction, or space charge.

Cell junction:- The area of immediate contact between two layers (positive and negative) of a photovoltaic cell. The junction lies at the centre of the cell barrier or depletion zone.

Chemical Vapour Deposition (CVD):- A method of depositing thin semiconductor films used to make certain types of photovoltaic devices. With this method, a substrate is exposed to one or more vaporized compounds, one or more of which contain desirable constituents. A chemical reaction is initiated, at or near the substrate surface, to produce the desired material that will condense on the substrate.

Cleavage of Lateral Epitaxial Films for Transfer (CLEFT):- A process for making inexpensive Gallium Arsenide (GaAs) photovoltaic cells in which a thin film of GaAs is grown atop a thick, single-crystal GaAs (or other suitable material) substrate and then is cleaved from the substrate and incorporated into a cell, allowing the substrate to be reused to grow more thin-film GaAs.

Combined collector:- A photovoltaic device or module that provides useful heat energy in addition to electricity.

Concentrator:- A PV module that uses optical elements to increase the amount of sunlight incident on a PV cell. Concentrating arrays must track the sun and use only the direct sunlight because the diffuse portion cannot be focused onto the PV cells. Efficiency is increased, but lifespan is usually decreased due to the high heat.

Concentrator (module, array, or collector):- An arrangement of photovoltaic cells that includes optical components such as lenses (Fresnel lens) to direct and concentrate sunlight onto a PV cell of smaller area. Concentrators can increase the power flux of sunlight hundreds of times.

Conduction Band (or conduction level):- An energy band in a semiconductor in which electrons can move freely in a solid, producing a net transport of charge.

Conversion efficiency (cell or module):- The ratio of the electric energy produced by a photovoltaic device (under one-sun conditions) to the energy from sunlight incident upon the cell.

Copper Indium Diselenide ($CuInSe_2$, or CIS):- A polycrystalline thin-film photovoltaic material (sometimes incorporating gallium (CIGS) and/or sulphur).

Crystalline Silicon:- A type of photovoltaic cell made from a slice of single-crystal silicon or poly-crystalline silicon.

Current at maximum power (I_{mp}):- The current at which maximum power is available from a module.

Cycle life:- Number of discharge-charge cycles that a battery can tolerate under specified conditions before it fails to meet specified criteria as to performance (e.g., capacity decreases to 80% of the nominal capacity).

Dangling Bonds:- A chemical bond associated with an atom on the surface layer of a crystal. The bond does not join with another atom of the crystal, but extends in the direction of exterior of the surface.

Dendrite:- A slender threadlike spike of pure crystalline material, such as silicon.

Dendritic Web Technique:- A method for making sheets of polycrystalline silicon in which silicon dendrites are slowly withdrawn from a melt of silicon whereupon a web of silicon forms between the dendrites and solidifies as it rises from the melt and cools.

Depletion Zone:- Same as cell barrier. The term derives from the fact that this microscopically thin region is depleted of charge carriers (free electrons and hole).

Diffuse insolation:- Sunlight received indirectly as a result of scattering due to clouds, fog, haze, dust, or other obstructions in the atmosphere. Opposite of direct insolation.

Direct insolation:- Sunlight falling directly upon a collector. Opposite of diffuse insolation.

Direct Beam Radiation:- Radiation received by direct solar rays. Measured by a pyrheliometer with a solar aperture of 5.7° to transcribe the solar disc.

Distributed systems:- Systems that are installed at or near the location where the electricity is used, as opposed to central systems that supply electricity to grids. A residential

Edge-Defined Film-Fed Growth (EFG):- A method for making sheets of polycrystalline silicon for photovoltaic devices in which molten silicon is drawn upward by capillary action through a mould.

Efficiency:- The ratio of power output of a Photovoltaic cell to the incident power from the sun or simulated sun sources under specified standard insolation conditions.

Electrodeposition:- Electrolytic process where a metal is deposited at the cathode from a solution of its ions.

Energy:- The ability to do work. Stored energy becomes working energy when we use it.

Energy Levels:- The energy represented by an electron in the band model of a substance.

Epitaxial Growth:- The growth of one crystal on the surface of another crystal. The growth of the deposited crystal is oriented by the lattice structure of the original crystal.

EVA:- (Ethylene Vinyl Acetate) An encapsulant between the glass cover and the PV cells in PV modules. It is durable, transparent, resistant to corrosion, and flame retardant.

Fill Factor:- The ratio of a photovoltaic cell's actual power to its power if both current and voltage were at their maxima. A key characteristic in evaluating cell performance.

Fixed Tilt Array:- A photovoltaic array set in at a fixed angle with respect to horizontal.

Flat-plate PV:- Refers to a PV array or module that consists of non-concentrating elements. Flat-plate arrays and modules use direct and diffuse sunlight, but if the array is fixed in position, some portion of the direct sunlight is lost because of oblique sun-angles in relation to the array.

Full Sun:- The full sun condition is the amount of power density received at the surface of the earth at noon on a clear day - about 100 mW/cm^2. Lower levels of sunlight are often expressed as ½sun or 0.1 sun. A figure of 0.5 sun means that the power density of the sunlight is one-half of that of a full sun.

Fresnel Lens:- An optical device that focuses light like a magnifying glass; concentric rings are faced at slightly different angles so that light falling on any ring is focused to the same point.

Gallium (Ga):- A chemical element, metallic in nature, used in making certain kinds of PV cells and semiconductor devices.

Gallium Arsenide (GaAs):- A crystalline, high-efficiency compound used to make certain types of PV cells and semiconductor material.

Grid-connected (PV system):- A PV system in which the PV array acts like a central generating plant, supplying power to the grid.

Heterojunction:- A region of electrical contact between two different materials.

Hole:- The vacancy where an electron would normally exist in a solid; behaves like a positively charged particle.

Homojunction:- Region between n-layer and p-layer in a single material, photovoltaic cell.

Incident light:- Light that shines onto the face of a PV cell or module.

Indium Oxide:- A wide band gap semiconductor that can be heavily doped with tin to make a highly conductive, transparent thin film. Often used as a front contact or one component of a heterojunction PV cell.

Infrared Radiation:- Electromagnetic radiation whose wavelengths lie in the range from 0.75 micrometer to 1000 micrometers; invisible long wavelength radiation (heat) capable of producing a thermal or photovoltaic effect, though less effective than visible light.

Irradiance:- The direct, diffuse, and reflected solar radiation that strikes a surface. Usually expressed in kW/m^2. Irradiance multiplied by time equals insolation.

Insolation:- Sunlight, direct or diffuse; from 'incident solar radiation.'

Interconnect:- A conductor within a module or other means of connection which provides an electrical interconnection between the PV cells. [UL 1703]

I-V curve:- A graphical presentation of the current versus the voltage from a photovoltaic device as the load is increased from the short circuit (no load) condition to the open circuit (maximum voltage) condition. The shape of the curve characterized cell performance.

I-V data:- The relationship between current and voltage of a photovoltaic device in the power-producing quadrant, as a set of ordered pairs of current and voltage readings in a table, or as a curve plotted in a Cartesian coordinate system. [ASTM E 1036]

Junction diode:- A semiconductor device with a junction and a built-in potential that passes current better in one direction than the other. All PV cells are junction diodes.

kilowatt-hour (kWh):- One thousand watts acting over a period of 1 hour. The kWh is a unit of energy. 1 kWh=3600 kJ.

Light-induced defects:- Defects, such as dangling bonds, induced in an amorphous silicon semiconductor upon initial exposure to light.

Light trapping:- The trapping of light inside a semiconductor material by refracting and reflecting the light at critical angles; trapped light will travel further in the material, greatly increasing the probability of absorption and hence of producing charge carriers.

Maximum Power Point (MPP):- The point on the current-voltage (I-V) curve of a module under illumination, where the product of current and voltage is maximum. [UL 1703] For a typical silicon

cell panel, this is about 17 volts for a 36-cell configuration. MPP tracking will typically increase power delivered to the system by 10% to 40%, depending on climate conditions and battery state of charge. For a typical silicon cell, this is at about 0.45 volts.

Microgroove:- A small groove scribed into the surface of a cell which is filled with metal for contacts.

Module:- A number of PV cells connected together, sealed with an encapsulant, and having a standard size and output power; the smallest building block of the power generating part of a PV array. Also called panel.

Monolithic:- Fabricated as a single structure.

Multicrystalline:- Material that is solidified at such as rate that many small crystals (crystallites) form. The atoms within a single crystallite are symmetrically arranged, whereas crystallites are jumbled together. These numerous grain boundaries reduce the device efficiency. A material composed of variously oriented, small individual crystals. (Sometimes referred to as polycrystalline or semicrystalline).

Multijunction device:- A photovoltaic device containing two or more cell junctions, each of which is optimized for a particular part of the solar spectrum, to achieve greater overall efficiency.

n-type semiconductor:- A semiconductor produced by doping an intrinsic semiconductor with an electron-donor impurity (e.g., phosphorous in silicon).

NOCT:- Nominal Operating Cell Temperature. The PV cell temperature at a reference environment defined as 800 W/m^2 irradiance, 20°C ambient air temperature, and 1 m/s wind speed with the cell or module in an electrically open circuit state.

Open-circuit voltage (Voc):- The maximum possible voltage across a photovoltaic cell or module; the voltage across the cell in sunlight when no current is flowing.

Peak load; peak demand:- The maximum load, or usage, of electrical power occurring in a given period of time, typically a day.

Peak Sun Hours:- The equivalent number of hours per day when solar irradiance averages 1,000 W/m^2. For example, six peak sun hours means that the energy received during total daylight hours equals the energy that would have been received had the irradiance for 6 hours been 1,000 W/m^2.

Peak Watt:- A unit used to rate the performance of PV cells, modules, or arrays; the maximum nominal output of a photovoltaic device, in watts (W_p) under standardized test conditions, usually 1,000 W/m^2 of sunlight with other conditions, such as temperature specified.

Photon:- A particle of light that acts as an individual unit of energy.

Photovoltaic (PV):- Pertaining to the direct conversion of light into electricity.

Photovoltaic (PV) array:- An interconnected system of PV modules that function as a single electricity-producing unit. The modules are assembled as a discrete structure, with common support or mounting. In smaller systems, an array can consist of a single module.

Photovoltaic (PV) cell:- The smallest semiconductor element within a PV module to perform the immediate conversion of light into electrical energy (dc voltage and current).

Photovoltaic (PV) conversion efficiency:- The ratio of the electric power produced by a photovoltaic device to the power of the sunlight incident on the device.

Photovoltaic (PV) efficiency:- The ratio of electric power produced by a cell at any instant to the power of the sunlight striking the cell. This is typically about 9% to 14% for commercially available cells.

Photovoltaic (PV) generator:- The total of all PV strings of a PV power supply system, which are electrically interconnected.

Photovoltaic (PV) module:- The smallest environmentally protected, essentially planar assembly of PV cells and ancillary parts, such as interconnections, terminals, [and protective devices such as diodes] intended to generate dc power under unconcentrated sunlight. The structural (load carrying) member of a module can either be the top layer (superstrate) or the back layer (substrate). [UL 1703]

Photovoltaic (PV) panel:- often used interchangeably with PV module (especially in one-module systems), but more accurately used to refer to a physically connected collection of modules (i.e., a laminate string of modules used to achieve a required voltage and current).

Photovoltaic (PV) peak watt:- Maximum "rated" output of a cell, module, or system. Typical rating conditions are 1000 watts per square meter of sunlight, 20°C ambient air temperature and 1 m/s wind speed.

Photovoltaic (PV) system:- Thee set of components for converting sunlight into electricity by the photovoltaic process, including the array and balance of system components.

Photovoltaic-thermal (PV/T) system:- A photovoltaic system that, in addition to converting sunlight into electricity, collects the residual heat energy and delivers both heat and electricity in usable form. Also called a total energy system.

Physical Vapour Deposition:- A method of depositing thin semiconductor photovoltaic films. With this method, physical processes, such as thermal evaporation or bombardment of ions, are used to deposit elemental semiconductor material on a substrate.

P-I-N:- A semiconductor photovoltaic (PV) device structure that layers an intrinsic semiconductor between a p-type semiconductor and an n-type semiconductor; this structure is most often used with amorphous silicon PV devices.

Polycrystalline:- See Multicrystalline.

Polycrystalline Silicon:- A material used to make photovoltaic cells, which consist of many crystals unlike single-crystal silicon.

PV:- Abbreviation for photovoltaic(s).

Pyronometer:- An instrument for measuring total hemispherical solar irradiance on a flat surface, or "global" irradiance; thermopile sensors have been generally identified as pyranometers, however, silicon sensors are also referred to as pyranometers.

Pyrheliometer:- An instrument used for measuring direct beam solar irradiance. Uses an aperture of 5.7° to transcribe the solar disc.

Recombination:- The action of a free electron falling back into a hole. Recombination processes are either radiative, where the energy of recombination results in the emission of a photon, or non-radiative, where the energy of recombination is given to a second electron which then relaxes back to its original energy by emitting phonons. Recombination can take place in the bulk of the semiconductor, at the surfaces, in the junction region, at defects, or between interfaces.

Resistive voltage drop:- The voltage developed across a cell by the current flow through the resistance of the cell.

Ribbon (Photovoltaic) Cells:- A type of photovoltaic device made in a continuous process of pulling material from a molten bath of photovoltaic material, such as silicon, to form a thin sheet of material.

Semiconductor:- Any material that has a limited capacity for conducting an electric current. Generally falls between a metal and an insulator in conductivity. Certain semiconductors, including silicon, gallium arsenide, copper indium diselenide, and cadmium telluride, are uniquely suited to the photovoltaic conversion process.

Semicrystalline:- See 'Multicrystalline.'

Series connection:- A way of joining photovoltaic cells or batteries by connecting positive leads to negative leads; such a configuration increases the voltage.

Series resistance:- Parasitic resistance to current flow in a cell due to mechanisms such as resistance from the bulk of the semiconductor material, metallic contacts, and interconnections.

Short-circuit current (I_{sc}):- The current flowing freely from a photovoltaic cell through an external circuit that has no load or resistance; the maximum current possible.

Silicon (Si):- A chemical element, atomic number 14, semi-metallic in nature, dark gray, an excellent semiconductor material. A common constituent of sand and quartz (as the oxide). Crystallizes in face-centred cubic lattice like a diamond. The most common semiconductor material used in making photovoltaic devices.

Single-crystal material:- A material that is composed of a single crystal or a few large crystals.

Solar cell:- See 'Photovoltaic cell.'

Solar constant:- The strength of sunlight; 1353 watts per square meter in space and about 1000 watts per square meter at sea level at the equator at solar noon.

Solar energy:- Electromagnetic energy transmitted from the sun (solar radiation). The amount that reaches the earth is equal to one billionth of total solar energy generated, or the equivalent of about 420 trillion kilowatt-hours.

Solar-grade silicon:- Intermediate-grade silicon used in the manufacture of PV cells. Less expensive than electronic-grade silicon.

Solar noon:- That moment of the day that divides the daylight hours for that day exactly in half. To determine solar noon, calculate the length of the day from the time of sunset and sunrise and divide by two. Solar noon may be quite a bit different from 'clock' noon.

Solar spectrum:- The total distribution of electromagnetic radiation emanating from the sun. The different regions of the solar spectrum are described by their wavelength range. The visible region extends from about 390 to 780 nanometres (a nanometre is one billionth of one meter). About 99 percent of solar radiation is contained in a wavelength region from 300 nm (ultraviolet) to 3,000 nm (near-infrared). The combined radiation in the wavelength region from 280 nm to 4,000 nm is called the broadband, or total, solar radiation.

Solar thermal electric:- Method of producing electricity from solar energy by using focused sunlight to heat a working fluid, which in turn drives a turbogenerator.

Split-spectrum cell:- A compound photovoltaic device in which sunlight is first divided into spectral regions by optical means. Each region is then directed to a different photovoltaic cell optimized for converting that portion of the spectrum into electricity. Such a device achieves significantly greater overall conversion of incident sunlight into electricity. See 'mulitjunction device.'

Sputtering:- A process used to apply photovoltaic semiconductor material to a substrate by a physical vapour deposition process where high-energy ions are used to bombard elemental sources of semiconductor material, which eject vapors of atoms that are then deposited in thin layers on a substrate.

Stand-alone (PV system):- An autonomous or hybrid photovoltaic system not connected to a grid. May or may not have storage, but most stand-alone systems require batteries or some other form of storage.

Stand-off mounting:- Technique for mounting a PV array on a sloped roof, which involves mounting the modules a short distance above the pitched roof and tilting them to the optimum angle.

Standard Test Conditions (STC):- Conditions under which a module is typically tested in a laboratory [IEC 1215]:

i. Irradiance intensity of 1000 W/square meter;
ii. AM1.5 solar reference spectrum; and
iii. A cell (module) temperature of 25 degrees C, plus or minus 2 degrees C.

Substrate:- The physical material upon which a photovoltaic cell is made.

Superstrate:- The covering on the sun side of a PV module, providing protection for the PV materials from impact and environmental degradation while allowing maximum transmission of the appropriate wavelengths of the solar spectrum.

Staebler-Wronski Effect:- The tendency of the sunlight to electricity conversion efficiency of amorphous silicon photovoltaic devices to degrade (drop) upon initial exposure to light.

String:- A number of photovoltaic modules or panels interconnected electrically in series to produce the operating voltage required by the load.

Substrate:- The physical material upon which a photovoltaic cell is applied.

Superstrate:- The covering on the sunny side of a photovoltaic (PV) module, providing protection for the PV materials from impact and environmental degradation while allowing maximum transmission of the appropriate wavelengths of the solar spectrum.

Thermal electric:- Electric energy derived from heat energy, usually by heating a working fluid, which drives a turbogenerator. See 'solar thermal electric.'

Thermophotovoltaic (TPV) device:- A device that converts secondary thermal radiation, re-emitted by an absorber or heat source, into electricity; The device is designed for maximum efficiency at the wavelength of the secondary radiation.

Thick-crystalline materials:- Semiconductor material, typically measuring from 200-400 microns thick, that is cut from ingots or ribbons.

Thin film:- A layer of semiconductor material, such as copper indium diselenide, cadmium telluride, gallium arsenide, or amorphous silicon, a few microns or less in thickness, used to make photovoltaic cells. Commonly called amorphous.

Thin Film Photovoltaic Module:- A photovoltaic module constructed with sequential layers of thin film semiconductor materials. See amorphous silicon.

Tilt Angle:- The angle at which a photovoltaic array is set to face the sun relative to a horizontal position. The tilt angle can be set or adjusted to maximize seasonal or annual energy collection.

Total internal reflection:- The trapping of light by refraction and reflection at critical angles inside a semiconductor device so that it cannot escape the device and must eventually be absorbed by the semiconductor.

Tracking array:- PV array that follows the path of the sun to maximize the solar radiation incident on the PV surface. The two most common orientations are:

i. one axis where the array tracks the sun east to west and
ii. two-axis tracking where the array points directly at the sun at all times.

Tracking arrays use both the direct and diffuse sunlight. Two-axis tracking arrays capture the maximum possible daily energy. Typically, a single axis tracker will give you 15% to 25% more power per day, and dual axis tracking will add about 5% to that. Depends somewhat on latitude and season.

Two-axis tracking:- A system capable of rotating independently about two axes (e.g., vertical and horizontal) and following the sun for maximum efficiency of the solar array.

Tunnelling:- Quantum mechanical concept whereby an electron is found on the opposite side of an insulating barrier without having passed through or around the barrier.

Ultraviolet:- Electromagnetic radiation in the wavelength range of 4 to 400 nanometres.

Vacuum Evaporation:- The deposition of thin films of semiconductor material by the evaporation of elemental sources in a vacuum.

Valence Band:- The highest energy band in a semiconductor that can be filled with electrons.

Valence Level Energy/Valence State:- Energy content of an electron in orbit about an atomic nucleus. Also called bound state.

Vertical Multijunction (VMJ) Cell:- A compound cell made of different semiconductor materials in layers, one above the other. Sunlight entering the top passes through successive cell barriers, each of which converts a separate portion of the spectrum into electricity, thus achieving greater total conversion efficiency of the incident light. Also called a multiple junction cell. See multijunction device and split-spectrum cell.

V_{mp}:- Voltage at maximum power . The voltage at which maximum power is available from a photovoltaic module.

V_{oc}:- Open-circuit voltage

Voltage at maximum power (V_{mp}):- The voltage at which maximum power is available from a module.

Wafer:- A thin sheet of semiconductor material made by mechanically sawing it from a single-crystal or multicrystal ingot or casting.

Watt-hour (Wh):- See 'Kilowatt-hour.'

Window:- A wide band gap material chosen for its transparency to light. Generally used as the top layer of a photovoltaic device, the window allows almost all of the light to reach the semiconductor layers beneath.

Work Function:- The energy difference between the Fermi level and vacuum zero. The minimum amount of energy it takes to remove an electron from a substance into the vacuum.

Zenith Angle:- the angle between the direction of interest (of the sun, for example) and the zenith (directly overhead).

Glossary of Capacitor terminology

AC voltage:- The sum of the dc and peak ac voltage applied to the capacitor should not exceed the rated dc voltage, nor should the rms voltage exceed the Corona Start Voltage.

Aerogel Capacitor:- these capacitors use carbon aerogel to attain immense electrode surface area, can attain huge values, up to thousands of farads. EDLCs can be used as replacements for batteries in applications where a high discharge current is required, e.g. in electric vehicles. They can also be recharged hundreds of thousands of times, unlike conventional batteries which last for only a few hundred or thousand recharge cycles. However, capacitor voltage drops faster than battery voltage during discharge, so a dc to dc converter may be used to maintain voltage and to make more of the energy stored in the capacitor usable.

Aluminium Electrolytic Capacitor:- are compact but 'lossy'. A capacitor made up of two aluminium electrodes separated by paper saturated with an electrolyte. The dielectric is the oxide of the anode. They are available in the range of less than 1µF to 1,000,000µF with working voltages over five hundred volts dc. The dielectric is a thin layer of aluminium oxide. They contain corrosive liquid and can burst if the device is connected backwards. The electrolyte will tend to dry out in the absence of a sufficient rejuvenating voltage, and eventually the capacitor will fail. Bipolar electrolytics contain two capacitors connected in series opposition and are used for coupling ac signals. Poor frequency and temperature characteristics make them unsuited for high-frequency applications.

Capacitance (Capacity):- That property of a system of conductors and dielectrics which permits the storage of electricity when potential difference exists between the conductors. A measure of the energy storage ability of a capacitor, given as C = k A/d, where A is the area of the electrodes, d is their separation, and k is a function of the dielectric between the electrodes. The formula yields a result in farads (F), but a farad is so large that the most commonly used values are expressed in microfarads (µf = 10^{-6}F) or picofarads (pf = 10^{-12}F). Capacitance is always positive.

Capacitive Reactance (X_c):- The opposition to the flow of alternating or pulsating current by a capacitor measured in ohms. The imaginary component of the impedance of a capacitor. The non-heating impedance component of the capacitor when ac flows: $X_c = 1/2\pi fC$.

Capacitor:- An electrical/electronic part that stores electrical charges. In its simplest form it consists of two conducting surfaces separated by a dielectric. A passive circuit element capable of storing electrical energy and releasing it at a predetermined time and at a predetermined rate.

Charge:- The amount of electricity present upon the capacitor's plates. Also, the act of forcing of electrons onto the capacitor's plates. See Coulomb.

Corona:- A luminous discharge due to ionization of the gas surrounding a conductor around which exists a voltage gradient exceeding a certain critical value. Any electrically detectable, field intensified ionization that does not result immediately in complete breakdown of the insulation and electrode system in which it occurs. A type of discharge -sometimes visible- in the dielectric of an insulation system caused by an electric field and characterized by the rapid development of an ionized channel which does not completely bridge the electrode. May be continuous or intermittent. Not a materials property, but related to the system, including electrodes. Its incidence can be reduced or avoided through special designs.

Corona Resistance:- The time that insulation will withstand a specified level field-intensified ionization that does not result in the immediate complete breakdown of the insulation.

Creepage:- Electrical leakage on a solid dielectric surface

Critical Voltage (of gas):- The voltage at which a gas ionizes and corona occurs, preliminary to dielectric breakdown of the gas.

Ceramic Capacitor:-This capacitor is so named because it contains a ceramic dielectric. One type of ceramic capacitor uses a hollow ceramic cylinder as both the form on which to construct the capacitor and as the dielectric material. The plates consist of thin films of metal deposited on the ceramic cylinder. The other type of ceramic capacitor is manufactured in the shape of a disk. After leads are attached to each side of the capacitor, the capacitor is completely covered with an insulating moisture-proof coating. Ceramic capacitors usually range in value from 1pF to 0.1µF and may be used with voltages as high as 30kV.

Coulomb:- A coulomb is the unit of electric charge.1 coulomb is the amount of electric charge transported by a current of one ampere in one second. It can also be defined in terms of capacitance and voltage, where one coulomb is defined as one farad of capacitance times one volt of electric potential difference.

Dielectric:- The insulating material between the plates of the capacitor. The material is chosen for its ability to permit electrostatic attraction and repulsion to take place across it. The material will have the property that energy required to establish an electric field is recoverable in whole or in part, as electric energy. In other words, a good dielectric material is a poor conductor of electricity while being an effective supporter of electrostatic fields.

Dielectric Absorption (DA):-That property of an imperfect dielectric whereby there is an accumulation of electric charges within the body of the material when it is placed in an electric field. An apparent 'recovery voltage' measured after the capacitor is discharged and expressed as a percent of the initial charge voltage. DA is due largely to the dipole moment of the dielectric and to lesser degree the migration of free electrons to the surface of the dielectric. A measure of the reluctance of a capacitor's dielectric to discharge completely - usually measured in percent of original charge.

Dielectric Constant:- That property of a dielectric which determines the electrostatic energy stored per unit volume for unit potential gradient.

Dielectric Loss:- The time rate at which electric energy is transformed into heat in a dielectric when it is subjected to a changing electric field.

Dielectric Loss Angle:- The difference between 90° and the dielectric phase angle.

Dielectric Phase Angle:- The angular difference in phase between the sinusoidal alternating potential difference applied to a dielectric and the component of the resulting alternating current having the same period as the potential difference.

Dielectric Power Factor:- The cosine of the dielectric phase angle (or sine of the dielectric loss angle).

Dielectric Strength:- The voltage which an insulating material can withstand before breakdown (puncture) occurs, usually expressed as a voltage gradient (such as volts per mil). The voltage figure used is the average RMS voltage gradient between two electrodes at the time of failure.

Displacement Current:- A current which exists in addition to ordinary conduction current in AC circuits. It is proportional to the rate of change of the electric field.

Disruptive Discharge:- The sudden and large increase in current through an insulation medium due to the complete failure of the medium under the electrostatic stress.

Dissipation Factor (DF) tanδ:-The tangent of the loss angle of the insulating material. A measure of the deviation from the ideal capacitance value. A measure of the power factor (or losses) of a capacitor, given as tanδ =DF = $2\pi \times fRC \times 100\%$, where R is the equivalent series resistance (ESR) of the capacitor, f is the frequency (Hz.), and C is capacitance (Farads). Dissipation Factor varies with frequency and temperature.

dv/dt:- Change in Voltage divided by Change in Time, usually expressed in Volts per us. Is the maximum allowed change in volts per microsecond at the rated voltage. The maximum voltage rise (or discharge) time a capacitor can withstand being damaged.

EDLC:- Electric Double Layer Capacitor - is a next-generation energy storage device that will be used as an auxiliary power supply and the combined use with photovoltaics equipment and hybrid electric cars. Also known as supercapacitors or ultracapacitors, have very high capacitance values but low voltage ratings. They use a molecule-thin layer of electrolyte, rather than a manufactured sheet of material, as the dielectric. As the energy stored is inversely proportional to the thickness of the dielectric, these capacitors have an extremely high energy density. The electrodes are made of activated carbon, which has a high surface area per unit volume, further increasing the capacitor's energy density.

Electrolytic Capacitor:- is used where a large amount of capacitance is required. As the name implies, an electrolytic capacitor contains an electrolyte. This electrolyte can be in the form of a liquid (wet electrolytic capacitor). The wet electrolytic capacitor is no longer in popular use due to the care needed to prevent spilling of the electrolyte. A dry electrolytic capacitor consists essentially of two metal plates separated by the electrolyte. In most cases the capacitor is housed in a cylindrical aluminium container which acts as the negative terminal of the capacitor. The positive terminal (or terminals if the capacitor is of the multisection type) is a lug (or lugs) on the bottom end of the container. The capacitance value(s) and the voltage rating of the capacitor are generally printed on the side of the aluminium case

A polarized capacitor exhibiting a high capacitance/volume ratio that consists of two electrodes immersed in an electrolyte, with a chemical film that acts as a dielectric on one or both electrodes. Electrolytic capacitors are made by winding either plain or etched foils on which an oxide has been formed on the surface of one (either anode or cathode) film. The etching of the foil increases the surface area and a considerable increase in capacitance is obtained.

Equivalent Series Resistance (ESR):- A resistive series element of the capacitor model found in both the AC and DC domains. Contributing factors: electrodes, leads, dielectric. This value can change with frequency, time, etc. A measure of the total lossiness of a capacitor which includes the leads, electrodes, dielectric losses, leakage (IR) and most important, the end spray connecting the leads to the metallised film. The lower the ESR the higher the current carrying ability the capacitor will have. It is related and dependant on temperature and frequency and generally when either these factors increase, a reduction in ESR results.

The sum of all the internal resistances of a capacitor measured in ohms. Expressed mathematically as ESR = $DF \times X_c$.

Farad:- A farad is defined as the amount of capacitance for which a potential difference of one volt results in a static charge of one coulomb. It has the base SI representation of $s^4 \cdot A^2 \cdot m^{-2} \cdot kg^{-1}$. Since an ampere is the rate of electrical flow (current) of one coulomb per second, an alternate definition is that a farad is the amount of capacitance that requires one second for a one ampere flow of charge to change the voltage by one volt.

The basic unit of a measure of a capacitor. A capacitor charged to 1 volt with a charge of 1 coulomb would have a capacitance of 1 Farad. 1 µF = .000001 Farads.

Film Capacitor:- Made from high quality polymer film (usually polycarbonate, polystyrene, polypropylene, polyester (Mylar), and for high quality capacitors polysulphone), and metal foil or a layer of metal deposited on surface. They have good quality and stability, and are suitable for timer circuits and for high frequencies.

Fixed Capacitor:- is constructed in such manner that it possesses a fixed value of capacitance which cannot be adjusted. A fixed capacitor is classified according to the type of material used as its dielectric, such as paper, oil, mica, or electrolyte.

I_{rms}:- The maximum rms ripple current in amps at a given frequency.

I_{peak}:-The maximum peak current in amps at +25°C for non-repetitive pulses or where the pulse time off is sufficient to allow cooling so overheating will not result.

Inductance ESL:- Some series inductance is present in all capacitor, which dominates impedance at very high frequencies. Most significant in aluminium electrolytic capacitors, with values usually less than a few tens of nH.

Insulation:- Material having a high resistance to the flow of electric current, which prevents leakage of current from a conductor.

Insulation Resistance (IR):- The ratio of the applied Voltage to the total current between two electrodes in contact with a specific insulator. A measure of the resistance to a dc current flow through the capacitor under steady state conditions. Values for film and ceramic capacitors are usually expressed in megohm-microfarads for a given design and dielectric. The actual resistance of the capacitor is obtained by dividing the megohm-microfarads by the capacitance.

A measure of the resistance to a dc current flow through the capacitor under steady state conditions. Values for film and ceramic capacitors are usually expressed in megohm-microfarads for a given design and dielectric. The actual resistance of the capacitor is obtained by dividing the megohm-microfarads by the capacitance.

The ratio of the dc voltage applied to the terminals of a capacitor and the resultant leakage current flowing through the dielectric and over its surface after the initial charging current has ceased expressed in megohms or as time constant megohm × microfarads.

Impedance (Z_c):- The total opposition offered to alternating or pulsating current measured in ohms. Impedance is the vector sum of the resistive and reactive series components of a capacitor expressed mathematically as $Z_c = \left(ESR^2 + \left(X_l - X_c\right)^2\right)^{\frac{1}{2}}$. Impedance is dominated by the capacitive reactance at low frequencies and by the inductive reactance at high frequencies. At the series resonant frequency $Z=ESR$.

Insulator:- A material of such low electrical conductivity that the flow of current through it can usually be neglected.

Ion, Ionization:- An electrified portion of matter of sub-atomic, atomic, or molecular dimensions such as is formed when a molecule of gas loses an electron (when the gas is stressed electrically beyond the critical voltage) or when a neutral atom or group of atoms in a fluid loses or gains one or more electrons. Ionization is the dissociation of an atom or molecule into positive or negative ions or electrons. Restrictively, the state of an insulator whereby it facilitates the passage of current due to the presence of charged particles usually induced artificially.

Joule (watt second):- Joule = ½×Capacitance (Farads)×Voltage^2. The Joule is a measure of the amount of energy delivered by one Watt of power in one second or 1 million watts of power in one microsecond. The Joule rating of a surge protection device is the amount of energy that it can absorb before it becomes damaged.

Leakage Current:- Measure of the stray direct current flowing through capacitor after dc voltage is impressed on it.

After charging a capacitor to a set voltage, initially, a high current flows which decreases rapidly until a constant small value is reached, the final leakage current. The leakage current value increases both with voltage and temperature. In the case of electrolytic capacitors, after a long storage period, the leakage current value can exceed the rated value and leakage measurement is after a short re-anodization period.

Metallised Capacitor:- A capacitor where a thin layer of metal is vacuum-deposited directly onto the dielectric.

Mica Capacitor:- is made of metal foil plates that are separated by sheets of mica (the dielectric). The whole assembly is encased in moulded plastic. Since the capacitor parts are moulded into a plastic case, corrosion and damage to the plates and dielectric are prevented. Also the moulded plastic case makes the capacitor mechanically stronger. Various types of terminals are used on mica capacitors to connect them into circuits. The terminals are also moulded into the plastic case.

Oil Capacitors (Self Healing):- these are often used in high-power electronic equipment. An oil-filled capacitor is nothing more than a paper capacitor that is immersed in oil. Since oil impregnated paper has a high dielectric constant, it can be used in the production of capacitors having a high capacitance value. Many capacitors will use oil with another dielectric material to prevent arcing between the plates. If arcing should occur between the plates of an oil-filled capacitor, the oil will tend to reseal the hole caused by the arcing. These are referred to as a self healing capacitor.

Overvoltage:- A voltage above the normal operating voltage of a device or circuit. In a dielectric withstand test, capacitors are overvoltage-tested (Hi-potted) at 1.5× or 2× its rated voltage to assure quality.

Partial Discharge:- A partial discharge is an electric discharge that only partially bridges the insulation between conductors when the voltage stress exceeds a critical value. Partial discharges may, or may not, occur adjacent to a conductor. Partial discharge is often referred to as corona but the term corona is preferably reserved for localized discharges in cases around a conductor, bare or insulated, remote from any other solid insulation.

Polychlorinated Biphenyls PCB:- Chemical pollutant formerly used in oil-filled capacitors which have been outlawed since the 1970's and are no longer used in the capacitor and transformer industries.

Permittivity:- Preferred term for dielectric constant.

Polycarbonate Resins:- Polymers derived from the direct reaction between aromatic and aliphatic dihydroxy compounds with phosgene or by the ester exchange reaction with appropriate phosgene derived precursors.

Polyester:- A resin formed by the reaction between a dibasic acid and a dihydroxy alcohol.

Polyethylene:- A thermoplastic material composed of polymers of ethylene.

Polymer:- A compound formed by polymerization which results in the chemical union of monomers or the continued reaction between lower molecular weight polymers.

Polymerize:- To unite chemically two or more monomers or polymers of the same kind to form a molecule with higher molecular weight.

Polypropylene:- A plastic made by the polymerization of high-purity propylene gas in the presence of an organometallic catalyst at relative low pressures and temperatures.

Polystyrene:- A thermoplastic produced by the polymerization of styrene (vinyl benzene).

Pulse Operation:- Capacitors subjected to dc pulses or non-sinusoidal voltages with fast rise or drop times (high *dv/dt*) will be exposed to high current. This current must be limited to within the maximum peak current allowed. These peak currents refer to an unlimited number of pulses charging or discharging the capacitors.

Rated Capacitance C_R:- The rated capacitance, defined at specific frequency and temperature, for example, 100 Hz and 20°C, is the capacitance of an equivalent circuit having capacitance and resistance series connected.

Rated Voltage V_R:- The rated voltage is the voltage value that can be applied continuously within the operating temperature range of capacitors. When using a capacitor with AC voltage superimposed on a DC voltage, the peak value of AC voltage plus the DC voltage must not exceed the rated voltage.

Ripple Current:- The total amount of alternating and direct current that can be applied to a capacitor under specific conditions without causing a failure. It depends mostly on the allowable temperature rise due to the ESR I^2R heat production. Since ripple current raises the core temperature, it is important in specifying operational life of the component.

Shelf Life:- The voltage free storage life, most important with electrolytic based capacitors. Period based on specified drift in ESR and impedance. At 20°C, the shelf life of a high voltage (>100V) electrolytic capacitor, is as short as two years.

Sparkover:- A disruptive discharge between electrodes of a measuring gap, such as a sphere gap or oil testing gap.

Surface Leakage:- The passing of current over the boundary surfaces of an insulator as distinguished from passage through its volume.

Surge:- A transient variation in the current and/or potential at a point in the circuit.

Surge Voltage (SV) V_p:- The maximum dc voltage a capacitor can tolerate under any circumstances for a short period of time without suffering any damage. The surge voltage is the maximum overvoltage including DC, peak AC and transients to which the capacitor can be subjected for short periods of time. Typically, not more then 30 seconds in any 5 minute period, at maximum operational temperature, where the charge is held for 30 seconds for 1000 cycles, then the capacitor is allowed to discharge without load for 5 minutes.

SuperCapacitors:- another word for Ultracapacitors - Made from carbon aerogel, carbon nanotubes, or highly porous electrode materials. Extremely high capacity. Can be used in some applications instead of rechargeable batteries.

Tantalum Capacitor:- compact, low-voltage devices up to several hundred µF, these have a lower energy density and are more accurate than aluminium electrolytics. These capacitors are comprised of a permeable tantalum centre section surrounded by tantalum pentoxide. A tantalum wire is inserted into the centre section and then extends axially from the component. There are many advantages of using tantalum capacitors over other types: They have higher volumetric efficiency (CV/cc); They have superior frequency characteristics; They are highly reliable and do not degrade over time. Tantalum capacitors do not lose capacitance like electrolytic capacitors. Unlimited shelf life.

Temperature Coefficient (TC):-The change in capacitance with temperature expressed linearly as parts per million per degree centigrade (ppm/°C), or as a percent change over a specified temperature range. Most film capacitors are not linear and TC is expressed in percent. The change in capacitance with temperature expressed linearly as parts per million per degree centigrade (ppm/°C), or as a percent change over a specified temperature range. Most film capacitors are not linear and TC is expressed in percent.

Thermal Conductivity:- Ability of a material to conduct heat.

Transients:- High voltage surges through an electrical system caused by lightning strikes to nearby transformers, overhead lines, or the ground. May also be caused by switching of motors and compressors, as well as by short circuits or utility system switching.

Voltage Sag:- Drop in voltage levels of electrical distribution system which interferes with the operation of electrical and electronic equipment. Commonly called brownout. Results when demand for electricity exceeds capacity of the distribution system.

Volumetric efficiency:- Energy density in µf-volts per cubic centimetre, from: (capacitance) X (working voltage) ÷ (volume). Longer capacitors are more efficient than shorter units, because of volume used by encapsulation and unused dielectric at the capacitor ends (the margins). Cylindrical units have a smaller volume than rectangular units, although rectangular units can be stacked more compactly.

Working voltage (WV_{dc}, WV_{ac}):- The maximum continuous voltage that should be applied to a capacitor. Rated voltages for dc and ac operation are usually not the same. The maximum dc voltage applied to a capacitor for continuous operation at maximum rated temperature.

X Capacitor:- RFI Capacitor used in positions where if failed would not be hazardous to anyone who touches the case of the equipment. The X capacitors are connected across the line conductors. There are three sub-classes of X capacitors: X1, X2 and X3. The most common is X2 sub-class, used for IEC-664 Installation Category II. The X2 capacitors are rated for peak pulse voltage in service of less or equal to 2.5KV.

Y Capacitor:- RFI Capacitor used in positions where if failed could be hazardous to somebody who touches the case of the equipment. The Y capacitors are connected between power lines and chassis/earth. There are four sub-classes of Y capacitors: Y1, Y2, Y3 and Y4. The most common is Y2 sub-class, used across a Basic or Supplementary insulation. The Y2 capacitors are rated for nominal working voltages less or equal to 250Vac and for peak impulse voltage before endurance test of less or equal to 5KV. Because Safety Standards stipulate maximum current towards earth for different applications, the capacitance of Y capacitors must be limited to a certain value depending on the type of equipment in which the capacitor is used.

Glossary of Thermoelectric terminology

Active heat Load:-The amount of heat (in Watts) being generated by the device that is on top of the TE Cooler. Typically, this is the input power of this device, voltage x current.

Alumina:- Ceramics made of aluminium oxide (Al_2O_3). These ceramics are used on most of our standard TECs. A positive of Al_2O_3 is that it is inexpensive and can be designed for snap states instead of dice, which considerably reduces production costs. Negative aspects of this material are its lower thermal conductivity and it is difficult to use in 3 to 6 stage coolers.

Ambient temperature:- Temperature of the air or environment surrounding a thermoelectric cooling system; sometimes called room temperature.

Aspect ratio:- The numerical ratio of the length (height) to cross-sectional area of a thermoelectric element. An element's L/A aspect ratio is inversely proportional to its optimum current.

BeO:- Ceramics made of beryllium oxide. Typically used in multi-stage coolers due to its higher thermal conductivity. The advantages to this material are that it enhances the thermal performance of the TE Cooler as well as makes it easier to assemble because of the high heat conductance. Disadvantages are that it is expensive and can be toxic when its dust is inhaled. The dust comes from dicing and sanding of the material, both of which are performed on a TE Cooler in its final condition. The risks of BeO sometimes prohibit it as an option.

Bismuth Telluride:- A thermoelectric semiconductor material that exhibits optimum performance in a 'room temperature' range. An alloy of bismuth telluride most often is used for thermoelectric cooling applications.

Bismuth Antimony:- A thermoelectric semiconductor material that exhibits optimum performance characteristics at relatively low temperatures.

Burn-in test:- A power cycling test performed by repeatedly powering on and off the TE Cooler for short intervals of time. The test is designed to detect latent manufacturing or material defects that would cause premature failure of the TE Cooler.

Cascaded module (multi-stage module):- A thermoelectric cooler configuration whereby one cooler is stacked on top of another so as to be thermally in series. This arrangement makes it possible to reach lower temperatures than can be achieved with a single-stage cooler.

Ceramic:- A patterned substrate (at least one side) for a TE Cooler. This material conducts heat and insulates electric current. Typically comprised of Al_2O_3 Al_2N_5 BeO
Thermal Conductivity (W/in °C) .051, 4.0, 6.5, CTE (10^{-6}/°C) 7.0 ,4.0, 9.0,

Coefficient of performance (COP):- A measures of the efficiency of a thermoelectric cooler, device or system. Mathematically, COP is the total heat transferred through the thermoelectric device divided by the electric input power (COP = P_c/W). COP ca be stated as COPR (Coefficient of Performance as a Refrigerator) or as COPH (Coefficient of Performance as a Heater).

Cold side of a thermoelectric module:- The side of a cooler that normally is placed in contact with the object being cooled. When the positive and negative cooler leads are connected to the respective positive and negative terminals of a dc power source, the cooler's cold side will absorb heat. Typically, the leads of a TE cooler are attached to the hot side.

Conduction (thermal):- The transfer of heat within a material caused by a temperature difference through the material. The actual material may be a solid, liquid or gas (or a combination) where heat will flow by means of direct contact from a high temperature region to a lower temperature region.

Convection (thermal):- The transfer of heat by air (gas) movement over a surface. Convection is a combined heat transfer process involving conduction, mixing action, and energy storage.

Couple:- A pair of thermoelectric elements consisting of one N-type and one P-type connected electrically in series and thermally in parallel. Because the input voltage to a single couple is quite low, a number of couples normally are joined together to form a "cooler."

Delta-T, ΔT:- The temperature difference between the cold and hot sides of a thermoelectric cooler.

ΔT Test:- Test performed in which the TE Cooler is placed on a temperature controlled base plate (typically 27°C) and powered at Imax. A thermocouple is pressed onto the top ceramic using a spring plunger and the cold side temperature as well as voltage is measured.

Efficiency:- For thermoelectric coolers, mathematical efficiency is the heat pumped by a cooler divided by the electrical input power; for thermoelectric generators, efficiency is the electrical output power from the cooler divided by the heat input (Q_c/Q_h). To convert to percent, multiply by 100. See definition of Coefficient of Performance.

Figure-of-merit, (Z factor):- A measure of the overall performance of a thermoelectric device or material. Material having the highest figure-of-merit also has the highest thermoelectric performance. A good thermoelectric material will have a high Z, high Seebeck coefficient and low thermal conductivity and resistively.
The Z is a direct measure of the cooling performance of a thermoelectric module. $Z = S^2/\rho\times\kappa$ where S is the Seebeck Coefficient, ρ is electrical resistivity and κ is the thermal conductivity of the thermoelectric material. Z is temperature dependent though, so, when comparing one module to another, they must be based on the same hot-side temperatures.

Heat load:- The quantity of heat presented to a thermoelectric device that must be absorbed by the device's cold side. The term heat load, when used by itself, tends to be somewhat ambiguous and it is preferable to be more specific. Terms such as active heat load, passive heat load or total heat load are more descriptive and less uncertain as to meaning.

Heat pump:- A general term describing a thermoelectric cooling device, often being used as a synonym for a thermoelectric cooler. In somewhat less common usage, the term heat pump has been applied to a thermoelectric device operating in the heating mode.

Heat pump capacity:- The amount of heat that a thermoelectric device is capable of pumping at a given set of operating parameters. Frequently, this term will be used interchangeably with the expression maximum heat pumping capacity. The two terms are not strictly synonymous, however, because maximum heat pumping capacity specifically defines the maximum amount of heat that a cooler will pump at the maximum rated input current and at a zero temperature differential.

Heat transfer coefficient:- A numerical value that describes the degree of coupling that exists between an object and a cooling or heating fluid. The heat transfer coefficient actually is an extremely complex value that encompasses many physical factors.

Hot side of a thermoelectric module:- The face of a thermoelectric cooler that usually is placed in contact with the heat sink. When the positive and negative cooler leads are connected to the respective positive and negative terminals of a DC power source, the cooler's hot side will reject heat. Normally, the wire leads are attached to the hot side ceramic substrate.

I_{max}:- Current which, the maximum ΔT is produced. Generally, it is not a good to operate a TE cooler at I_{max} because the amount of input power increases significantly without a significant change in ΔT. 70 - 80 % of I_{max} is usually an optimal operating condition.

Interstage temperature:- The temperature between specific stages or levels of a multi-stage or cascade cooler.

Joule heating:- The passage of an electrical current through a conductor or material due to the internal resistance, resulting in Heat

Kinetic viscosity:- The ratio of a fluid's viscosity to its density; typically units are centimetres squared per second and feet squared per second.

Latent heat:- Thermal energy required to cause a change of state of a substance such as changing water into ice or water into steam.

Lead telluride:- A thermoelectric semiconductor that exhibits its optimum performance within a temperature range of 250-450°C. Lead telluride is used most often for thermoelectric power generation applications.

Maximum heat pump capacity (maximum P_c):- The maximum quantity of heat that can be absorbed at the cold face of a thermoelectric cooler when the temperature differential between the cold and hot cooler faces is zero and when the cooler is being operated at its rated optimum current. P_{max} is a significant thermoelectric cooler/device specifications.

Maximum temperature differential (maximum ΔT):- The largest difference that can be obtained between the hot and cold faces of a thermoelectric cooler when heat applied to the cold face is effectively zero. ΔT_{max} or D_{max} is one of the significant thermoelectric cooler/device specifications.

Metallisation:- The conductive copper pattern printed on the ceramics.

Module:- A thermoelectric cooling component or device fabricated with multiple thermoelectric couples that are connected thermally in parallel and electrically in series.

Multi-stage module (cascade module):- A thermoelectric configuration whereby one TEC is mechanically stacked on top of another in series. This arrangement makes it possible to reach lower temperatures than can be achieved with a single-stage cooler.

Optimum current:- The specific level of electrical current that will produce the greatest heat absorption by the cold side of a thermoelectric cooler. At the optimum current, a thermoelectric cooler will be capable of pumping the maximum quantity of heat; maximum temperature differential (ΔT_{max}) typically occurs at a somewhat lower current level.

Passive heat loads:- The amount of non-active heat (in Watts) being applied on the TE cooler. This includes conductance through wires that extend from the cold side of the TE cooler to the ambient, the convective loads from the surrounding atmosphere (note: convective loads are present in Nitrogen, Argon, and Xenon, but are not present in a vacuum).

Peltier effect:- The phenomenon whereby the passage of an electrical current through a junction consisting of two dissimilar metals results in a cooling effect; when the direction of current flow is reversed heating will occur.

Q_{max}:- The maximum amount of heat (in Watts) that a TE cooler can pump. This occurs when the ΔT is zero. Only for multi-stage coolers operating near a ΔT_{max} condition.

Seebeck Coefficient:- The Seebeck Coefficient is a measure of the electrical voltage potential that exists in an electrical conductor whose ends are maintained at two different temperatures and current is not flowing. It is an intrinsic property and has units of V/K. Thermocouples used for temperature measurement utilize this principle.

Seebeck effect:- The phenomenon whereby an electrical current will flow in a closed circuit made up of two dissimilar metals when the junctions of the metals are maintained at two different temperatures. A common thermocouple used for temperature measurement utilizes this principle.

Silicon-germanium:- A high temperature thermoelectric semi-conductor material that exhibits its optimum performance within a temperature range of 500-1000°C. Silicon-Germanium material most often is used for special thermoelectric power generation applications that utilize a radioisotope/nuclear heat source.

Single-stage module:- The most common type of thermoelectric cooling module using a single layer of thermoelectric couples connected electrically in series and thermally in parallel. Single-stage coolers will produce a maximum temperature differential of approximately 70°C under a no-load condition.

Thermal coefficient of expansion:- A measure of the dimensional change of a material due to a temperature change. Common measurement units include centimetre per centimetre per °C.

Thermal conductance:- The amount of heat a given object will transmit per unit of temperature. Thermal conductance is independent of the physical dimensions, i.e., cross-sectional area and length of the object. Typical units include watts per degree Celsius.
Thermal conductivity:- The amount of heat a material will transmit per unit of temperature based on the material's cross-sectional area and thickness.
Thermal grease:- A grease-like material used to enhance heat transfer between two surfaces by filling in the microscopic voids caused by surface roughness. Most thermal greases, also known as Transistor Heat Sink Compound or Thermal Joint Compound, are made from silicone grease loaded with zinc oxide. Non-silicone based compounds are also available which in most cases are superior but more expensive than silicone-based alternatives.
Thermal shock:- Thermal Shock also is referred to as temperature cycling in some MIL specs. In a thermal shock test, the TE cooler (not powered throughout test) is placed in a hot chamber (for example, 85°C) for a set time (for example, 30 minutes). The part is then transferred to the cold chamber (for example, -40°C) for the same time. This cycle is repeated several times depending on the requirement.
Thermoelectric:- A term used to denote not only the products produced but also the basic scientific principle upon which products are designed.
Thermoelectric generator:- A device that directly converts energy into electrical energy based on the Seebeck Effect. Bismuth telluride-based thermoelectric generators have very low efficiencies (generally not exceeding two or three percent) but may provide useful electrical power in certain applications.
Thermoelectric heat pump:- Another name for a thermoelectric module or thermoelectric cooler. The term Heat Pump has been used by some specifically to denote the use of a thermoelectric module in the heating mode, but this usage is uncommon.
Thermoelectric material:- An alloy of materials that produce thermoelectric properties.
Thomson Coefficient:- If the ends of an electrical conductor are held at two different temperatures, a voltage potential is created because there will be a tendency for electrons at the hot end of the conductor to drift towards the cold end of the conductor. When an external current is applied, so that electrical carriers flow from cold end to the hot end, the electrical carriers must absorb heat to maintain equilibrium with the temperature. If the external current was applied from hot to cold, the carriers would release heat to maintain temperature equilibrium. The Thomson Coefficient is a measure of the voltage per difference in temperature, and with the application of an external current is a measure of the heat generated or absorbed per unit temperature difference per unit current.
Usually, the Thomson effect is intrinsic to the material. However, the Thomson effect can also be extrinsically applied to a conductor by varying the material properties along the length of the conductor. This can actually improve the cooling performance as compared to the usual isotropic material. The Thomson effect is really more complex than that described above.
Thomson effect:- The phenomena whereby a reversible evolution or absorption of heat occurs at opposite ends of a conductor having a thermal gradient when an electrical current passes through the conductor.
V_{max}:- The optimum voltage the maximum ΔT is produced.

Glossary of Fan Cooling and other Heating and Cooling terminology

A-Coil:- A heat exchanger consisting of two diagonal coils that are joined together in a manner that looks like the letter 'A'.
Absorber:- The blackened surface in a collector that absorbs the solar radiation and converts it to heat energy.
Absorptance:- The ratio of solar energy absorbed by a surface to the solar energy striking it.
Active System:- A solar heating or cooling system that requires external mechanical power to move the collected heat.
Air flow volume:- The amount of air the system circulates, expressed in cubic feet per minute, cfm.
Air Handler/Coil Blower:- The part of an air conditioner or heat pump that moves cooled or heated air throughout the ductwork. An air handler is usually a furnace or a blower coil.
Air System:- Solar domestic hot water systems employing air-type collectors. Hot air generated by these collectors is fan forced through an air-to-liquid heat exchanger with the potable water being pumped through the liquid section of the exchanger. The heated water is then circulated through the storage tank in a similar fashion to the liquid collector system. Air does not need to be protected from freezing or boiling, is non-corrosive, and is free. However, air ducts and air handling units require greater space than piping, and air leaks are difficult to detect.
Air-Type Collector:- A collector that uses air as the heat transfer fluid.
Altitude:- The angular distance from the horizon to the sun.
Ambient Temperature:- The temperature of the surrounding air.

Auxiliary Heat:- The extra heat provided by a conventional heating system for periods of cloudiness or intense cold when a solar heating system cannot provide enough.
Azimuth:- The angular distance between true south and the point on the horizon directly below the sun.
Ball bearing:- The most reliable bearing system in fans. Extremely high temperature load, extremely low starting torque at low temperatures, no loss of lubricant.
Blower:- This frequently refers to large radial and axial fans with dimensions that are usually larger than 120x120mm.
Burn-in:- The running in of fans reduces the otherwise unavoidable early failures and thus increases the reliability of a fan. The burn-in process of fans is integrated in the production process so that errors that may occur, immediately result in "Corrective Action" and do not reach the customer.
BTU:- A British thermal unit is a unit of heat energy. British thermal unit; the amount of heat required to raise or lower the temperature of one pound of water one degree Fahrenheit. The higher the Btu rating, the greater the heating capacity of the system.
BTUh:- British thermal units per hour. 12,000 BTUh equals one ton of cooling.
Calorie:- The quantity of heat needed to raise the temperature of one gram of water one degree Celsius.
Capacity:- The output or producing ability of cooling or heating systems. The ability of a heating or cooling system to heat or cool a given amount of space. For heating, this capacity is usually expressed in British thermal units BTUs. For cooling it is usually given in tons.
CFM:- Abbreviation for cubic feet per minute of air flow, a standard measurement of airflow. This measurement indicates how many cubic feet of air pass by a stationary point in one minute. The higher the number, the more air is being forced through the ductwork by the system.
Check Valve:- A check valve is a mechanical device normally applied to a piping system which allows fluid to flow in only one direction.
Closed Loop:- An underground heat exchanger piping system usually of polyethylene or polybutylene designed to allow the extraction or rejection of heat to the earth by the circulation of fluid within the tubing.
Coefficient of Heat Transmission:- The rate of heat loss in BTU per hour through a square foot interface or other surface when the difference between inner and outer air temperatures is one degree Fahrenheit.
Coefficient of Performance (COP):- Heating capacity divided by electrical energy consumed. (for example,15 kW output / 4.5 kW input = COP of 3.3) The coefficient of performance of a heating system is the electrical ratio of the heat got out divided by the heat put in.
Collector:- A device that collects solar radiation and converts it to heat.
Collector Efficiency:- The ratio of usable heat energy extracted from a collector to the solar energy striking the cover.
Compressor-Watts:- Compressor electricity consumption.
Compressor:- The heart of an air conditioning or heat pump system. It is the part of the unit that pumps refrigerant in order to meet the cooling requirements of the system. It is the refrigeration component which increases the density, temperature and pressure of entering refrigerant through compression and discharges a hot dense gas.
Condensate:- Vapour that liquefies due to the lowering of its temperature to the saturation point.
Condenser:- The heat rejecting mechanism in a heat pump usually in the form of a refrigerant-to-air coil or a refrigerant-to-water coil. Refrigeration heat exchanger where the refrigerant gives up its heat during condensation from a vapour to a liquid.
Condenser coil:- In an air conditioner, the coil dissipates heat from the refrigerant, changing the refrigerant from vapour to liquid. In a heat pump system, the coil absorbs heat from the outdoors.
Condenser fan:- The fan that circulates air over the air-cooled condenser.
Concentrating Collector:- A device which concentrates the sun's rays on an absorber surface which is significantly smaller than the overall collector area
Conductance:- The rate of heat flow (in BTUs per hour) through an object when a 1° F.
Conduction:- The flow of heat due to temperature variations within a material.
Conductivity:- A measure of the ability of a material to permit conduction of heat flow through it.
Convection:- The motion of fluid such as gas or liquid by which heat may be transported.
Cover Plate:- A sheet of glass or transparent plastic placed above the absorber in a flat plate collector.
Cupro-nickel:- 90% copper / 10% nickel alloy which has high corrosion resistance to water containing salt, sulphur, chlorides and other dissolved minerals.
Damper:- Located in ductwork, this movable plate opens and closes to control and regulate airflow. Dampers can be used to balance airflow in a duct system. They are also used in zoning to regulate airflow to certain regions. Dampers are used to direct air to the areas that need it most.
dB:- A decibel is a unit used to measure the relative intensity of sound.
Degree Day:- The number of degrees that the mean temperature for that day is below 65° F. (for example, mean temperature of 40°F for the day 65-40=25 degree days). A unit that represents a 1 °F deviation from some fixed reference point (usually 65°F.) in the mean daily outdoor temperature.

Dehumidifier:- An air cooler that removes moisture from the air.
Delta T:- Difference between LWT and EWT
Desuperheater:- A heat exchanger and pump system which removes a small portion of heat from the compressor discharge gas and typically transfers it to a hot water tank.
Design Heat Load:- The total heat loss from a system under the most severe cold conditions likely to occur.
Design Temperature:- The temperature close to the lowest expected for a location, used to determine the design heat load.
Diffuser:- A grille over an air supply duct having vanes to distribute the discharging air in a specific pattern or direction.
Diffuse Radiation:- Indirect sunlight that is scattered from air molecules, dust and water vapour.
Direct Radiation:- Solar radiation that comes straight from the sun, casting shadows on a clear day.
Downflow:- Air enters at the top or bottom of the unit and is discharged vertically out the bottom.
Downflow furnace:- A furnace that intakes air at its top and discharges air at its bottom.
Drain back system:- The solar heat transfer fluid automatically drains into a tank by gravity. Drain back systems are available in one or two tank configurations. A heat exchanger is necessary, because the tap inlet pressure would prevent draining. The heat transfer fluid in the collector loop may be distilled or tap water if the loop plumbing is copper. If the plumbing is threaded galvanized pipe, inhibitors may be added to prevent corrosion. Most inhibitors are non-potable and require a double wall heat exchanger. The pump used must be sized to overcome a static head.
Drain pan:- This also referred to as a condensate pan. This is a pan used to catch and collect condensate (in residential systems vapour is liquefied on the indoor coil, collected in the drain pan and removed through a drain line).
Dry bulb temperature:- Heat intensity, measured by a dry bulb thermometer.
Dry bulb thermometer:- An instrument that measures air temperature independently of humidity.
DX:- Direct expansion; a system in which heat is transferred by the direct expansion of refrigerant.
Drain down System:- Potable water is circulated from the storage tank through the collector loop. Freeze protection is provided by solenoid valves opening and dumping the water at a preset low temperature.
Collectors and piping are pitched so that the system can drain down, and are assembled to withstand 100 psi tap water line pressures. Pressure reducing valves are recommended when tap water pressure is greater than the working pressure of the system.
Dual Condenser: - A heat pump system which has the capability to switch, usually automatically, between an air and a water heat exchanger. Full capacity hot air or hot water output is available.
Electronic Air Cleaner:- An electronic device that filters out large particles and bioaerosols in the air.
EMC filter grid:- Protective guard with shield, to highly reduce unwanted emission (or influence) of high frequency radiation. The airflow performance is reduced by approximately 10 to 15%. EMC filter and dust protection filters can be combination parts.
Emittance:- A measure of the propensity of a material to emit thermal radiation.
Energy Efficiency Ratio (EER):- Cooling capacity in BTU/hr divided by electrical energy consumed in watts, in steady state.
Eutectic Salts:- A group of materials that melt at low temperatures, absorbing large quantities of heat.
EAT:- Entering air temperature.
EER:- Energy efficiency ratio.
ELT:- Entering liquid temperature.
EWT:- Entering water or fluid temperature.
Evaporator:- The heat absorbing mechanism or heat exchanger in a heat pump. Refrigerant changes phase from a liquid to a gas in this exchanger, absorbing heat energy from the surrounding media in the process.
Evaporator coil:- The half of an air conditioning system located inside. This is a tubing coil in which a volatile liquid evaporates and absorbs heat. This is where the refrigerant evaporates as it absorbs heat from the internal environment air that passes over the coil.
Evaporator Temperature:- The temperature on evaporator side when Freon is converted from a liquid to a vapour (gas).
Fan:- Any device that creates air currents. Electromechanical component for creating airflow that dissipates air heated by thermal loss in a device. In comparison to convection, the heat output is improved by factor 3. Depending on the design, the airflow can be discharged axially (straight through the fan) or radially (discharged at the side).
Fan accessories:- A term for components that are additionally required for a fan: finger guard, filter grid, EMC guard, connecting cable, etc.
Fan Coil:- A unit that includes a cooling and/or heating coil and a fan to move air through the ductwork. Filters for the circulation air and accessories to introduce outside ventilation air may also be included.
Fan-Watts:- Blower motor electricity consumption.

Filter:- Any device that removes impurities through a straining process.
Filter grid:- Protective grid with a replaceable dust filter that protects the fan and the device against dust collecting quickly in an environment that is subject to a high accumulation of dust. The airflow performance is reduced by approx. 25% by a filter guard.
Finger guard:- Safety device of wire or plastic for protecting against injuries to fingers in large fans or for protecting against damage in small fans. The airflow performance is reduced.
Flue:- Any vent or passageway that carries the products of combustion from a furnace.
Flat Plate Collector:- A solar collection device in which sunlight is converted into heat on a plane surface without the aid of reflecting surfaces to concentrate the rays.
Flow IGPM:- Liquid flow.
Forced Convection:- The transfer of heat by the flow of fluids (such as air or water) driven by fans, blowers or pumps.
Freon:- Trade name for a series of synthetic chemicals or refrigerants used in refrigeration systems. Each refrigerant is designed to change phase at specific temperatures and pressures which will produce the desired cooling effect required for a specific application. The refrigerant absorbs energy as it evaporates and releases energy during condensation.
Full Package:- Self contained heat pump which has an integrated blower and compressor.
Full-Condensing Heat Exchanger:- A heat exchanger with enough surface area to condense all the hot refrigerant gas produced by a heat pump to its liquid state thereby transferring all the heat produced by the unit.
Galvanic Corrosion:- A condition caused as a result of a conducting liquid making contact with two different metals which are not properly isolated physically and/or electrically.
Geothermal Energy:- Heat energy stored in the earth's crust by the absorption of solar energy and by conduction with the earth's hot interior.
Getters:- A column or cartridge containing an active metal which will be sacrificed to protect some other metal in the system against galvanic corrosion.
Glaubers Salt:- Sodium sulphate a eutectic salt that melts at 90°F and absorbs about 104 Btu per pound as it does so.
Gravity Convection:- The natural movement of heat that occurs when a warm fluid rises and a cool fluid sinks under the influence of gravity.
Ground Loop:- A series of heat exchange pipes containing an antifreeze solution which are buried either vertically or horizontally in the earth.
Ground Source:- A heat pump which utilizes the earth as its source of energy.
HAB:- Heating mode: heat absorption capacity from the ground or water
Cooling mode: heat absorption capacity from the inside air (total cooling load)
Headers:- The pipe that runs across the edge of an array of solar collectors, gathering or distributing the heat transfer fluid from, or to the risers in the individual collectors. This insures that equal flow rates and pressure are maintained.
Heat Capacity:- A property of a material denoting its ability to absorb heat.
Heat Exchanger:- A component which transfers heat energy from one medium to another. For example, heat could be transferred, in a geothermal heat pump system, from water-to-air or from water-to-water etc. and vice versa. An area, box or coil where heat flows from the warmer to the colder fluid or surface. The transfer heat from one fluid to another without the fluids coming into direct contact with each other. A device, such as a coiled copper tube immersed in a tank of water, that is used to transfer heat from one fluid to another through a separating wall.
Heat Gain:- Heat added to the conditioned space by infiltration, solar radiation, occupant respiration, lighting, and operating equipment.
Heat Loss:- The rate of heat transfer from a heated space to the external environment.
Heat Pump:- A mechanical-compression cycle refrigeration system that can be reversed to either heat or cool the controlled space. A heat pump is an HVAC unit that heats or cools by moving heat. During the winter, a heat pump draws heat from outdoor air and circulates it through the air ducts. In the summer, it reverses the process and removes heat from the space and releases it outdoors. A mechanical device that transfers heat from one medium to another, thereby cooling the first and warming the second.
Heat sink:- A medium or container to which heat flows. The area or media where heat is deposited.
Heat Source:- A medium or container from which heat flows. The area or media from which heat is removed, for example water, air, etc.
Heat Storage:- A device or medium that absorbs collected solar heat and stores it for use during periods of inclement or cold weather.
Heat Storage Capacity:- The amount of heat which can be stored by a material.
Heating Season:- The period from early fall to late spring (in the northern hemisphere) during which additional heat is needed to maintain an environment.
Hybrid Solar Energy System:- A system that uses both active and passive methods in its operation.
Heat Transfer:- The movement of heat energy from one point to another. The means for such movement are conduction, convection, and radiation.

Humidifier:- A machine that adds water vapour to the air to increase humidity.
Humidity:- The presence of water vapour in the air.
Humidity, absolute:- Weight of water vapour per cubic foot of dry air, expressed as grains of moisture per cubic foot.
Humidity, relative:- The amount of moisture in the air expressed as a percentage of the maximum amount that the air is capable of holding at a specific temperature.
Horizontal Flow:- Air enters at the end or any side of the unit and is discharged horizontally out the other end or any side of the unit.
Humidistat:- An automatic device used to maintain humidity at a fixed or adjustable set point.
Hydrodynamic bearing:- Sintered sleeve bearing systems. The lubrication effect is similar to the typical sintered bearing. Especially treating the shaft and bearing tube can achieve extremely stable lubrication with a lower influence of temperature and wear.
Indirect System:- A solar heating or cooling system in which the solar heat is collected externally and transferred internally using ducts or piping and, usually fans or ducts.
Infrared Radiation:- Electromagnetic radiation from the sun that has wavelengths slightly longer than visible light.
Insolation:- The total amount of solar radiation direct, diffused and reflected-striking a surface exposed to the sky.
Insulation:- A material with high resistance (R-value) to heat flow.
IGPM:- Water flow in Imperial Gallons
kWh - kilowatt hours:- Electrical term - 1 kWh equals the use of 1000 watts for one hour.
kW out:- Heat pump capacity in kW's
Langley:- A measure of solar radiation; equal to one calorie per square centimetre.
Latent:- The load created by moisture in the air, including from outside air infiltration and that from internal sources.
Latent Heat:- A type of heat, which when added to or taken from a substance, does not change the temperature of the substance. Instead, the heat energy enables the substance to change its state.
Liquid Type Collector:- A collector using a liquid as the heat transfer fluid.
Liquid-to-Air Heat Pump:- A heat pump which absorbs heat from a liquid and distributes the energy in the form of hot forced air.
Liquid-to-Liquid heat pump:- A heat pump which absorbs heat from a liquid and distributes the energy in the form of hot water.
LWT:- Leaving water temperature.
LAT:- Leaving air temperature.
LLT:- Leaving liquid temperature.
Mechanical Cooling:- Conventional cooling provided by a compressor operated refrigeration device. Term can be interchanged with 'active cooling'.
Natural Convection:- See Gravity Convection.
Nocturnal Cooling:- The cooling of a building or heat storage device by the radiation of excess heat into the night sky.
One-Tank Closed-Loop System:- A conventional DHW tank, usually electrically heated, is converted to a solar DHW storage tank by installing an external heat exchanger coil. The lower electrical element is removed, leaving the uppermost of the usual two elements to provide auxiliary water heating and to achieve good stratification (layering of hotter water over progressively colder water).
Open System:- Some part of the System is open to the atmosphere, or system contains fresh or changeable water or air.
Open Loop:- A system where water is pumped from a water source for use in a heat pump.
Output:- Heat pump capacity in Btu/Ton.
Oversized Evaporator:- A technique of employing a larger than normal evaporator (heat absorption device) in a geothermal heat pump in order to obtain greater heat exchange and thus better performance from the unit.
PSI:- Pounds per square inch.
PSIA:- Pounds per square inch, absolute.
PSIG:- Pounds per square inch gauge.
Package Heat Pump:- A heat pump which has all components (compressor, blower and heat exchangers etc.) in one cabinet.
Passive System:- A solar heating or cooling system that uses no external mechanical power to move the collected solar heat.
Percentage of Possible Sunshine:- The percentage of daytime hours during which there is enough direct solar radiation to cast a shadow.
Photosynthesis:- The conversion of solar energy to chemical energy, by the action of chlorophyll in plants and algae.
Photovoltaic Cells:- Semiconductor devices that convert solar energy into electricity.

Pyranometer:- An instrument for measuring solar radiation.
Radial Fan:- Special shape of a fan where the air is suctioned in axially, turned 90°, radially, and thus discharged vertical to the axis. Also known as a Turbofan. The distinguishing features of radial fans are their extremely high compression rigidity, which renders them especially suitable for ventilation of heat sinks and general cooling applications where space is limited.
Radiation:- The flow of energy through open space via electromagnetic waves, such as visible light.
Radiant Panels:- Panels with integral passages for the flow of warm fluids, either air or liquids. Heat from the fluid is conducted through the metal and transferred to the environment by thermal radiation.
Passive Cooling:- A process whereby cold water (less than 50ºF) is pumped directly to a finned air coil (much like the radiator of a vehicle) so that when the heat pump fan is operated, cooling and dehumidification are provided without the operation of a compressor driven refrigeration system.
Radiant Floor Heating:- Process of embedding tubing (cross-linked polyethylene, polybutylene, etc.) directly in a concrete floor so that hot water can be pumped through the tubing for the purpose of heating the building via the flooring.
Reflected Radiation:- Sunlight that is reflected from the surrounding environment onto a surface exposed to the sky.
Refrigerant:- A chemical that produces a cooling effect while expanding or vaporizing. Most air conditioning systems contain R-22 refrigerant, which is scheduled to be in production until the year 2020. Its used in approximately 95 percent of air conditioning equipment. A liquid such as Freon is use in cooling devices to absorb heat from surrounding air or liquids as it evaporates. A naturally occurring or man made liquid which absorbs and releases heat energy in a refrigeration device by changing phase from a liquid to a gas and vice versa in response to the influence of a refrigeration compressor.
Refrigerant Charge:- The required amount of refrigerant in a system.
R-410A Refrigerant:- A chlorine-free refrigerant that meets environmental guidelines.
Resistance, or R Value:- The tendency of a material to retard the flow of heat.
Retrofitting:- The application of a solar heating or cooling system to an existing site.
Reversing Heat Pump:- A heat pump in which the condenser and evaporator coils of the unit reverse roles in response to a reverse in the direction of the flow of refrigerant in the machine.
Risers:- The flow channels or pipes that distribute the heat transfer liquid across the face of an absorber.
Scroll Compressor:- A specially designed compressor that works in a circular motion, as opposed to a reciprocating up-and-down piston action.
Seasonal Coefficient of Performance (SCoP):- Is the average CoP over the entire heating period.
Seasonal Efficiency:- The ratio, over an entire heating period, of solar energy collected and used to the solar energy striking the collector.
Seasonal Energy Efficiency Ratio (SEER):-The average cooling efficiency over an entire cooling period.
Self-contained System:- A refrigerating system that can be moved without disconnecting any refrigerant lines; also know as a package unit.
Selective Surface:- A surface that absorbs radiation of one wavelength (fe.g., sunlight) but emits little radiation of another wavelength (for example, infrared); used as a coating for absorber plates.
Sensible:- The internal heat gain due to heat conduction, convection, and radiation from the external into the internal, and from appliances.
Sensible Heat:- That heat which, when added to or taken away from a substance, causes a temperature rise or fall.
Sensor:- Any device that reacts to a change in the conditions being measured, permitting the condition to be monitored and controlled.
SEPA:- Acronym for the terms, Silent, Economic, Powerful, Advanced
Setpoint:- The temperature or pressure at which a controller is set with the expectation that this will be a nominal value depending on the range of the controller.
Shading Coefficient:- The ratio of the solar heat gain through a specific glazing system to the total solar heat gain through a single layer of clear double-strength glass.
Simple Payback Factor (heating):- Subtract the installation cost of the least expensive (less efficient) system from the installation cost of the more expensive (more efficient) heating system. This value is the increased cost of installing the more efficient system. Calculate the yearly energy savings by installing the more efficient system. Take the increased cost to install divided by the yearly energy savings and the result is the number of years required for the more efficient system to pay for itself.
Sink Temperature:- This is the temperature of the media (water or air) into which the heat pump must reject its heat.
Sleeve bearing:- Sleeve bearings of porous, sintered iron or bronze alloys are used in fans. The liquid lubricant is stored in the sintered pores and is discharged when the shaft rotates. Due to the hydrodynamic effect, a lubricant cycle is created that only functions freely in a relatively tight temperature range (approximately 0 to 60°C). Due to surface errors, micro-contamination and natural wear during running the fan in and running down, reliability is considerably lower than with ball bearings. Sleeve bearings are frequently used due to their favourable price.

Silencer System:- Carrier Silencer System ensures quite operation inside and out, typically achieved using quiet motor mounts, a compressor sound blanket, forward swept fan blades, a laminated sound elimination compressor mounting plate, and integrated silencer airflow baffles.
Split System:- Split heat pumps are two part refrigeration systems which have separate evaporator / air handler and compressor / condenser sections. Commonly employed in air-to-air systems where the condenser section is located externally while the evaporator / air handler is located inside the conditioned structure.
Spine Fin Coil:- All-aluminium outdoor coil with a spine fin design which provides greater heat exchanging capabilities (meaning higher efficiencies) and is more resistant to corrosion than traditional copper/aluminium.
Solar Constant:- The average intensity of solar radiation reaching the earth outside the atmosphere; amounting to 1395 W/m².
Solar Radiation (Solar Energy):- Electromagnetic radiation emitted by the sun.
Source Temperature:- This is the temperature of the media (water or air) from which the heat pump extracts its heat.
Specific Heat:- The quantity of heat, in BTU, needed to raise the temperature of one pound of a material 1°F.
Standby Heat Loss:- Heat lost though the storage tank and piping walls.
Sun Path Diagram:- A circular projection of the sky vault, similar to a map, that is used to determine solar positions and to calculate shading.
Thermal Capacity:- The quantity of heat needed to warm a collector to its operating temperature.
Thermal Mass or Thermal Inertia:- The tendency of a structure with large quantities of heavy materials to remain at the same temperature or to fluctuate only slowly; also the overall heat storage capacity of the building.
Thermal Radiation:- Electromagnetic radiation emitted by a warm body.
Thermostat:- A thermostat consists of a series of sensors and relays that monitor and control the functions of a heating and cooling system.
Thermidistat:- Monitors temperature and humidity and adjusts heating or cooling system to maintain the desired levels.
Thermistor:- Sensing device which changes its electrical resistance according to temperature. Used in the control system to generate input data on collector and storage temperatures.
Thermosyphoning:- The process that makes water circulate automatically between a warm collector and a cooler storage tank above it. (See Gravity Convection).
Tilt Angle:- The angle that a flat plate collector surface forms with the horizontal plane.
Ton (of refrigeration):- The amount of energy it takes to convert 2000 lbs. of water at 32º F to ice at 32º F during a 24 hour period. Calculation: 2000 lbs.H2O x 144 Btu/lb. = 288,000 Btu's in 24 Hrs. Divide by 24 hrs = 12,000 BTU/hr. Therefore a 'ton' of cooling is a measure of heat energy which is roughly equivalent to 12,000 BTU's.
Temperature difference.:- Difference between ELT and a LLT.
Trickle Type Collector:- A collector in which the heat transfer liquid flows through metal tubes which are fastened to the absorber plate by solder, clamps or other means. (See Collector).
Tube-in-Plate-Absorber:- A metal absorber plate in which the heat transfer fluid flows through passages formed in the plate itself.
Two-stage heating / Two-stage cooling:- Two-stage heating and cooling is considered to be more efficient, because it operates at low speed most of the time. However, when more heating or air conditioning is required, it switches to the next stage for maximum performance.
TX Valve:- A temperature and pressure controlled device for metering refrigerant in a heat pump or other refrigeration device.
U-Factor:- The factor representing resistance to heat flow of various materials.
Ultraviolet Radiation:- Electromagnetic radiation with wavelengths slightly shorter than visible light.
Upflow:- Air enters at the bottom of the unit and is discharged vertically out the top.
Upflow Furnace:- A furnace in which air is drawn in through the sides or bottom and discharged out the top.
Vacuum:- A pressure below atmospheric pressure. A perfect vacuum is 30 inches Mercury (elemental symbol 'Hg').
Ventilator:- A system that exchanges old, recirculated indoor air with fresh, filtered outside air.
Water-to-Water:- A heat pump which extracts heat from water in one area and transfers the heat usually at a higher temperature to another body of water. For example, extracting heat from a 50°F. source and using it to heat domestic hot water at 120° F.

Glossary of Magnetic terminology

Absolute Permeability:- The permeability of a magnetic material expressed in actual physical units, not relative to permeability of free space. The permeability of magnetic materials is rarely expressed in terms of absolute permeability. The usual mode is in terms of relative permeability.
Aging:- Change in magnetic properties, especially B_r, with time.
Air gap:- A low permeability gap in the flux path of a magnetic circuit. A non-magnetic discontinuity in a ferro-magnetic circuit. Often air, but inclusive of other materials such as paint, aluminium, etc.
Air gap volume V_g:- The useful volume of air or nonmagnetic material between magnetic poles; measured in cubic centimetres.
Amorphous:- Refers to magnetic materials that are metallurgically non-crystalline in nature.
Anisotropic magnet:- A magnet having a preferred direction of magnetic orientation, so that the magnetic characteristics are optimum in one preferred direction.
Anisotropy:- Having different properties depending on the inspected direction. Magnets which are anisotropic, or have an easy axis of magnetization, have their anisotropy developed by two methods: Shape and Magnetocrystalline.
Material that have a preferred magnetization direction. These materials are typically manufactured in the influence of strong magnetic fields, and can only be magnetized through the preferred axis.
Anneal:- A high-temperature conditioning of magnetic material to relieve the stresses introduced when the material was formed. To prevent oxidation, the anneal is usually performed in a vacuum or inert-gas atmosphere.
Antiferromagnetic:- Materials in which the internal magnetic moments line up antiparallel, resulting in permeabilities slightly greater than unity; unlike paramagnetic substances, these materials exhibit hysteresis and have a Curie Temperature. Examples include manganese oxide, nickel oxide and ferrous sulphide.
Area of the air gap, A_s:- or the cross sectional area of the air gap perpendicular to the flux path, is the average cross sectional area of that portion of the air gap within which the application interaction occurs. Area is measured in sq. cm. in a plane normal to the central flux line of the air gap.
Area of the magnet, A_m:- The cross sectional area of the magnet perpendicular to the central flux line, measured in sq. cm. at any point along its length. In design, A_m is usually considered the area at the neutral section of the magnet.
Barkhausen Effect:- The series of irregular changes in magnetization that occur when a magnetic material is subjected to a change in magnetizing force.
Bonded Magnets:- Consisting of powdered permanent magnet material, usually isotropic ceramic ferrite or neodymium iron-boron, and a polymer binder, typically rubber or epoxy. This magnet material can be moulded into complex shapes.
$B_d / \mu_o H_d$:- Slope of the operating line, is the ratio of the remnant induction, B_d, to a demagnetizing force, H_d. It is also referred to as the permeance coefficient, shear line, load line and unit permeance.
$B_d H_d$:- Energy product, indicates the energy that a magnetic material can supply to an external magnetic circuit when operating at any point on its demagnetization curve; measured in kiloJoules per cubic meter (kJ/m^3).
BH_{max} Maximum energy product:- The maximum product of $B_d H_d$ which can be obtained on the demagnetization (normal) curve, that is, in the second (fourth) quadrant of the hysteresis loop.
BH Curve:- See Demagnetization Curve.
BH Loop:- A hysteresis loop across four quadrants.
B_g, Magnetic induction in the air gap:- The average value of magnetic induction over the area of the air gap, A_g. Also defined as the magnetic induction measured at a specific point within the air gap; measured Tesla.
B_i (or J), Intrinsic induction:- The contribution of the magnetic material to the total magnetic induction, *B*. It is the vector difference between magnetic induction in the material and magnetic induction that would exist in a vacuum under the same field, *H*.
B_m:- Maximum induction.
B_r, Residual induction:- The magnetic induction which corresponds to zero applied field (magnetizing force) in a magnetic material after saturation in a closed circuit; measured in Tesla.
Carbonyl Iron:- A relatively expensive iron powder used in low-permeability, high frequency powdered iron cores.
Ceramic Ferrite:- A relatively inexpensive permanent magnet material with moderate coercivity and low energy product that is composed of strontium or barium oxide and iron oxide.
Closed Circuit:- This exists when the flux path external to a permanent magnet is confined within high permeability materials that compose the magnet circuit.
Closed circuit condition: A condition that exists when the external flux path of a permanent magnet is confined with high permeability material.

Coercive Force, H_c:- The demagnetizing force, measured in At/m, necessary to reduce observed induction, *B*, to zero after the magnet has previously been brought to saturation. It is expressed in At/m.

Coercive force, H_k:- The value of H_{ci} at $0.9B_r$. This value gives an indication of the squareness of the intrinsic curve. The more square the intrinsic curve, the closer the material is to being ideal. H_k values that approach the H_{ci} values are considered extremely good materials.

Coercive force of a material, H_c:- Equal to the demagnetizing force required to reduce residual induction, B_k to zero in a magnetic field after magnetizing to saturation; measured in At/m.
The material characteristic of coercivity is taken as the maximum coercivity - that value of *H* required to reduce the residual induction to zero after the material has been saturated (fully magnetized).

Coercivity, H_{ci} or iH_c:- The resistance of a magnetic material to demagnetization. It is equal to the value of *H* where the intrinsic curve intersects the H axis in the second quadrant of the hysteresis loop. It is expressed in At/m.

Core Loss:- Power lost in a magnetic material when flux density changes. Also called iron losses or excitation losses, mainly consisting of hysteresis and eddy current losses.

Curie Temperature, T_c:- The transition temperature above which the alloy loses its magnetic properties. It is not the maximum serviceable temperature, which is usually much lower. The temperature at which the parallel alignment of elementary magnetic moments completely disappears, and the material is no longer able to hold magnetization. Most references state that the ferromagnetic material becomes paramagnetic (weakly magnetic).

Current Density:- The amps per unit of cross-section in the conductor.

DC Bias:- Direct Current (DC) applied to the winding of a core in addition to any time-varying current. Inductance with DC bias is a common specification for powder cores. The inductance decreases or rolls-off gradually and predictably with increasing DC bias.
DC Stress:- Annealing a magnetic material in the presence of a DC magnetic field to enhance magnetic properties.

Demagnetization curve:- The second (or fourth) quadrant of a major hysteresis loop generally describing the behaviour of magnetic characteristics in actual use. Also known as the *B-H* Curve. That portion of the hysteresis loop which lies between the residual induction point, B_r, and the coercive force point, H_c (normal curve) or H_{ci} (intrinsic curve). Points on this curve are designated by the coordinates B_d and H_d.

Demagnetized:- A material condition where a ringing AC field has reduced the remanent induction to or near zero. A ringing AC field is a continually decreasing sinusoidal field. A pulsed DC field can be used to achieve gross demagnetization, but with much effort and with residual local magnetization.

Diamagnetic Material:- A material with magnetization directed opposite to the magnetizing field, so that the relative permeability is less than one; metallic bismuth is an example.

Dipole (Magnetic):- An arrangement of one or more magnets to form a magnet system that produces a magnetic field with one pair of opposite poles.

Direction of magnetization:- Refers to the 'easy axis' or the axis of choice for the direction of alignment. Most rings are aligned axially so the direction of magnetization is through the axis (or thickness). Other possibilities for rings would include 'across the diameter' and 'radial'.

Distributed Air Gap:- Major feature of powder cores. It is the cumulative effect of many small gaps distributed evenly throughout the core. In a typical MPP core, the number of separate air gaps results from the use of powder to construct the core and numbers in the millions. The result is minimal fringing flux density compared to a core with one or two air gaps in the magnetic path. (Flux that passes around a discrete air gap and through the sides of a core is fringing. Fringing flux enters the surrounding winding and causes a substantial amount of eddy current loss.)

Domains:- Areas in a magnetic alloy which have the same orientation. The magnetic domains are regions where the atomic moments of atoms cooperate and allow for a common magnetic moment. It is the domains which are rotated and manipulated by an external magnetizing field to create a useful magnet which has a net magnetic moment. In un-magnetized material the domains are un-oriented and cancel each other out. In this condition there is no net external field.

Eddy currents:- Circulating electrical currents that are induced in electrically conductive elements when exposed to changing magnetic fields, creating an opposing force to the magnetic flux. Eddy currents can be harnessed to perform useful work (such as dampening of movement), or may be unwanted consequences of certain designs, which should be accounted for of minimized.

Eddy Current Loss:- Core loss associated with the electrical resistivity of the magnetic material and induced voltages within the material. Eddy currents are inversely proportional to material resistivity and proportional to rate of change of flux density. Eddy current and hysteresis losses are the two major core loss factors. Eddy current loss becomes dominant in powder cores as the frequency increases.

Electrical Resistivity:- The electrical resistance to current flow in ohms per unit length of the material being evaluated.

Electromagnet:- A magnet, consisting of solenoid with a permeable material such as iron core, which has a magnetic field existing only during the time of current flow through the coil.

Energy Product:- Indicates the energy that a magnetic material can supply to an external magnetic circuit when operating at any point on its demagnetization curve. Calculated as $B_d \times H_d$, and measured in kiloJoules per cubic meter (kJ/m^3).

Epoxy Impregnated:- Cut cores are impregnated with an epoxy to make the core rigid. No insulative purpose is intended.

Epstein Test:- A standardized method of evaluating unprocessed thin-gauge alloy for core loss and permeability.

Excitation Current:- The current which produces magnetic energy (or flux) in an inductor.

Ferrimagnetic Material:- An antiparallel alignment of adjacent atomic moments is present as in antiferromagnetic materials, but the moments are not equal. The response to an external magnetic field is therefore large, although smaller than that for a ferromagnetic material. Ferrites are the most important example of this class of material.

Ferrites:- A soft ferrite material that has lower permeability with very low eddy current loss. The common ferrites are nickel-zinc, manganese-zinc and magnesium-zinc ferrite.

Ferromagnetic material:- A material whose permeability is very much larger than 1 (from 60 to several thousands times 1), and which exhibits hysteresis phenomena. A material in which internal magnetic moments spontaneously line up parallel to each other to form domains, resulting in relative permeabilities considerably higher than unity (in practice, 1.1 or more); examples include iron, nickel and cobalt.

Ferromagnetism:- Ferromagnetic materials have atomic fields that align themselves parallel with externally applied fields creating a total magnetic field much greater than the applied field. Ferromagnetic materials have permeabilities much greater than 1. Above the Curie temperature, the ferromagnetic materials become paramagnetic.

Flux:- The condition existing in a medium subjected to a magnetizing force. This quantity is characterized by the fact that an electromotive force is induced in a conductor surrounding the flux at any time the flux changes in magnitude. The MKS unit of flux is the Weber.

Flux Density:- *Magnetic, B* - The fundamental magnetic force field. Flux means to flow (around a current carrying conductor, for example) and "density" refers to its use with an enclosed area and Faraday's Law to determine induced voltage. Also called the 'induction field.' From Faraday's Law, the MKSA unit of flux density is a volt-second per square meter per turn or Tesla.

Flux loss:- Refers to the change (loss) in magnetic strength of a magnet, which occurs as a result of temperature stabilization. Also known as irreversible loss. Once it occurs, the only way to regain the flux loss is to re-magnetize the magnet. Under normal circumstances, flux loss is limited to a few percent.

Flux meter:- An instrument that measures the change of flux linkage with a search coil or Helmholtz coil. A flux meter is basically a voltage integrator, which is an integrating device totalizing the voltage output with respect to time.

Fringing fields:- Leakage flux particularly associated with edge effects in a magnetic circuit. The field(s) associated with the divergence of the flux from the shortest path between poles in a magnetic circuit. Where flux passes from a high permeability into a lower permeability material, the flux redistributes.

Gauss:- Lines of magnetic flux per square centimetre, cgs unit of flux density, equivalent to lines per square inch in the English system, and Webers per square meter or Tesla in the SI system. One Tesla is equal to one Weber per square centimetre (metre).

Gauss meter:- An instrument that measures the instantaneous value of magnetic induction, *B*. Its principle of operation is usually based on one of the following: the Hall effect, nuclear magnetic resonance (NMR), or the rotating coil principle.

Gilbert:- The unit of magneto motive force, *F*, in the cgs electromagnetic system. MKS equivalent is ampturns, At

Grain Oriented:- Silicon steel or other granular magnetic material that has a preferred direction of magnetization.

Hall Effect Transducer:- A device which produces a voltage output dependent upon an applied DC voltage and an incident magnetic field. The magnitude of the output is a function of the field strength and the angle of incidence with the Hall device.

Hard Ferrite:- Same as ceramic ferrite.

Hard Magnetic Material:- A permanent magnet material that has an intrinsic coercivity greater than or equal to about 24 kA/m. A ferromagnetic material that retains its magnetization when the magnetizing field is removed; a magnetic material with significant coercivity.

H_c, Coercive Force, or Coercivity:- Is equal to the demagnetizing field required to reduce the B field in the magnet to zero after the magnet has been fully saturated; measured A/m.

H_{ci}, Intrinsic Coercive Force, or Intrinsic Coercivity:- That value of *H* corresponding to the remnant induction, B_d; on the demagnetization curve, measured in At/m. Represents the ability of the

magnetic materials to resist demagnetization. It is equal to the demagnetizing field that reduces the B field in the magnet to zero (from saturation).

H_d, B_d:- Operating point on demagnetisation curve.

Henry:- A unit of inductance.

H_{mv}, H_m:- That value of H corresponding to the recoil induction, B_m,; measured in At/m. Common symbol for maximum applied magnetizing force.

H_o:- The magnetic field strength at the point of the maximum energy product BH_{max}; measured in At/m.

H_s:- Net effective magnetizing force, is the minimum magnetizing force required in the material, to magnetize to saturation measured in At/m.

Hysteresis and Hysteresis Loss:- Hysteresis is the tendency of a magnetic material to retain its magnetization. Hysteresis causes the graph of magnetic flux density versus magnetizing force to form a loop rather than a line. The area of the loop represents the difference between energy stored and energy released per unit volume of material per cycle. This difference is called hysteresis loss. It is one of two major loss mechanisms in inductor cores; the other is eddy current loss. Hysteresis loss is measured at low frequency to distinguish it from eddy current loss.

Hysteresis loop:- A closed curve obtained for a material by plotting (usually to rectangular coordinates) corresponding values of magnetic induction, *B*, for ordinates and magnetizing force, *H*, for abscissa when the material is passing through a complete cycle between definite limits of either magnetizing force, *H*, or magnetic induction, *B*. If the material is not driven to saturation, it is said to be on a minor loop.

Hysteresis Loop, Major:- Of a material is the closed loop obtained when the material is cycled between positive and negative saturation.

Hysteresis, Magnetic:- The property of a magnetic material by virtue of which the magnetic induction for a given magnetizing force depends upon the previous conditions of magnetization.

Induction, B:- The magnetic flux per unit area of a section normal to the direction of flux. Expressed in Tesla.

Induced Flux Density:- The flux density generated in a core (or soft magnetic material) by the applied MMF.

Inductance:- Inductance is the ratio of voltage to time rate of change of current. By definition, it has dimensions of volt-seconds per ampere. A volt-second per ampere is called a Henry.

Inductance Factor A_L:- Core constant used to calculate inductance based on the number of winding turns squared. Value is given in millihenries per 1000 turns squared, which is the same as nanohenries per turn squared.

$$L = A_L N^2 \text{ nanohenries}$$

Induction, B:- Magnetic induction, B, is the magnetic field induced by an applied field, *H*. It is measured as the flux per unit area normal to the direction of the magnetic path.

Induction Curve, Normal:- A graph depicting the relation between normal induction B and magnetizing force H, where B corresponds to the sum of the externally applied field, H, and the magnetic flux from the magnetic material, B_i.

Inductor:- A coil that has significant self inductance, typically many turns of wire and with a permeable core. It is a device that stores and releases electromagnetic energy.

Initial Permeability:- The relative permeability of a magnetic material at a very low flux level.

Insulator, Insulation:- Opposite of conductor, that is, does not conduct an electrical current. In soft magnetic cores, refers to electrical insulation between adjacent laminations, layers of thin gauge tape, or powder particles. Also associated with some of the finishes, which have dielectric capacity, applied to cores.

Intrinsic Coercive Force, H_{ci}:- Measured in At/m, this is a measure of the materials inherent ability to resist demagnetization. It is the demagnetization force corresponding to zero intrinsic induction in the magnetic material after saturation. Practical consequences of high H_{ci} values are seen in greater temperature stability for a given class of material, and greater stability in dynamic operating conditions.

Intrinsic coercive force of a material, H_{ci}:- Indicates its resistance to demagnetization. It is equal to the demagnetizing force which reduces the intrinsic induction, B_i, in the material to zero after magnetizing to saturation; measured in At/m. This quantity is used to gage the field required to magnetize a material and its ability to resist demagnetization.

Intrinsic Coercivity:- Same as H_{ci}. Indicates a material's resistance to demagnetization. It is equal to the demagnetizing force which reduces the intrinsic induction, B_i, in the material to zero; measured in At/m. As for coercivity, the maximum value of intrinsic coercivity is obtained after the material has been saturated (fully magnetized).

Intrinsic Demagnetization Curve:- The hysteresis loop corresponding to *B* versus *H* where *B* is the magnetization resulting from only the magnetic material. For the Normal Curve, B corresponds to the sum of the externally applied field and the field of the magnetic material. The second quadrant portion of the hysteresis loop generated when intrinsic induction B_i is plotted against applied field *H*, which is mathematically related to the normal curve; most often used to determine the effects of demagnetizing (or magnetizing) fields. Also known as the intrinsic B versus H curve.

Intrinsic induction, B_i (or J):- The contribution of the magnetic material to the total magnetic induction, *B*. It is the vector difference between the magnetic induction in the material and the magnetic induction that would exist in a vacuum under the same field strength, *H*. This relation is expressed by the equation: $B_i = B - H_{em}$ where: B_i = intrinsic induction in Tesla; *B* = magnetic induction in Tesla; H_{em} = field strength in kA/m.

Irreversible (flux) loss:- Defined as partial demagnetization of the magnet, caused by exposure to high or low temperatures external fields or other factors, such as mechanical shock. These losses are recoverable by remagnetisation. Magnets can be stabilized against irreversible losses by partial demagnetization induced by temperature cycles or by external magnetic fields. Stabilization results in the loss prior to placing the magnet in the application and the application is designed around the output of the stabilized magnet.

Isotropic:- A magnetic material that has the same magnetic properties in all directions. Such a material may be magnetized in any direction since it does not have a preferred alignment direction. Also known as unoriented material. Most magnetic materials are anisotropic as cast or powdered: each crystallite has a preferred direction of magnetic orientation. If the particles are not physically oriented during manufacture of the magnet, this results in a random arrangement of the particles and magnetic domains and produces isotropic magnet properties. Conversely, orienting the material during processing results in an anisotropic magnet.

Isotropic Magnet:- A magnet material whose magnetic properties are the same in any direction, and which can therefore be magnetized in any direction without loss of magnetic characteristics.

J:- see B_i Intrinsic induction.

J_s:- see B_{is}, Saturation intrinsic induction.

Keeper:- A piece (or pieces) of soft iron that is placed on or between the pole faces of a permanent magnet to decrease the reluctance of the air gap and thereby reduce the flux leakage from the magnet. It also makes the magnet less susceptible to demagnetizing influences.

Keepers:- A keeper is a high permeability material, typically mild steel, which is installed on a magnet or magnetic assembly to reduce the reluctance of the magnetic circuit. This reduces the overall leakage fields generated by the magnet or magnetic assembly. Keepers are typically installed to help the magnet or magnetic assembly resist demagnetization during handling, transportation, or storage. Keepers are typically found on Alnico magnets and Alnico magnetic assemblies.

Knee (of the demagnetization curve):- The point at which the *B-H* curve ceases to be linear. All magnet materials, even if their second quadrant curves are straight line at room temperature, develop a knee at some temperature. Alnico 5 exhibits a knee at room temperature. If the operating point of a magnet falls below the knee, small changes in H produce large changes in B, and the magnet will not be able to recover its original flux output without re-magnetization.

Leakage factor, k_f:- Accounts for flux leakage from the magnetic circuit. It is the ratio between the magnetic flux at the magnet neutral section and the average flux present in the air gap. $k_f = (B_m A_m)/(B_g A_g)$

Leakage flux:- The flux, ø, whose path is outside the useful or intended magnetic circuit; measured in Weber. That portion of the magnetic flux that is lost through leakage in the magnetic circuit due to saturation or air-gaps, and is therefore unable to be used.

Leakage Inductance:- The inductance associated with the leakage flux of a core coil.

Legg's Equation:- An expression for total core loss at low flux densities. The sum of hysteresis loss, residual loss and eddy current loss. The equation is:

$$R_{ac} / \mu L = a B_{max} f + cf + ef^2$$

where

R_{ac} = effective resistance due to core losses
μ = permeability of the core
L = inductance in henries
a = hysteresis loss coefficient
B_{max} = maximum flux density in Tesla
f = frequency
c = residual loss coefficient
e = eddy current loss coefficient

Length of the air gap, ℓ_g:- The length of the path of the central flux line of the air gap; measured in centimetres. It is important to distinguish between the magnetic length of the gap and the physical length; for magnetic circuit calculations, any nonmagnetic material in the flux path is equivalent to air and contributes to the (magnetic) gap.

Length of the magnet, ℓ_m:-The total length of magnet material traversed in one complete revolution of the centreline of the magnetic circuit; measured in centimetres. The distance between the magnetic poles. (Measured in centimetres when using the cgs system for calculations).

Litz Wire:- A special type of wire that consists of many strands (sometimes hundreds) of magnet wire

woven together to form a single conductor. This type of wire offers advantages over single strand at high frequency.

L_m / D Dimension ratio:- The ratio of the length of a magnet to its diameter, or the diameter of a circle of equivalent cross-sectional area. For simple geometries, such as bars and rods, the dimension ratio is related to the slope of the operating line of the magnet, B_d/H_a.

Load line:- Graphically, a line drawn from the origin of the demagnetization curve with a slope of *B/H*, the intersection of which with the second quadrant *B-H* curve represents the operating (working) point, H_d, B_d, of the magnet. Graphic representation of permeance. Also see permeance coefficient.

Mag amp (Magnetic Amplifier):- A device that utilizes a square loop core material to provide a series impedance. The impedance is switched off at a predetermined time during a voltage pulse.

Magnet:- Any object that can sustain an external magnetic field.

Magnetic Bias:- A constant magnetic field on which is superimposed a variable, often sinusoidal, perturbation magnetic field in devices like magnetic bearings (Tesla (T)).

Magnet Wire:- Copper or aluminium wire with electrical insulating material applied to the surface to prevent continuity between adjacent turns in a winding.

Magnetic Assemblies:- A combination of materials, magnetic and non-magnetic, which form a particular solution. Incorporates a permanent magnet as the flux generator and usually relies on mild steel to conduct the flux to the workface. Allows for better means of mounting-tapped holes, threads, press fits, etc.

Magnetic circuit:- An assembly consisting of some or all of the following: permanent magnets, ferromagnetic conduction elements, air gaps through or around which the magnetic flux path passes.

Magnetic Energy:- The product of the flux density *B* in a magnetic circuit and the (de)magnetizing force *H* required to reach that flux density. See Energy Product.

Magnetic field strength, H:- (magnetizing or demagnetizing force), The measure of the vector magnetic quantity that determines the ability of an electric current, or a magnetic body, to induce a magnetic field at a given point; measured in At/m.

Magnetic Flux, ø:- A contrived but measurable concept that has evolved in an attempt to describe the flow of a magnetic field. The total magnetic induction over a given area. When the magnetic induction, *B*, is uniformly distributed over an area *A*, Magnetic Flux = *BA*. Is a contrived but measurable concept that has evolved in an attempt to describe the flow of a magnetic field. Mathematically, it is the surface integral of the normal component of the magnetic induction, *B*, over an area, *A*.

$$ø = \iint B \cdot dA$$

where:

ø = magnetic flux, in Weber
B = magnetic induction, in Tesla
dA = an element of area, in square centimetres

When the magnetic induction, *B*, is uniformly distributed and is normal to the area, *A*, the flux, *ø=BA*.

Magnetic Flux Density, B: A vector quantifying a magnetic field, so that a particle carrying unit charge experiences unit force when travelling with unit velocity in a direction perpendicular to the magnetic field characterized by unit magnetic flux density (Tesla (T)).

Magnetic induction, B_o:- Magnetic induction at the point of the maximum energy product BH_{max}; measured in Tesla.

Magnetic Induction in the Air Gap, B_g:- The average value of magnetic induction over the area of the air gap, A_g; or it is the magnetic induction measured at a specific point within the air gap; measured in Tesla.

Magnetic induction, B:- The magnetic field induced by a field strength, H, at a given point. It is the vector sum, at each point within the substance, of the magnetic field strength and resultant intrinsic induction. Magnetic induction is the flux per unit area normal to the direction of the magnetic path. The flux density within a magnetic material when driven by an external applied field or by its self demagnetizing field, which is the vector sum of the applied field and the intrinsic induction (Tesla (T)).

Magnetic Length:- The effective distance between the north and south poles within a magnet, which varies from 0.7 (Alnico) to 1.0 (NdFeB, SmCo, hard ferrite) times the physical length of the magnet.

Magnetic (path) Length:- The physical length of the magnet dimension which corresponds to the direction the magnet is magnetized. This may or may not be the magnet's orientation direction. The length of the closed path that magnetic flux follows around a magnetic circuit. Ampere's Law determines it.

Magnetic Line of Force:- An imaginary line representing a magnetic field, which at every point has the direction of the magnetic flux at that point. Flux is a vector quantity having both magnitude and

direction.

Magnetic Losses:- The loss of flux in a magnetic circuit, primarily due to leakage and fringing.

Magnetic Orientation:- The preferred direction of magnetization for an anisotropic magnetic material.

Magnetic saturation:- Of a material exists when an increase in magnetizing force, *H*, does not cause an increase in the intrinsic magnetic induction, *B*, of the material.

Magnetic Stabilization:- The act of purposely demagnetizing a magnet with reverse fields or a change in temperature so that no irreversible losses are experienced when the magnet operates under similar conditions in the field.

Magnetic Susceptibility:- The ratio of the magnetization to the applied external field; an indicator of how easily a material is magnetized.

Magnetizing field (H):- An applied magnetic field used to drive another material to a condition of being magnetized. It may be applied by current through a coil of wire or by using permanent magnets to generate the applied field.

Magnetizing Force, H:- The magnetomotive force per unit length at any point in a magnetic circuit. An applied magnetic field used to drive another material to a condition of being magnetized. It may be applied by current through a coil of wire or by using permanent magnets to generate the applied field. Measured in At/m.

Magnetomotive Force, F:- (magnetic potential difference), Analogous to voltage in electrical circuits, this is the magnetic potential difference between any two points. Most commonly produced by a current flowing through a coil of wire where its magnitude is proportional to the current, and to the number of turns.

$$F = NI$$

where *I* is in amperes and *N* is the number of turns
The line integral of the field strength, *H*, between any two points, p_1 and p_2.

$$F = \int_{p_1}^{p_2} H\, dl$$

F = magneto motive force in At
H = field strength in At/m
dl = an element of length between the two points, in centimetres.
The rationalized unit is the ampere-turn (*ni*).
Magnetomotive force may also result from a magnetized body.

Magnetostriction:- The expansion and contraction of a magnetic material with changing magnetic flux density. The saturation magnetostriction coefficient has the symbol □s. It is change of length divided by original length (a dimensionless number) and is measured at the saturation flux density. Magnetostriction causes audible noise if the magnetostriction is sufficiently large and the applied field is AC and in the audible frequency range, e.g., 50 or 60 Hz.

Major hysteresis loop:- Material closed loop obtained when the material is cycled between positive and negative saturation. For a magnetic material, the loop generated as intrinsic or magnetic induction (B_i or *B*) is plotted with respect to applied field *H* when the material is driven from positive saturation to negative saturation and back, showing the lag of induction with respect to applied field.

Manganese-Zinc Ferrites:- A soft magnetic material used in powder cores and characterized by very low eddy current loss. Used for transformer and inductor cores. Compared to nickel-zinc ferrites, they have higher saturation flux density but with greater loss with high frequency current.

Maximum Energy Product, BH_{max}:- The point on the Demagnetization Curve where the product of *B* and *H* is a maximum and the required volume of magnet material required to project a given energy into its surroundings is a minimum. Measured in kiloJoules per cubic meter (kJ/m^3).

Maximum Operating Temperature, T_{max}:- The maximum operating temperature, also known as maximum service temperature, is the temperature at which the magnet may be exposed to continuously with no significant long-range instability or structural changes. A proposed magnetic definition is that the hysteresis normal curve is substantially a straight line in the second quadrant up to the T_{max} temperature and becomes curved above T_{max}. Note that this temperature is a function of the operating point of the magnet, and not an absolute value.

Maxwell:- The unit of magnetic flux in the cgs electromagnetic system. One Maxwell is one line of magnetic flux. MKS equivalent is Weber.

Mean Length Turn:- The average length of a single turn in the winding of the device.

Minor Hysteresis Loop:- A hysteresis loop generated within the major hysteresis loop when a magnetic material is not driven to full positive or negative saturation.

MMF:- Magneto-motive force.

MMF Drops:- The portions of a magnetic circuit that "consume" the applied MMF. Analogous to voltage drop in an electrical circuit.

Mu-metal:- A nickel-iron alloy typically containing more than 65% nickel used for shielding magnetic flux. The name of the material refers to the Greek letter, μ (mu), which is the symbol for magnetic

permeability. Mu-metal has a high value of magnetic permeability.

Multifilar Winding:- A winding technique where a single turn consists of two or more stands of magnet wire operating in parallel. This reduces some of the second order effects associated with a single strand of wire. Typical would be a bifilar, trifilar, etc.

Neodymium-Iron-Boron (NdFeB):− A high energy magnetic material composed of the three nominal elements and other additives, characterized by a high residual induction and high coercivity. NdFeB has a high magnetic temperature coefficient, which is undesirable for high temperature use.

Net permeability:- The permeability of a magnetic circuit when all materials, air gaps, and applied mmf's are taken into account; it is the same as effective permeability.

Neutral section:- part of a permanent magnet defined by a plane passing through the magnet perpendicular to its central flux line at the point of maximum flux.

Nickel-Zinc Ferrites:- A soft ferrite material that has lower permeability with very low eddy current loss. The other common ferrites are manganese-zinc and magnesium-zinc.

Normal Demagnetization Curve:− The second quadrant portion of the hysteresis loop generated when magnetic induction *B* is plotted against applied field *H*, which is mathematically related to the intrinsic curve; used to determine the performance of a magnet in a magnetic circuit. Also known as the normal *B* versus *H* curve.

North pole:- is the pole of a magnet which, when freely suspended, would point to the north magnetic pole of the earth. The definition of polarity can be a confusing issue, and it is often the best to clarify by using "north seeking pole" instead of "north pole" in specifications.

Oersted:- The unit of magnetic field strength, H, in the cgs electromagnetic system. One Oersted equals a magneto motive force of one gilbert per centimetre of flux path. A cgs unit of measure used to describe magnetizing force. The SI systems is Ampere turns per meter, (At/m).

Open circuit condition:- Exists when a magnetized magnet is by itself with no external flux path of high permeability material.

Operating (load) line:- For a given permanent magnet circuit it is a straight line passing through the origin of the demagnetization curve with a slope of negative B_d / H_d. Although the slope is negative, by convention the values are usually referred to in the absolute value of the slope. (Also known as permeance coefficient line.)

Operating point:- That point on a demagnetization curve of a permanent magnet defined by the coordinates H_d, B_d or that point within the demagnetization curve defined by the coordinates B_m H_m.

Orientation direction:- The direction in which an anisotropic magnet should be magnetized in order to achieve optimum magnetic properties. Also known as the axis, easy axis, or angle of inclination.

Oriented (anisotropic) material:- One that has better magnetic properties in a given direction. Material with a preferred direction of magnetization. This type of material should be magnetized only through this preferred direction. Trying to magnetize through the other directions will result in substantial losses in magnetic properties, and the data provided will not be valid.

Paramagnetic material:- A material having a permeability slightly greater than 1. Sodium, Potassium, Oxygen and the rare earth elements are examples.

Permalloy:- 4-79 Molybdenum Permalloy. A high permeability alloy of 4% molybdenum, 79% nickel, 17% iron used to make tape-wound and laminated cores and other components in a magnetic circuit. See Mu-Metal.

Permanent Magnet Material:- Shaped piece of ferromagnetic material, which once having been magnetized, shows definite resistance to external demagnetizing forces, that is, requires a high demagnetizing force to remove the residual magnetism.

Permeability, μ:- The general term used to express various relationships between magnetic induction, *B*, and the field strength, *H*. The ratio of the ability of a material to carry magnetic flux in comparison to air or a vacuum, the permeability of which is, by definition, one.

Permeability, Incremental:- The ratio of change in magnetic flux density to change in magnetic field (magnetizing force).

$\mu = (1/\mu_o)\Delta B/\Delta H$ in MKSA units

$\mu = \Delta B/\Delta H$ in CGS units

The magnetic field variations are small or incremental and can be in addition to a steady (DC) bias field. For magnetic powder core data, permeability is incremental permeability unless otherwise noted. Because of the distributed air gap in powder cores, the initial permeability and incremental permeability, without bias, are essentially the same.

Permeability, Initial:- The limit of incremental permeability as a changing unbiased magnetizing force approaches zero. Because of the distributed gap in powder cores, the initial permeability and incremental permeability without bias are essentially the same.

Permeability, Normal, *μ*:- The ratio of the normal induction to the corresponding magnetizing force.

In the cgs system, the flux density in a vacuum is numerically equal to the magnetizing force and, consequently, the magnetic permeability is numerically equal to the ratio of the flux density to the magnetizing force. Thus:

$\mu = B/H$

In a non-isotropic (anisotropic) medium the permeability is a function of the orientation of the medium, since, in general, the magnetizing force and the magnetic flux are not parallel.

Permeability of Free Space μ_o:- The permeability of a volume occupied by a vacuum. Sometimes called the magnetic constant. Free space permeability is an arbitrary constant used with relative permeability to define the magnetic field (magnetizing force), *H*, and account for the contribution of a magnetic material to total flux density. In the MKSA system, it has a magnitude of $4\pi \times 10^{-7}$ and dimensions of Henries per meter. In the CGS System, free space permeability has a magnitude of 1 and no dimensions. The MKSA free space permeability was chosen so that the practical units for electrical measurements match the ones used for relating magnetic quantities to voltage and current.

Permeability, Recoil:- The ratio of change in flux density as a function of incremental change in applied field (H) in the vicinity of *H*=0. It has no dimensions.

$\mu_o\mu_R = B/H$ in MKS units.

$\mu_r = B/H$ in CGS units.

Permeameter:- An instrument that can measure, and often record, the magnetic characteristics of a specimen.

Permeance, P:- The reciprocal of the reluctance, *R*, measured in weber/At. analogous to conductance in electrical circuits. Indicates the ease with which magnetic flux will follow a certain path, which can be approximated by calculations based purely on magnetic circuit geometry.

Permeance Coefficient, P_c:- Ratio of the magnetic induction, B_d, to its self demagnetizing force, H_d. $P_c = B_d / \mu_o H_d$. This is also known as the 'load line', 'slope of the operating line', or operating point of the magnet, and is useful in estimating the flux output of the magnet in various conditions. As a first order approximation, $B_d / H_d = L_m / L_g$, where L_m is the length of the magnet, and L_g is the length of an air gap that the magnet is subjected to. P_c is therefore a function of the geometry of the magnetic circuit.

Polarity:- The characteristic of a particular pole at a particular location of a permanent magnet. Differentiates the North from the South Pole.

Poles, North and South Magnetic:- The north pole of a magnet, or compass, is attracted toward the north geographic pole of the earth (which is actually, by definition, a magnetic south pole), and the south pole of a magnet is attracted toward the south geographic pole of the earth. The north-seeking pole of a compass or of a magnet is designated by the letter "N", and the other pole by the letter "S". The N (north) pole of the magnet will attract the S (south) pole of another magnet: unlike poles attract.

Pole pieces:- Ferromagnetic materials placed on magnetic poles used to shape and alter the effect of lines of flux.

Polymer bonded magnets:- Magnet powder is mixed with a polymer such as epoxy to form a carrier matrix. The magnets are moulded by compression, extrusion, or injection into a certain shape. Solidification occurs by curing instead of sintering.

Quenching:- A rapid cooling process which follows sintering or solid solutioning.

Rare Earths:- A family of elements in the periodic table having an atomic number from 57 to 71, and including 21 and 39. They are also known as the lanthanide series, which includes lanthanum, cerium, praseodymium, neodymium, samarium, europium, gadolinium, terbium, dysprosium, holmium, erbium, thulium, ytterbium, lutetium, scandium, and yttrium.

Rare-Earth Magnet:− A magnet that has any of the rare-earth elements in its composition; typically stronger than other magnet materials, these include neodymium iron boron and samarium cobalt.

Recoil induction, B_m:- The magnetic induction that remains in a magnetic material after magnetizing and conditioning for final use; measured in Tesla.

Recoil induction B_{rl}, is the magnetic induction that remains in a magnetic material after magnetizing and conditioning for final use; measured in Tesla.

Recoil permeability, μ_R:- $\mu_R = \chi+1$, or permanent permeability, is the average slope of the recoil hysteresis loop. Also known as the minor loop. Of a permanent magnet is defined by a plane passing through the magnet perpendicular to its central flux line at the point of maximum flux.

Relative Permeability:- The ratio of permeability of a medium to that of a vacuum. In the cgs system, the permeability is equal to 1 in a vacuum by definition. The permeability of air is also for all practical purposes equal to 1 in the cgs system.

Rectangularity Ratio:- See squareness ratio.

Reluctance factor, f:- Accounts for the apparent magnetic circuit reluctance. This factor is required due to the treatment of *H*, and *H*, as constants.

Relative Permeability:- The permeability of a material compared with the permeability of free space. This is what normally is specified as material permeability.

Reluctance:- Analogous to electrical resistance, it is the quantity that determines the magnetic flux, ø, resulting from a given magnetomotive force, F.

$R = F / ø$

where: R = reluctance, in At/weber
F = magnetomotive force, in At
$ø$ = flux, in webers

Remnant or *residual*:- The flux density that remains in a magnetic material after an applied magnetic field (magnetizing force) is removed.

Remnant induction, B_d:- Any magnetic induction that remains in a magnetic material after removal of an applied saturating magnetic field, H_s. (B_d is the magnetic induction at any point on the demagnetization curve: measured in Tesla.)

Remanence, B_d:- The magnetic induction that remains in a magnetic circuit after the removal of an applied magnetizing force. If there is an air gap in the circuit, the remanence will be less than the residual induction, B_r.

Residual Flux:- The flux that remains in a core when the applied MMF is returned to a value of zero.

Residual induction (or flux density), B_r:- The magnetic induction corresponding to zero magnetizing force in a magnetic material after saturation in a closed circuit; measured in Tesla. The point at which the hysteresis loop crosses the B axis at zero magnetizing force, and represents the maximum flux output from the given magnet material. By definition, this point occurs at zero air gap, and therefore cannot be seen in practical use of magnet materials.

Return path:- Conduction elements in a magnetic circuit, which provide a low reluctance path for the magnetic flux. Soft magnetic material such as iron or various steels are used to carry or channel the magnetic flux to the gap or working region for interaction with other components. This conductor of magnetic flux is referred to as the return path. It is usually designed to minimize fringing and leakage flux.

Reversible Loss:- A decrease in magnetic induction B of a permanent magnet when subjected to thermal or magnetic demagnetization that is fully recovered (without remagnetisation) when the detrimental conditions are removed (Tesla (T)).

Reversible temperature coefficient:- A measure of the reversible changes in flux caused by temperature variations. These are spontaneously regained when the temperature is returned to its original point. Magnetic saturation of a material exists when an increase in magnetizing force produces no increase in intrinsic induction. The temperature coefficient is a factor which describes the reversible change in a magnetic property with a change in temperature. The magnetic property spontaneously returns when the temperature is cycled to its original point. Most materials exhibit a non-linear response with temperature. It usually is expressed as the percentage change per unit of temperature.

Samarium Cobalt:- A brittle, high energy magnetic material that is best known for its performance at high temperatures. It comes in two compositions: $SmCo_5$ and a higher energy Sm_2Co_{17}.

Saturation - This is the flux density of maximum material magnetization. Magnetization M is the contribution of a magnetic material to the total flux density.

$B = \mu_o (H+M)$ in MKSA units.
$B = H+4\pi M$ in CGS units.

Saturation magnetization is the maximum value of magnetization. Also, the term saturation is sometimes used as a reference to the decrease of permeability with increasing magnetizing force. In an inductor, this corresponds to a decrease of inductance with current.

Saturation:- A condition where the increase in applied external field yields no increase in induction. When this condition is met, all of the elementary magnetic moments have the same alignment. This condition is important in permanent magnet alloys and in Ferromagnetic alloys. Magnet alloys must always be magnetized to saturation. The magnet may not be used at this level, but before conditioning and stabilization the magnet must always first be magnetized to saturation. Usually saturation should not be exceeded in Ferromagnetic alloys which comprise the yoke or return path elements of a magnetic circuit. If ferromagnetic elements are saturated there will be flux leakage in the system and a redesign should be considered.
The condition under which all elementary magnetic moments have become oriented in one direction. A ferromagnetic material is saturated when an increase in the applied magnetizing force produces no increase in induction. Saturation flux densities for steels are in the range of 1.6 to 2 Tesla.

Saturation Flux Density, B_{sat}:- The flux density at which a material saturates.

search coil:- A coiled conductor, usually of known area and number of turns, that is used with a flux meter to measure the change of flux linkage with the coil.

Saturation intrinsic induction, B_{is}, (or J):- The maximum intrinsic induction possible in a material.

Second quadrant curve:- The second quadrant curve is the demagnetization portion of the hysteresis loop created with a permeameter. In a permeameter, magnets are magnetized to saturation in the first quadrant and then demagnetized to plot the second quadrant curve. The second quadrant curve is the intrinsic curve starting at B_r and ending at H_{ci}. From this intrinsic curve, the extrinsic (normal) curve is calculated to derive the line which extends from B_r to H_c.

Self Demagnetizing Field:- A field inside a permanent magnet that is opposed to its own magnetization, which is due to the internal coupling of its poles following the introduction of an air gap in the magnetic circuit (Tesla (T)).

Sintered:- A sintered magnet is comprised of a compacted powder which is then subjected to a heat treat operation where the full density and magnetic orientation is achieved. Sintering occurs at elevated temperatures, typically between 1100 and 1200°C.

Sintered Iron:- Powdered iron that has been pressed and sintered into a structural form. This type of material occasionally is used in a magnetic
application, but they normally exhibit excessive core losses.

Sintered Magnets:- Magnets that are made from powdered materials that are pressed together, and then heated in an oven to produce desired shapes and magnetic properties.

Skewing Of The Loop:- When air gap is added to the magnetic path, the hysteresis loop is made to lean over (permeability is reduced); it is said to be skewed or sheared.

Soft Magnetic Material:- Shaped piece of ferromagnetic material that once having been magnetized is easily demagnetized, i.e. requires a slight coercive force to remove the resultant magnetism. A material with low coercivity and high permeability. Generally accepted as having a coercivity of less than 24 kA/m though most soft materials used in inductors have coercivity of under 0.8 kA/m.

Square Loop:- Refers to a hysteresis loop where the difference between B_m and B_r of a material is quite small, resulting in a rectangular appearance of the intrinsic curve.

Stabilization:- Exposure of a magnet to demagnetizing influences expected to be encountered in use in order to prevent irreversible losses during actual operation. Demagnetizing influences can be caused by high or low temperatures, or by external magnetic fields.

Temperature Coefficient of B_r:- A factor, which describes the reversible change in a magnetic property with change in temperature. Expressed as percent change per unit of temperature.
The magnetic property spontaneously returns when the temperature is cycled to its original point so long as a limit condition is not exceeded – see note below. It usually is expressed as the percentage change per unit of temperature over a specified temperature range.
Above (or below) a critical temperature, dependent upon the material and its magnetic characteristics and magnetic circuit, an irreversible loss may take place which is recovered when the magnet is re-saturated.

Temperature Stabilization:- After manufacture, many types of hard and soft magnetic materials can be thermally cycled to make them less sensitive to subsequent temperature extremes.

Tesla:- MKSA (SI) unit for magnetic flux density, defined by Faraday's Law. A Tesla represents a volt-second per square meter per turn. One Tesla is equal to one Weber per square metre. One Tesla equals 10,000 Gauss

T_{max}, T_m, or Maximum service temperature:- The maximum temperature to which the magnet may be exposed with no significant long-range instability or structural changes. A proposed magnetic definition is that the normal hysteresis curve is a straight line in the second quadrant up to the T_{max} temperature; the line begins to show curvature (a 'knee') once T_{max} is exceeded.

Unoriented (isotropic) material:- Material with equal magnetic properties in all directions.

Weber:- The practical unit of magnetic flux. It is the amount of magnetic flux which, when linked at a uniform rate with a single-turn electric current during an interval of 1 second, will induce in this circuit an electromotive of force of 1 volt. 1 Weber = 10^8 Maxwells.

Bibliography

The following books are recommended for reading and reference purposes.

A

Acha, E. (Editor)
Power electronic control in electrical systems, Newnes, 2002.. p. 443.

Agrawal, J. P.
Power electronic systems: theory and design, Prentice Hall, 2001, p. 562.

Ahmed, A.
Power Electronics for Technology, Prentice Hall, 1999, p. 427.

Akagi, H.
Instantaneous power theory and applications to power conditioning, Wiley, 2007. p. 379.

Arrilliaga, J. *et al.*
Power System Harmonics, Wiley, 1985. p. 336.

Ang, S. S.
Power-Switching Converters, Dekker, 1995, p. 412.

Antognetti, P. *et al.*
Semiconductor Device Modeling with SPICE, McGraw-Hill, 1988, p. 389.

Asghar, M. S. J.
Power Electronics, Prentice Hall, 2004, p. 500.

Atkinson, P.
Thyristors and their Applications, Mills and Boon, 1972, p. 128.

B

Baliga, B. J.
Modern Power Devices, Krieger, 1992, p. 476.

Baliga, B. J.
Power Semiconductor Devices, PWS Publishing, 1996, p. 624.

Barnes, M.
Practical Variable Speed Drives and Power Electronics, 2003, p. 304.

Barton, T. H.
Rectifiers, Cycloconverters, and AC Controllers, Clarendon Press, 1994, p. 687.

Basso, C. P.
Switch-mode power supply SPICE cookbook, McGraw-Hill, 2001, p. 263.

Bar-Lev, A.
Semiconductors and Electronic Devices, Prentice-Hall, 1979.

Batarseh, I.
Power Electronic Circuits, Wiley, 2004, p. 574.

Bausiere, R. *et al.*
Power Electronic Converters: DC-DC Conversion, Springer-Verlag, 1992, p. 402.

Bedford, B. D. *et al.*
Principles of Inverter Circuits, Wiley, 1964, p. 430.

Benda, V. *et al.*
Power Semiconductor Devices: theory and applications, Wiley, 1999, p. 419.

Benysek. G.
Improvement in the Quality of Delivery of Electrical Energy using Power, Springer, 2007, p. 191.

Bergveld, H. J. (editor), *et al*.
Battery management systems: design by modeling, Kluwer Academic, 2002. p. 295.

Berutti, A..*et al*.
Practical Guide to Quality Power for Sensitive Electronic Equipment, Intertec Electrical Group, 1993, p. 100.

Billings, K. H.
Switchmode Power Supply Handbook, 2nd Ed., McGraw-Hill, 1999, p. 450.

Bird, B. M. *et al.*
An Introduction to Power Electronics, 2nd Ed., Wiley, 1993, pp. 374.

Bimbhra. P. S.
Power Electronics, Khanna Publishers, 2004.

Blaabjerg, F. et al.
Power Electronics for Modern Wind Turbines, Morgan and Claypool, 2006, pp. 60.

Blicher, A.
Thyristor Physics, Springer-Verlag, 1976.

Boldea, I. *et al.*
Electric Drives, CRC Press, 1999, pp. 411.

Bose, B. K.
Power Electronics and AC Drives, Prentice-Hall, 1986, pp. 402.

Bose, B. K. (Editor)
Microcomputer Control of Power Electronics and Drives, IEEE, 1987, p. 465.

Bose, B. K. (Editor)
Power Electronics and Variable Frequency Drives: Technology and Applications, Wiley, 2001.

Bose, B. K.
Modern power electronics and AC drives, Prentice Hall, 2002, p. 711.

Van den Bossche, A.
Inductors and Transformers for Power Electronics, CRS, 2005, p. 480.

Bradley, D. A.
Power Electronics, 2nd Ed., Chapman and Hall, 1995.

Brichant, F.
Forced Commutated Inverters, MacMillan, 1984.

Brown, M.
Practical Switching Power Supply Design, Academic Press, 1990, p. 240.

Brown, M.
Power Supply Cookbook, 2nd Ed., Newnes, 2001, p. 265.

C

Carr, J. J.
DC Power Supplies: A Technician's Guide, TAB Books, 1996.

Carrow, Robert S.
Electronic Drives, New York : TAB Books, c 1996. ix, 326 p.

Chetty, P. R. K.
Switch-Mode Power Supply Design Handbook, TAB Books, 1986, p. 179.

Cheron, Y.
Soft Commutation, Chapman and Hall, 1992, p. 233.

Chrysiss, G.
High Frequency Switching Power Supplies: Theory and Design, 2nd Ed., McGraw-Hill, 1989, p. 287.

Clark, J. W.
AC Power Conditioners: design and applications, Academic Press, 1990, p. 208.

Crandall, E.
Power Supply Testing Handbook: Strategic Approaches in Test Cost Reduction, Chapman and Hall, 1997, p. 320.

Crompton, T. R.
Battery Reference Book, 2nd Ed., Butterworth Heinemann, 1995, p. 752.

Crowder, R. M.
Electric drives and electromechanical systems, Butterworth-Heinemann, 2006.p. 292.

Csaki, F. *et al.*
Power Electronics, Akademiai Kiado, 1975, p. 708.

Csaki, F. *et al.*
Power Electronics, Problem Manual, Akademiai Kiado, 1979, pp. 474.

Cuk, S, *et al.*
Advances in Switched-Mode Power Conversion, Volume III, TESLAco, 1983. p. 353.

D

Datta, S. K.
Power Electronics and Controls, Reston, 1985, p. 196.

Davis, B.
Understanding DC Power Supplies, Prentice-Hall, 1983, p. 232.

Davis, R. M.
Power Diode and Thyristor Circuits, Cambridge University Press, 1971, p. 265.

De, G.
Converter Engineering, An Introduction to Operations and Theory, Wiley, 1975.

De, G.
Principles of Thyristorised Converters, Oxford and IBH Publishing Co., 1982.

Dewan, S. B. *et al.*
Power Semiconductor Drives, Wiley, 1984, p. 354.

Dote, Y. *et al.*
Intelligent Control: power electronic systems, Oxford University Press, 1998, p. 209.

Dubey, G. K.
Power Semiconductor Controlled Drives, Prentice-Hall, 1989.

Dubey, G. K. *et al.*
Thyristorised Power Controllers, Wiley, 1986, p. 825.

E

Eckert, J. K. (editor),
Filters and Power Conditioning, Interference Control Technologies, 1988.

Ehsani, M. (Editor)
Modern electric, hybrid electric, and fuel cell vehicles : fundamentals, theory, and design, CRC Press, 2005. p. 395.

Eichenauer, C. J.
Power Electronics and RF Power Systems Analysis - Program Examples in BASIC and C, Prentice Hall, 1993, p. 429.

El-Hawary, M. E.
Principles of Electric Machines with Power Electronic Applications, 2nd Ed., Wiley, 2002, p. 496.

El-Sharkawi, M. A.
Electric energy : an introduction, 2nd ed., CRC Press, 2009. p. 470.

Emadi, A.
Handbook of Automotive Power Electronics and Motor Drives, CPC, 2005, p. 704.

Erickson, R. W. *et al.*
Fundamentals of Power Electronics, 2nd Ed.; Kluwer Academic, 2001, p. 883.

F

Ferenczi, O.
Power Supplies:
Part A: Linear Power Supplies, DC-DC Converters;
Part B: Switched-Mode Power Supplies, Elsevier Science, 1987, p. 744.

Ferreira, J.A.
Electromagnetic Modeling of Power Electronic Converters, Kluwer, 1989, p. 173.

Fewson, D.
Introduction to Power Electronics, Oxford University Press, 1998, p. 192.

Finney, D.
The Power Thyristor and its Applications, McGraw-Hill, 1980, p. 57.

Fishenden, M. *et al.*
An Introduction to Heat Transfer, Oxford University Press, 1982.

Fisher, M. J.
Power Electronics, PWS-Kent, 1991, p. 491.

Flanagan, W. M.
Handbook of Transformer Design and Application, McGraw-Hill, 1993.

Fluke, J. C.
Controlling Conducted Emissions by Design, Van Nostrand-Reinhold, 1991, p. 334.

Fransua, A., and R. Magureanu,
Variable Frequency AC Motor Drive Systems, Peter Peregrinus Ltd., 1988.

G

Gates Energy Products,
Rechargeable Batteries Applications Handbook, Butterworth Heinemann, 1992.

General Electric
General Electric SCR Manual, 6th Ed. 1979

General Electric Industrial Controls
The Power Thyristor and its Applications, McGraw-Hill, 1980.

Gentry, F. W. *et al.*
Semiconductor Controlled Rectifiers, Prentice-Hall, 1964.

Geyger, W. A.
Magnetic-Amplifier Circuits, Basic Principles, Characteristics, and Applications, 2nd Ed, McGraw-Hill, 1957.

Ghandi, S. K.
Semiconductor Power Devices, Wiley, 1977.

Gingrich, H. W.
Electrical Machinery, Transformers, and Control, Prentice-Hall, 1979.

Goldman, A.
Magnetic components for power electronics, Kluwer, 2002, p. 286.

Goldman, A.
Handbook of modern ferromagnetic materials, Kluwer, 1982, p. 646.

Gottlieb, I. M.
Principles and Applications of Inverters and Converters, H. W. Sams, 1977, p. 192.

Gottlieb, Irving M.
Solid State Power Electronics, 1st ed., H. W. Sams, 1979, p. 296.

Gottlieb, I. M.
Practical Power-Control Techniques, H.W. Sams, 1987, p. 248.

Gottlieb, I. M.
Power Control with Solid State Devices, TAB Books, 1987, p. 372.

Gottlieb, I. M.
Design and Build Electronic Power Supplies, TAB Books, 1991, p. 163.

Gottlieb, I. M.
Regulated Power Supplies, 4th Ed, TAB Books, 1992, p. 460.

Gottlieb, I. M. *et al.*
Electronic Power Control, Macmillan/McGraw-Hill, 1993, p. 265.

Gottlieb, I. M.
Power Supplies, Switching Regulators, Inverters, and Converters, 2nd Ed., TAB Books, 1994, p. 479.

Graf, R. F.
The Modern Power Supply and Battery Charger Circuit Encyclopedia, TAB Books, 1992, p. 133.

Graf, R. F.
Converter and Filter Circuits, Newnes, 1997, p. 184.

Grant, D. A. *et al.*
Power MOSFETS - Theory and Application, Wiley, 1989, p. 504.

Grossner, N. R.
Transformers for Electronic Circuits, 2nd Ed, McGraw-Hill, 1983, p. 400.

Grover, F. W.
Inductance Calculations Working Formulas and Tables, Dover, 1946, p. 286.

Gurevich, V.
Electronic devices on discrete components for industrial and power engineering, CRC Press, 2008. p. 419.

Gwyther, H. F. G.
Solving Problems in Electrical Power and Power Electronics, Wiley, 1988, p. 203.

Gyugyi, L. G. *et al.*
Static Power Frequency Changers, Theory, Performance, and Applications, Wiley, 1976, p. 442.

H

Hart, D. W.
Introduction to Power Electronics, Prentice Hall, 1997, pp. 418.

Hempel, H. P.
Power Semiconductor Handbook, Semicron International, 1980.

Heumann, K.
Basic Principles of Power Electronics, Springer-Verlag, 1986, pp. 294.

Hnatek, E. R.
Design of Solid-State Power Supplies, 3rd Ed., Van Nostrand-Reinhold, 1989, pp. 639.

Hoffman, A. *et al.*
Thyristor-Handbook, Siemens, 1965.
Hoft, R. G. *et al.*
Semiconductor Power Electronics, Krieger, 1991, pp. 384.
Holmes, D. G. *et al.*
Pulse width modulation for power converters: principles and practice, Wiley, 2003, pp. 724.
Hughes, A.
Electric motors and drives: fundamentals, types, and applications, 2nd ed., Newnes, 1993. p. 339.

J

Jain, A.
Power Electronics and Its Applications, Penran.
Jacob, J. M.
Power Electronics: principles & applications, Delmar Thomson Learning, 2002, pp. 525.
Jamasb, T. *et al*. (Editors)
Future electricity technologies and systems, Cambridge University Press, 2006. p. 426.

K

Karady, G.
High-Power Electronic Devices in Advances in Electronics and Electron Physics, Vol. 41, Academic Press, 1976.
Kassakian, J. G. *et al.*
Principles of Power Electronics, Addison-Wesley, 1991, p. 738.
Kazimierczuk, M. K. *et al.*
Resonant Power Converters, Wiley, 1995, p. 481.
Kazmierkowski, M. P. *et al.*
Automatic Control of Converter-fed Drives, Elsevier, 1994, p. 559.
Kazmierkowski, M. P. (Editor)
Control in power electronics: selected problems, Academic Press, 2002, p. 518.
Kenjo, T.
Power Electronics for the Microprocessor Age, Oxford University Press, 1990, p. 349.
Kilgenstein, O.
Switched-Mode Power Supplies in Practice, Wiley, 1996, p. 338.
Kiltie, O.,
Design Shortcuts and Procedures for Electronic Power Transformers and Inductors, Printed by the author, 1981. p. 274.
Kimbark, E. W.
HVDC Transmission, Wiley, 1965.
Kislovski, A. S.
Introduction to Dynamic Analysis of Switching DC-DC Converters, BWV Engineering, Switzerland, 1985, p. 400.
Kislovski, A. S. *et al.*
Dynamic Analysis of Switching-Mode DC/DC Converters, Van Nostrand Reinhold, 1991, p. 404.
Kloss, A.
A Basic Guide to Power Electronics, Wiley, 1984, p. 227.
Krause, P. C. *et al.*
A*nalysis of electric machinery and drive systems*, 2nd ed., Wiley-IEEE Press, 2002. p. 613.
Krein, P. T.
Elements of Power Electronics, Oxford University Press, 1998, p. 766.
Kubat, M.
Power Semiconductors, Springer-Verlag, 1984.
Kularatna, N.
Power Electronics Design Handbook, Newnes, 1998, p. 300.
Kularatna, N.
Modern component families and circuit block design, Newnes, 2000, p. 452.
Kusko, A.
Solid State D.C. Motor Drives, MIT Press, 1968.
Kusko, A.
Computer Aided Design of Magnetic Circuits, MIT Press, 1969.
Kusko, A.
Emergency-Standby Power Systems, McGraw-Hill, 1988.

Kustom, R. L.
Thyristor Networks for the Transfer of Energy between Superconducting Coils, University of Wisconsin Press, 1980, p. 116.

L

Lander, C. W.
Power Electronics, 3rd Ed., McGraw-Hill, 1993, pp. 480.
Larson, B.
Power Control Electronics, Prentice-Hall, 1983, pp. 164.
Lazar, J.
Park Vector Theory of Line-Commutated Three-Phase Bridge Converters, Omikk, 1987.
Lee, Fred C. (Editor),
Power Electronics Technology and Applications II, IEEE, 1997.
Lee, R. *et al.*
Transformers and Circuits, 3rd Ed., Wiley, 1988, pp. 480.
Lee, Y. S.
Computer-Aided Analysis and Design of Switch-Mode Power Supplies, Dekker, 1993, pp. 522.
Lenk, J. D.
Complete Guide to Electronic Power Supplies, Prentice Hall, 1990, pp. 272.
Lenk, J. D.
Simplified Design of Linear Power Supplies, Butterworth-Heinemann, 1995, pp. 224.
Lenk, R.
Practical Design of Power Supplies, McGraw-Hill, 1998, pp. 267.
Leonhard, W.,
Control of Electrical Drives, Springer-Verlag, 1985.
Lilienstein, F. M.
Magnetics Engineering Fundamentals and Computer-Aided Design, Van Nostrand and Reinhold, 1993, p. 468.
Lines, D. *et al.*
Building Power Supplies, 2nd Ed., Master, 1997, p. 124.
Lowther, D. A. *et al.*
Computer-Aided Design of Magnetics, Springer-Verlag, 1986.
Lin F. L. *et al.*
Advanced DC/DC Converters, CRC, 2003, p. 792.
Luo, F. L. et al.
Digital Power Electronics and Applications, 2005, p. 464.
Lyshevski, S. E.
Electromechanical systems and devices, CRC Press, 2008. p. 565.
Lyshevski, S. E.
Electromechanical systems, electric machines, and applied mechatronics,: CRC Press, 2000. p. 782.

M

Mardiguian, M.
Controlling Radiated Emissions by Design, 2nd Ed., Kluwer, 2001, p. 338.
Mardiguian, M.
EMI troubleshooting techniques, McGraw-Hill, 2000, p. 301.
Marston, R. M.
110 Thyristor Project Using SCRs and Triacs, Iliffe, 1972, p. 138.
Marston, R. M.
Power Control Circuits Manual, 2nd Ed., Newnes, 1997, p. 220.
Martinelli, R. M. (Editor),
Smart Power for Power Conversion, Intertec Communications, 1988, p. 384.
Masar, S. A.,
Electric Machines and Power Systems, McGraw-Hill, 1995.
Massobrio, G.
Semiconductor device modeling with SPICE, 2nd Ed., McGraw-Hill, 1993, p. 479.
Mazda, F. F.
Thyristor Control, Newnes-Butterworth, 1973.
Mazda, F. F.
Power Electronics Handbook: Components, Circuits, and Applications, Butterworth, 1990, p. 417.
Mazda, F. F.
Power Electronics Handbook, 3rd Ed., Newnes, 1997, p. 441.

McLyman, W. T.
Magnetic core selection for transformers and inductors: a user's guide to practice and specification, 2nd Ed., Dekker, 1997, p. 630.
McLyman, W. T.
Designing Magnetic Components for High Frequency DC-DC Converters, KG Magnetics, 1993, p. 433.
McLyman, W. T.
High reliability magnetic devices: design and fabrication, Dekker, 2002.
McMurray, W.
The Theory and Design of Cycloconverters, MIT Press, 1972.
Middlebrook, R. D. et al.
Advances in Switched-mode Power Conversion, 2nd Ed, TESLAco, 1983. p. 533.
Miller, J. M.
Propulsion systems for hybrid vehicles, Published/Created: Stevenage, UK : Institution of Electrical Engineers, c2004, xvi, 455 p.
Miller, R.
Industrial electricity & motor controls, McGraw-Hill, c2008. xiii, 448 p.
Mitchell, D. M.
DC-DC Switching Regulator Analysis, McGraw-Hill, 1988, pp. 163.
Mohan, N. et al.
Power Electronics: converters, applications, and design, 4th Ed., Wiley, 2007, pp. 824.
Moltgen, G.
Converter Engineering: An Introduction to Operation and Theory, Wiley, 1984, pp. 165.
Moncrief, W. A.
Single-phase to Three-phase Electric Power Converters : A Concise Application Guide, The Association, 1996.
Montrose, M. I.
Printed circuit board design techniques for EMC compliance: a handbook for designers, 2nd Ed., IEEE, 2000, pp. 307.
Moorthi, V. R.
Power Electronics: Devices, Circuits and Industrial Applications, Oxford University Press, 2005, p. 1028.
Morrison, R.,
Grounding and Shielding Techniques in Instrumentation, 3rd Edition, Wiley, 1986. p. 172.
Motto, J. W., Jr. (Editor)
Introduction to Solid State Power Electronics, Westinghouse Electric Corporation, Semiconductor Division, 1977, pp. 143.
Mullard Ltd.
Power Engineering Using Thyristors; Vol. 1, Techniques of Thyristor Power Control, 1970.
Murphy, J. M. D.
Thyristor Control of AC Motors, Pergamon, 1973.
Murphy, J. M. D. *et al.*
Power Electronic Control of AC Motors, Pergamon, 1988, pp. 524.

N
Nasar, S. A.
Electric Machines and Power Systems, McGraw-Hill, 1995.
Nave, M. J.
Power Line Filter Design for Switched-Mode Power Supplies, Van Nostrand Reinhold, 1991, pp. 210.
Neiman, L. R., et al.
DC Transmissions in Power Systems, Israel program for Scientific Translations, 1967.
Newell, W. E. *et al.*
Introduction to Solid State Power Electronics, Westinghaouse Electric Corp., 1977.
Newman, M.
Industrial Electronics and Controls, Wiley, 1986.
Norris, Bryan (editor),
Microprocessors and Microcomputers and Switching Mode Power Supplies: McGraw-Hill, 1978. p. 216.
Novotny, D.W. *et al*.
Vector Control and Dynamics of AC Drives, Oxford University Press, 1996. p. 440
Nowicki, J. R.
Power Supplies for Electronic Equipment, Vols. I and 2, CRC, 1971.

O
Ohno, E. *et al.*
Introduction to Power Electronics, Clarendon, 1988, p. 304.
Ohshima, R. *et a*l. (editors),
Servo Sensors (Elements and Applications), Intertec Communications, 1988. p. 331.
Ott, H. W.
Noise Reduction Techniques in Electronic Systems, 2nd Ed., Wiley, 1988, p. 426.
Oxner, E. S.
Power FETs and Their Applications, Prentice-Hall, 1982.
Ozenbaugh, R. L.
EMI Filter Design, 2nd Ed., Dekker, 2001, p. 319.

P
Paice, D. A.
Power Electronics Converter Harmonics: Multipulse Methods for Clean Power, Wiley, 2001, p. 222.
Pearman, R. A.
Power Electronics: Solid State Motor Control, Reston, 1980.
Pearman, R. A.
Solid-State Industrial Electronics, Reston, Inc. 1984.
Pearman, R. A.
Electric Machinery and Transformer Technology, Saunders College Pub., 1994, pp. 636.
Pejovic, P.
Three-phase diode rectifiers with low harmonics : current injection methods, Springer, 2007. p. 318.

Pelly, B. R.
T*hyristor Phase Controlled Converters and Cycloconverters*, Wiley, 1971.
Phillips, A. B.
Transistor Engineering, McGraw-Hill, 1962.
Pierre, Edward R.,
Welding Processes and Power Sources, 1st ed, Power Publications Co. 1968. p. 263.
Pillai, S. K.,
A First Course on Electrical Drives, Wiley/Hasted Press, 1982.
Platt, S.
Magnetic Amplifiers, Theory and Applications, Prentice-Hall, 1958.
Podoski, J.
Power Conditioning Devices; Analytical Survey, Library of Congress Aerospace Technology Division, Washington, 1966, pp. 158.
Pressman, A. I.
Switching and Linear Power Supply, Power Converter Design, Hayden, 1977, pp. 372.
Pressman, A. I.
Switching Power Supply Design, 2nd Ed., McGraw-Hill, 1998, pp. 682.

Q
Querciolo, V.
Pulse Width Modulated Power Supplies, Elsevier Science, 1993.

R
Rajagopalan, V.
Computer-Aided Analyses of Power Electronics Systems, Dekker, 1987, pp. 541.
Ramamoorty, M.
Introduction to Thyristors and Their Application, McMillan, 1978.
Ramírez, H. S. R. *et al*.
Control Design Techniques in Power Electronics Devices, 2006, pp. 423.
Ramshaw, R. S.
Power Electronics, Chapman & Hall, London, 1975.
Ramshaw, R. S.
Power Electronics Semiconductor Switches, 2nd Ed., Klewer, 1993, pp. 458.
Ramshaw, R. *et al*.
PSpice simulation of power electronic circuits: an introductory guide, Kluwer, 1996, pp. 400.
Rashid, Muhammad H. (Editor)
Power electronics handbook, Academic Press, 2001, pp. 895.

Rashid, M. H.
Power Electronics, Circuits, Devices and Applications, 3rd Ed., Pearson, 2004, pp. 912.
Rashid, M. H.
SPICE for Power Electronics and Electric Power, Prentice Hall, 1993, pp. 394.
RCA
Solid State Power Circuits, RCA Designers Guide Technical Series SP-52, 1971.
Reddy, R. S.
Fundamentals of Power Electronics, CRC, 2000, pp. 190.
Rissik, H.
Mercury Arc Current Converters, Pitman, 1963.
Rombaut, C. *et al*.
Power Electronic Converters - AC/AC Conversion, McGraw-Hill, 1987, pp. 340.
Ross, J. N.
The Essence of Power Electronics, Prentice Hall, 1997, pp. 217.
Rossetti, N.
Managing Power Electronics: VLSI and DSP-Driven Computer Systems, Wiley, 2005, pp. 383.

S

Salon, S. *et al*.
Power Electronics & Motor Drives, IEEE, 2000
Sandler, S. M.
SMPS simulation with SPICE 3, McGraw-Hill, 1997, p. 187.
Sands, Leo G.
Batteries and Electronic Power Supplies, 2nd Ed., Scranton, International Correspondence Schools, 1968.
Sarjeant, W. *et al*.
High-Power Electronics, TAB, 1989, p. 392.
Saunders, C. W.
Power Electronics, McGraw-Hill, 1981.
Schaeffer, J.
Rectifier Circuits, Wiley, 1965.
Schwartz, Francisc. C.,
Power Processing, NASA SP-244, NASA Electronics Research Center, 1971.
Schwarz, Francisc C.,
Power Electronics, Delft University Press, 1974 p.16.
Seguier, G.
Power Electronic Converters - AC/DC Conversion, McGraw-Hill, 1986, p. 372.
Seguier, G.
Power Electronic Converters : DC-AC Conversion, Springer-Verlag, 1993, p. 444.
Sen, P. C.
Thyristor DC Drives, Wiley, 1981. Krieger, 1991, p. 307.
Sen, P. C.
Principles of Electric Machines and Power Electronics, 2nd Ed., Wiley, 1997, p. 615.
Severns, R. *et al.* (Editors)
MOSPOWER Applications, Siliconix, 1984.
Severns, R. P. *et al*.
Modern DC-DC Switchmode Power Conversion Circuits, Van Nostrand Reinhold, 1985, p. 342.
Shaffer, R.
Fundamentals of Power Electronics With Matlab, Charles River Media, 2006, p. 384.
Sholes, G. J.
Handbook of Rectifier Circuits, Ellis Horwood, 1980.
Shenkman, Arieh L
Transient analysis of electric power circuits handbook, Springer, 2005., p. 569.
Shepard, J. D.
Power Supplies, Reston Pub. Co., 1984, p. 178.
Shepard, W.
Thyristor Control of AC Circuits, Crosby Lockwood Staples, 1975.
Shepard, W.
Energy Flow and Power Factor in Nonsinusoidal Circuits, Cambridge University Press, 1979.
Shepard, W. *et al*.
Power Electronics and Motor Control, 2nd Ed., Cambridge University Press, 1995, p. 539.
Simões, M. G.
Alternative energy systems : design and analysis with induction generators, CRC Press, 2008. p. 433.

Singh, R. *et al.*
Cryogenic operation of silicon power devices, Kluwer, 1998, pp. 148.
Sira R. *et al*.
Control design techniques in power electronics devices, Springer, 2006. p. 423.
Sittig, R. *et al.*
Semiconductor Devices for Power Conditioning, Plenum, 1982.
Skinner, A. J.
Four Quadrant Inverter Technologies for High Frequency UPS, ERA Technology, 1992, p. 34.
Skvarenina, T. L. (Editor)
The power electronics handbook, CRC, 2002, p. 664.
Slonim, M. A.
Theory of Static Converter Systems - Part A: Steady-State Processes, Elsevier, 1984.
Slonim, M. A.,
Theory of Static Converter Systems : Mathematical Analysis and Interpretation, Elsevier, 1984.
Smith, D. C.
High Frequency Measurements and Noise in Electronics Circuits, Van Nostrand and Reinhold, 1993, p. 231.
Smith, S.
Magnetic Components - Design and Application, Van Nostrand and Reinhold, 1984.
Snelling, E. C.
Soft Ferrites, Iliffe, 1969.
Snelling, E. C. *et al.*
Ferrites for Inductors and Transformers, Research Studies Press, 1983.
Steinberg, D. S.
Cooling techniques for Electronic Equipment, 2nd Ed., Wiley, 1991, p. 483.
Stevens, R. E.
Electrical Machines and Power Electronics, Van Nostrand and Reinhold, 1983.
Storm, H. F.
Magnetic Amplifiers, Wiley, 1955.
Strzelecki, Ryszard. Benysek, Grzegorz.(Editors)
Power electronics in smart electrical energy networks, Springer, 2008., p. 414.
Subrahmanyam, V.
Power Electronics, Wiley, 2006. p. 863.
Sueker, K. H.
Power Electronics Design: A Practitioner's Guide, Elsevier, 2005, p. 272.
Sugandhi, R. K. *et al*.
Thyristors - Theory and Application, Wiley, 1981.
Sum, K. K.
Switch Mode Power Conversion, Basic Theory and Design, Dekker, 1984, p. 324.
Sze, S. M.
Semiconductor Devices and Technology, Wiley, 1985.

T

Takeuchi, T. J.
Theory of SCR Circuits and Applications to Motor Control, Tokyo Electrical Engineering College Press, 1968.
Taraseiskey, H.
Power Hybrid Circuit Design and Manufacture, Dekker, NY, 1996, p. 344.
Tarter, R. E.
Principles of Solid-State Power Conversion, H.W. Sams, 1985, p. 592.
Tarter, R. E.
Solid-State Power Conversion Handbook, Wiley, 1993, p. 719.
Taylor, P. D.
Thyristor Design and Realization, Wiley, 1987.
Thollot, P. A. (editor),
Power Electronics Technology and Applications, IEEE, 1992, p. 429.
Thomson-CSF Semiconductor Div.,
The Power Transistor in its Environment, 1978.
Thorborg, K.
Power Electronics, Chatwell-Bratt, Lund, 1993, p. 504.
Tihanyi, L.
Electromagnetic Compatibility in Power Electronics, Butterworth Heinemann, 1995, p. 403.
Toliyat, H. et al.
DSP-Based electromechanical motion control,: CRC Press, 2003. p. 360.

Towers, T. D.
Practical Solid State DC Supplies, G/L Tab Books, 1977, p. 192.
Traister, R. J.
44 Power Supplies for Your Electronic Projects, TAB, 1987, p. 244.
Traister, R.
Voltage Regulation Circuit Manual, Academic Press, 1989, p. 152.
Trzynadlowski, A. M.
Introduction to Modern Power Electronics, Wiley, 1998, p. 433.
Trzynadlowski, A. M.
The Field Orientation Principle in Control of Induction Motors, Kluwer Academic Press, 1993. p. 255.
Tse, C. K.
Complex behavior of switching power converters, CRC, 2004, p. 262.
Tuinenga, P. W.
SPICE : A Guide to Circuit Simulation and Analysis Using PSPICE, 3rd Ed., Prentice Hall, 1995, p. 288.
Takuechi, T. J. *Theory of SCR Circuits and Application to Motor Control,* Tokyo Electrical Engineering College Press, 1968.

U
Uhlman, E.
Power Transmission by Direct Current, Springer-Verlag, 1975.

V
Valentine, R. (Editor)
Motor control electronics handbook, McGraw-Hill, 1998. p. 704.
Vance, E. F.,
Coupling to Shielded Cables, Wiley, reprinted by Robert E. Krieger Publishing Co. 1987. p. 194.
Vithayathil, J.
Power Electronics: Principles and Applications, McGraw-Hill, 1995, p. 632.
Vorpérian, V.
Fast analytical techniques for electrical and electronic circuits, Cambridge University Press, 2002. p. 476.

W
Waggoner, R. M. (editor)
Practical Guide to Quality Power for Sensitive Electronic Equipment, 2nd Ed., EC&M Books, 1997, p. 144.
Walker, N. E.
The Design Analysis Handbook - A Practical Guide to Design Validation, DBS, 1994, p. 245.
Wang, Z.
Current-*mode Analog Integrated Circuits and Linearization Techniques in CMOS Technology*, Hartung-Gorre, 1990, p. 235.
Wells, R.
Solid-State Power Rectifiers, Granada.
Westinghouse Electric Corp.,
Westinghouse Thyristor Handbook.
Whitaker, Jerry C.
P*ower vacuum tubes handbook*, 2nd ed., CRC Press, 1999. p. 710.
Whittington, H. W. *et al.*
Switched-Mode Power Supplies - Design and Construction, 2nd Ed., Wiley, 997, p. 236.
Williams, T.
EMC for Product Designers, 3rd Ed., Newnes, 2001. p. 360.
Wood, P.
Switching Power Converters, Van Nostrand Reinhold, 1981, p. 446.
Wu, K. C.
Pulse width modulated DC/DC converters, Kluwer, 1997, p. 234.

Physical Constants

Angstrom	Å	10^{-10} m = 10^{-1} nm = 10^{-4} µm		
Avogadro's number	N	6.022×10^{23}	atom per mole	/mol
Bohr radius	a_o	0.529177	Angstrom	
		52.9177		pm
Boltzmann's constant	k	1.38×10^{-23}	Joule per Kelvin	J/K
Electronic charge, eV	q	1.602×10^{-19}	Coulomb	C
Free electron rest mass	m_e	9.11×10^{-31}	kilogram	kg
Acceleration – gravity	g	9.80665		m/s^2
Permeability of free space	μ_o	$4\pi\times10^{-7}$	Henry per metre	H/m
Permittivity of free space	ε_o	8.854×10^{-12}	Farad per metre	$1/\mu_o c^2$
Planck's constant	h	6.626×10^{-34}	Joule second	J s
Proton rest mass	M_p	1.67×10^{-27}	kilogram	kg
Speed of light in vacuum	c	2.998×10^{8}	metre per second	m/s
Standard Atmospheric Pressure		1.01325×10^5 Pa		Pa or N/m^2
Stefan-Boltzmann constant	σ	5.671		$W/(m^2K^4)$
Thermal voltage @ 300K	V_t	0.02586	kT/q	V
Wavelength of 1 eV quantum	λ	1.23977 µm		µm

Silicon Material parameters

Bandgap @ 300K	E_g	1.12		eV
Breakdown field	E_{br}	3×10^7		V/m
Density	ρ	2.33		g/cm^3
Intrinsic concentration @ 300K	n_i	1.0×10^{10}		cm^{-3}
Electron affinity	X	4.05		V
Mobility @ 300K	μ_n	1400	electrons	$cm^2/V\text{-}s$
	μ_p	450	holes	
Relative dielectric constant	$\varepsilon_s/\varepsilon_o$	11.9		
Thermal conductivity @ 300K	χ	1.5		W/cmK

Metal	resistivity @ 20°C ρ µΩ m	temperature co-efficient /K
Copper	0.01724	0.0039
Silver	0.0159	0.0041
Aluminium	0.0280	0.0043
Nichrome	1.080	0.0001
Tin	0.120	0.00046
Tantalum	0.1245	0.0038
Tungsten	0.0565	0.0045
Brass	0.062 – 0.078	

Derived electromagnetic units

Energy	joule	J	$kg\ m^2/s^2$
Charge	coulomb	C	A s
Voltage	volt	V	$J/C = kg\ m^2/(A\ s^3)$
Magnetic flux	weber	Wb	$V\ s = kg\ m^2/(A\ s^2)$
Magnetic flux density	tesla	T	$V\ s/m^2 = kg/(A\ s^2)$
Magnetic field intensity	amp-turn/metre	A/m	A/m
Resistance	ohm	Ω	$V/A = kg\ m^2/(A^2\ s^3)$
Inductance	henry	H	$Vs/A = kg\ m^2/(A^2\ s^2)$
Capacitance	farad	F	$C/V = A^2s^4/(kg\ m^2)$
Power	watt	W	$VA = J/s = kg\ m^2/s^3$

INDEX