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 Stefan-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).

***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.

The one-dimensional model for general molecular (non-radiation) heat transfer is given by

$$P_d = -\lambda A \frac{\delta T}{\delta \ell} + \rho_m A \ell \frac{\delta T}{\delta t} \qquad \text{(W)} \qquad (5.2)$$

where $\delta T = T_2 - T_1$ or ΔT, is the temperature difference between two 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).

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

$$P_d = \frac{\lambda}{\ell} A \Delta T \qquad \text{(W)} \qquad (5.3)$$

***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

$$P_d = h A \Delta T \qquad \text{(W)} \qquad (5.4)$$

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 ranging from 10 to 100. Values for forced liquid convection are 50 to 20,000, while the h 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

$$P_d = 1.35 A \sqrt[4]{\frac{\Delta T^5}{\ell}} = 1.35 A \sqrt[4]{\frac{\Delta T}{\ell}} \Delta T = hA\Delta T \qquad \text{(W)} \qquad (5.5)$$

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*

$$P_d = h A \Delta T = 3.9 \sqrt{\frac{v}{\ell}} A \Delta T \qquad \text{(W)} \qquad (5.7)$$

- *for turbulent flow*

$$P_d = h A \Delta T = 6.0 \sqrt[5]{\frac{v^4}{\ell}} A \Delta T \qquad \text{(W)} \qquad (5.8)$$

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

$$R_\theta = \frac{\Delta T}{P_d} = \frac{1}{hA} = \frac{\ell}{\lambda A} \quad \text{(K/W)} \qquad (5.9)$$

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)

$$P_d = \frac{\Delta T}{R_{\theta j\text{-a}}} = \frac{\hat{T}_j - T_a}{R_{\theta j\text{-a}}} \quad \text{(W)} \qquad (5.11)$$

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

$$P_d = \frac{\Delta T}{R_{\theta j\text{-c}}} = \frac{\hat{T}_j - T_c}{R_{\theta j\text{-c}}} \quad \text{(W)} \qquad (5.12)$$

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 (and magnetic) 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

$$R_{\theta j\text{-a}} = R_{\theta j\text{-c}} + \frac{R_{\theta c\text{-a}} \times (R_{\theta c\text{-s}} + R_{\theta c\text{-a}})}{R_{\theta c\text{-a}} + R_{\theta c\text{-s}} + R_{\theta s\text{-a}}} \quad \text{(K/W)} \qquad (5.16)$$

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

$$R_{\theta j\text{-a}} = R_{\theta j\text{-c}} + R_{\theta c\text{-a}} \quad \text{(K/W)} \qquad (5.17)$$

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:

$$R_{\theta j\text{-a}} = R_{\theta j\text{-c}} + R_{\theta c\text{-s}} + R_{\theta s\text{-a}} \quad \text{(K/W)} \qquad (5.18)$$

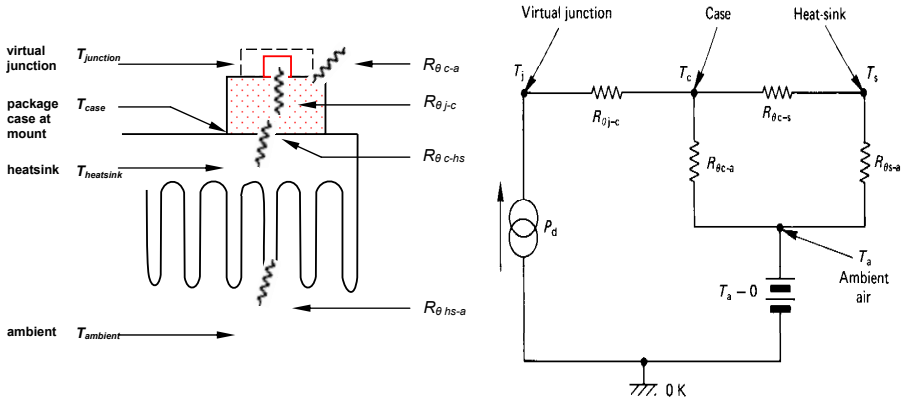


Figure 5.1. *Semiconductor thermal dissipation equivalent circuit.*

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 other than 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.

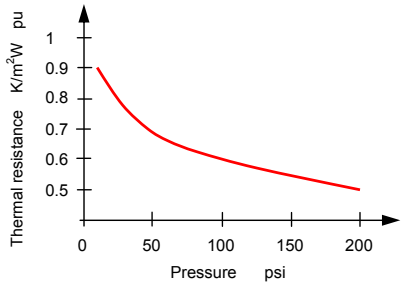


Figure 5.2. *Exemplary dependence of the gasket thermal resistance on clamping pressure.*

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 performance (that is, thermal, electrical, and physical properties), general appearance, and 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 5.3a 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 to 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 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 cm²/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
$Aℓ_2O_3$ / 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	
$Aℓ_2O_3$ 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 along 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, PCMs, 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. PCMs 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 $Aℓ_2O_3$, 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.

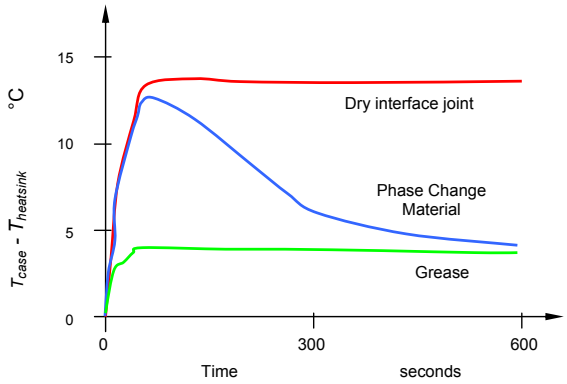


Figure 5.3. *Performance of a PCM, compared to a dry interface joint and thermal grease.*

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.

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_2O_3	0.15

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.

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 (see equation (5.22)) 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).

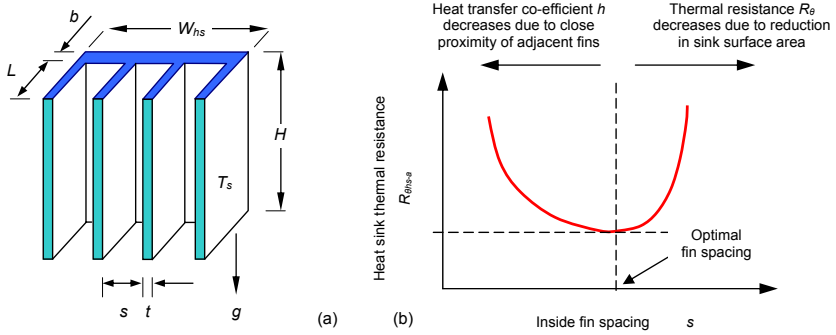


Figure 5.4. *Heat sink dimension parameters and thermal resistance dependence on fin spacing.*

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

$$\eta_f = \frac{\tanh(m_r \times H)}{m_r \times H} \quad \text{in which} \quad m_r = \sqrt{\frac{h \times P_\square}{\lambda \times A_x}} \qquad (5.22)$$

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 (= $1/v$ 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

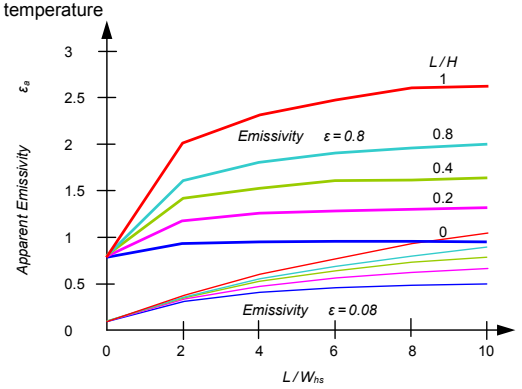


Figure 5.5. *Apparent emissivity ε_a of a channel heatsink of two different surface emissivities for different number of fins and dimensions.*

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 diffused 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.

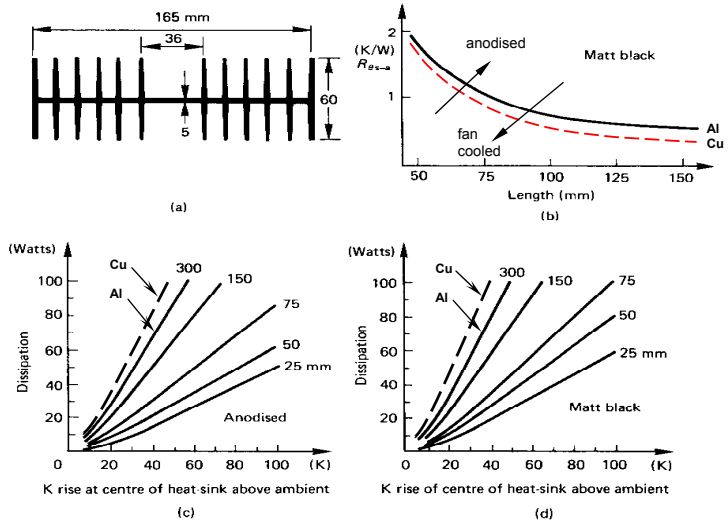


Figure 5.6. *Heat-sink typical data (for aluminium and copper):*
(a) cross-section view; (b) heat-sink length versus thermal resistance for a matt black surface finish;
(c) temperature rise versus dissipation for an anodised finish and different lengths; and
(d) as for (c) but with a matt black surface 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.

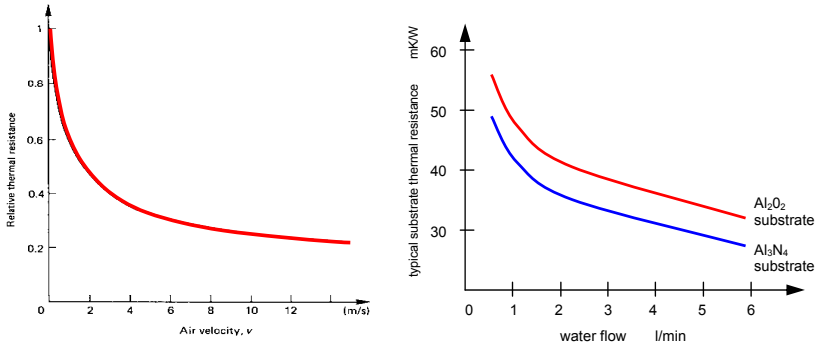


Figure 5.7. *Improved cooling with:*
(a) forced air cooled heat-sink - relative thermal resistance improvement with surface airflow and
(b) thermal resistance dependence on liquid cooling flow rate and substrate material.

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 5.7b.

Direct copper bonding
Direct copper bonding DCB is a process in which copper (on each side) and a ceramic material, usually either aluminium oxide ($A\ell_2O_3$) or aluminium nitride ($A\ell N$), are fused (bonded) together at high temperature.
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. 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 thermal cycling capabilities (up to 50,000 cycles). (See Appendix 5.27)

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 a 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 5.32 and 5.33 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 AℓN 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, as shown in figure 5.8c, 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 ΔT = 80K, 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.

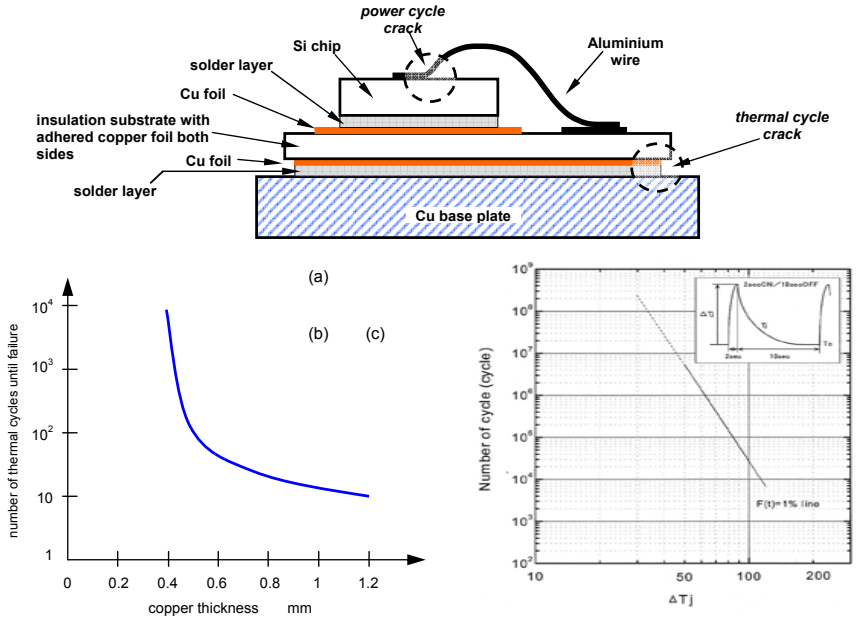


Figure 5.8. *Direct copper bonding:*
(a) sectional view of power module substrate showing boundary regions where power cycle cracking and thermal cycle cracking, occur; (b) copper thickness affect on power failure; and (c) power life thermal cycling.

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, figure 5.9a. 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

$$P_d = \frac{\hat{T}_j - T_c}{R_{\theta j-c}} = \frac{T_c - T_a}{R_{\theta c\text{-}s} + R_{\theta s\text{-}a}} \quad \text{(W)} \qquad (5.27)$$

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.

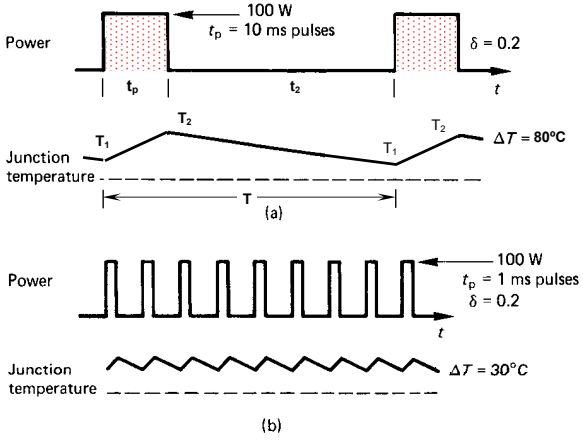


Figure 5.9. *Waveforms illustrating that peak junction temperature is a function of switching frequency: (a) lower switching frequency with 10 ms pulse and a 20 per cent duty cycle and (b) high frequency and 1 ms pulse with a duty cycle the same as in (a).*

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 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 a single 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

$$P_d = -\frac{\lambda A}{\ell} T + \rho_m A \ell \frac{\delta T}{\delta t} \qquad \text{(W)} \qquad (5.28)$$

The time domain solution for the temperature rise is

$$\Delta T(t) = \Delta \hat{T} \times \left(1 - e^{-t/\tau}\right) \qquad (5.29)$$

where the maximum temperature eventually attained if the power pulse were maintained, above ambient, is

$$\Delta \hat{T} = \frac{P_d \ell}{\lambda A} = \frac{P_d}{h A} = P_d R_\theta \qquad \text{(K)} \qquad (5.30)$$

and the thermal time constant

$$\tau = \frac{\rho_m \ell^2}{\lambda} = \frac{\text{thermal capacity, J/K}}{\text{power per K, W/K}} \qquad \text{(s)} \qquad (5.31)$$

The transient thermal impedance Z_θ is defined as

$$Z_\theta = r(t_p) R_\theta = \frac{\Delta T}{P_d} = \frac{\Delta \hat{T} \times \left(1 - e^{-t_p/\tau}\right)}{\Delta \hat{T} / R_\theta} = \left(1 - e^{-t_p/\tau}\right) R_\theta \qquad (5.32)$$

That is, thermal resistance R_θ is modified by the factor $r(t_p)$ to yield transient thermal impedance Z_θ:

$$r(t_p) = \left(1 - e^{-t_p/\tau}\right) \qquad (5.33)$$

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 the 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

$$Z_\theta = r(t_p) R_\theta = \frac{\Delta T}{P_d} = \frac{100\,\text{K}}{1103.3\,\text{W}} = 0.091\,\text{K/W}$$

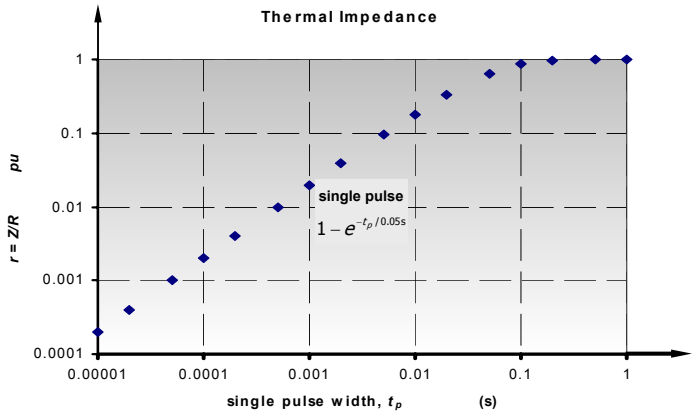
Thus the thermal resistance R_θ is modified, or normalized, by

$$r(10\text{ms}) = \frac{Z_\theta}{R_\theta} = \frac{0.0001\,\text{K/W}}{0.5\,\text{K/W}} = 0.181$$

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

$$Z_\theta(t_p, \delta) = \delta + (1 - \delta) z(t_p) \qquad \text{(K/W)} \qquad (5.34)$$

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, $t_p < 1\mu s$, 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.

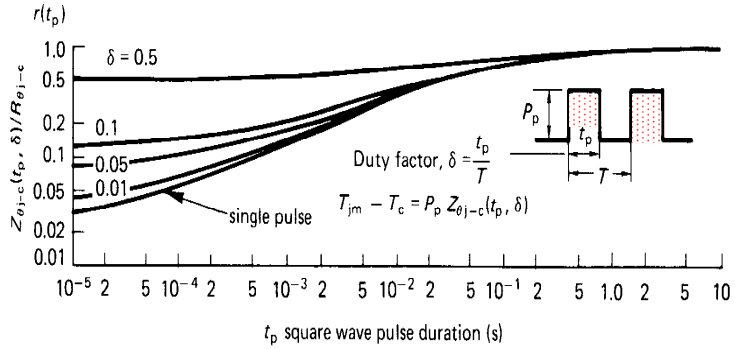


Figure 5.10. *Transient thermal impedance curves; normalised with respect to the steady state thermal resistance, $R_{\theta j\text{-}c}$.*

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 t_p / T is zero, $\delta = 0$.

$$\Delta T_{j-c} = P \times Z_{\theta j-c} = P \times r(t_p) \times R_{\theta j-c}$$

From figure 5.10, for a single 20μs pulse $r(t_p = 20\mu s) = 0.035$. 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

$$T(t) = \Delta\hat{T} - (\Delta\hat{T} - T_1) e^{-t/\tau} \qquad (5.38)$$

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

$$\Delta T_j = T_2 - T_1 = \Delta\hat{T} \frac{(1 - e^{-t_p/\tau})(1 - e^{-t_2/\tau})}{1 + e^{-T/\tau}} \qquad (5.41)$$

The maximum variation in junction temperature occurs for square-wave power, that is $t_p = t_2 = ½T$, $\delta = ½$:

$$\Delta T_j^{max} = \Delta\hat{T} \tanh\left(\frac{T}{4\tau}\right) \qquad (5.42)$$

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

$$\tau = \frac{\text{thermal capacity, J/K}}{\text{power per K, W/K}} = \frac{0.02}{1/½} = 0.01 \quad (\text{s})$$

Worst 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, $T = 20$ms, $\Delta T_j^{max} = 23.1$K

at 300Hz, $T = 3.33$ms, $\Delta T_j^{max} = 4.1$K

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, *f*=2kHz?

Solution

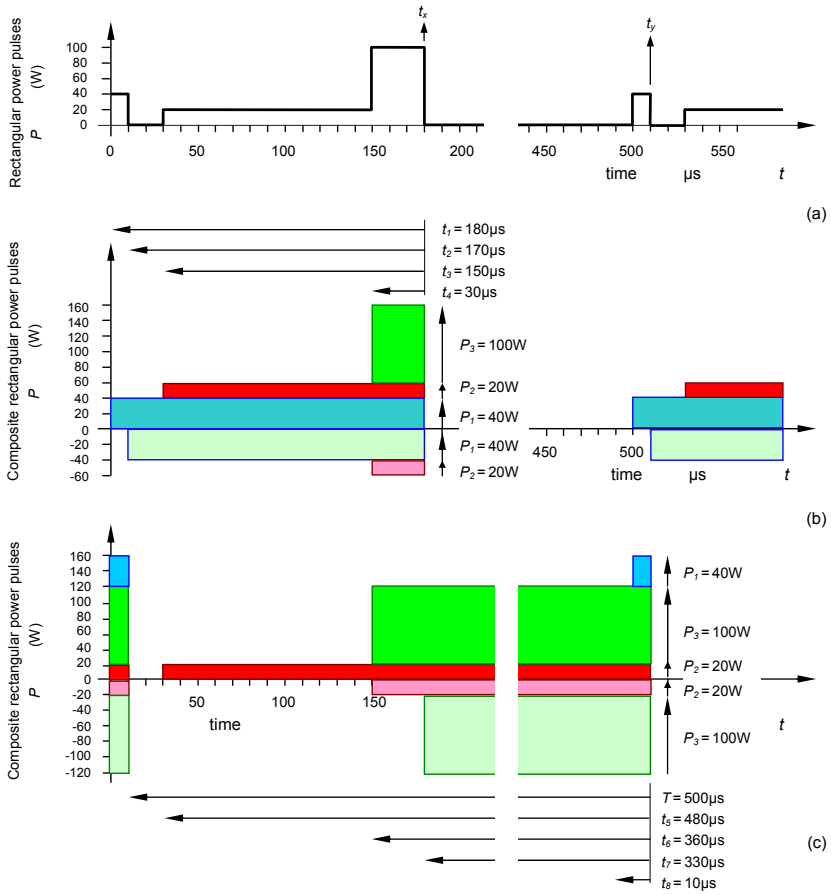


Figure 5.11. *Composite power pulses: (a) original rectangular pulse; (b) composite rectangular pulse, reference t_x; and (c) composite rectangular pulse, reference t_y.*

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.

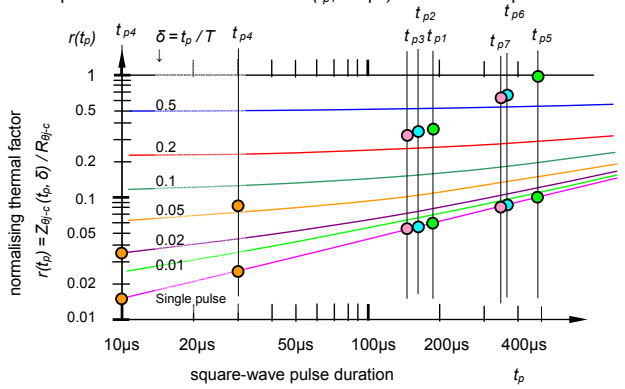


Figure 5.12. *Normalise junction to case-mount thermal impedance characteristics.*

The junction temperature (rise) at t_x, is given by

$$T_{j\text{-}c}^{t_x} = P_1 Z_{\theta j\text{-}c}^{tp1} - P_1 Z_{\theta j\text{-}c}^{tp2} + P_2 Z_{\theta j\text{-}c}^{tp3} - P_2 Z_{\theta j\text{-}c}^{tp4} + P_3 Z_{\theta j\text{-}c}^{tp4}$$

The junction temperature (rise) at t_y, is given by

$$T_{j\text{-}c}^{t_y} = P_2 Z_{\theta j\text{-}c}^{tp5} - P_2 Z_{\theta j\text{-}c}^{tp6} + P_3 Z_{\theta j\text{-}c}^{tp6} - P_3 Z_{\theta j\text{-}c}^{tp7} + P_1 Z_{\theta j\text{-}c}^{tp8}$$

t_x repetitive

$$T_{j\text{-}c}^{t_x} = 40 \times Z_{\theta j\text{-}c}^{tp1} - 40 \times Z_{\theta j\text{-}c}^{tp2} + 20 \times Z_{\theta j\text{-}c}^{tp3} - 20 \times Z_{\theta j\text{-}c}^{tp4} + 100 \times Z_{\theta j\text{-}c}^{tp4}$$
$$= 40 \times 0.38 \times 1\text{K/W} - 40 \times 0.36 \times 1\text{K/W} + 20 \times 0.35 \times 1\text{K/W} - 20 \times 0.075 \times 1\text{K/W} + 100 \times 0.075 \times 1\text{K/W}$$
$$= 13.8\text{K}$$

t_x single pulse

$$T_{j\text{-}c}^{t_x} = 40 \times Z_{\theta j\text{-}c}^{tp1} - 40 \times Z_{\theta j\text{-}c}^{tp2} + 20 \times Z_{\theta j\text{-}c}^{tp3} - 20 \times Z_{\theta j\text{-}c}^{tp4} + 100 \times Z_{\theta j\text{-}c}^{tp4}$$
$$= 40 \times 0.06 \times 1\text{K/W} - 40 \times 0.055 \times 1\text{K/W} + 20 \times 0.05 \times 1\text{K/W} - 20 \times 0.025 \times 1\text{K/W} + 100 \times 0.025 \times 1\text{K/W}$$
$$= 2.7\text{K}$$

t_y repetitive

$$T_{j\text{-}c}^{t_y} = 20 \times Z_{\theta j\text{-}c}^{tp5} - 20 \times Z_{\theta j\text{-}c}^{tp6} + 100 \times Z_{\theta j\text{-}c}^{tp6} - 100 \times Z_{\theta j\text{-}c}^{tp7} + 100 \times Z_{\theta j\text{-}c}^{tp8}$$
$$= 20 \times 0.96 \times 1\text{K/W} - 20 \times 0.73 \times 1\text{K/W} + 100 \times 0.73 \times 1\text{K/W} - 100 \times 0.67 \times 1\text{K/W} + 40 \times 0.035 \times 1\text{K/W}$$
$$= 12.2\text{K}$$

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

$$\frac{10\mu s\times 40\text{W} + 120\mu s\times 20\text{W} + 30\mu s\times 100\text{W}}{500\mu s} = 11.6\text{W}$$

The average junction to case mounting temperature rise is

$$T_{j-c} = P_{ave}\times R_{\theta j-c} = 11.6\text{W}\times 1\text{K/W} = 11.6\text{K}$$

♣

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 duration 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. Worst 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.

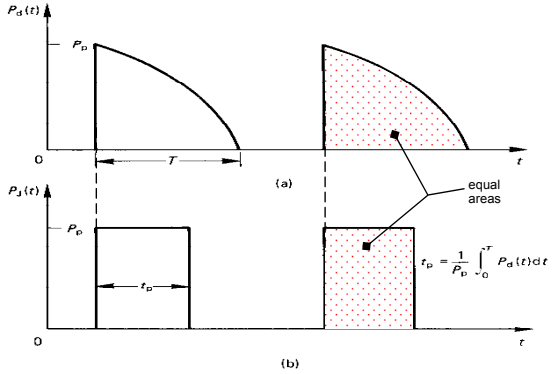


Figure 5.13. *Conversion of non-rectangular power pulse (a) into equivalent rectangular pulse (b).*

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 of the 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, $T = 1000\mu s$

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 drop, $T_{j\text{-}c}$.

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.

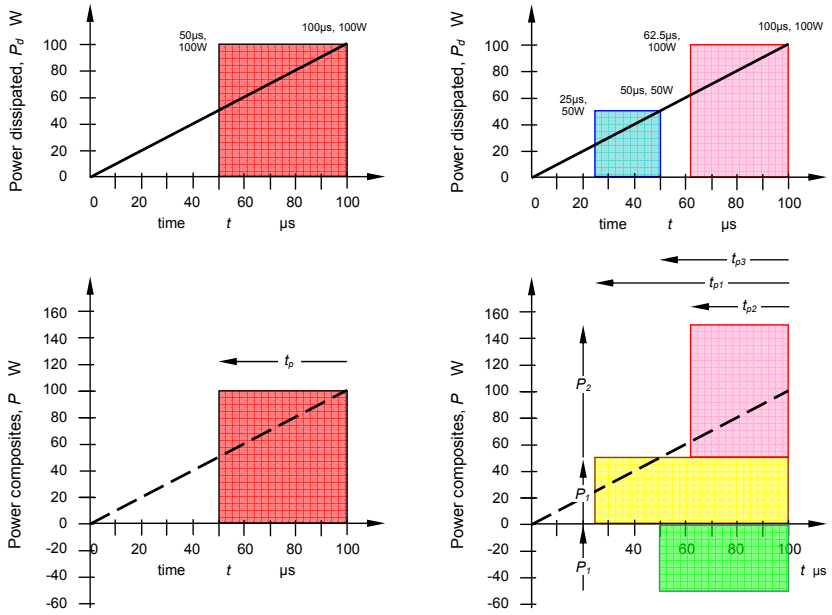


Figure 5.14. *Transient thermal impedance curves; normalised with respect to the steady state thermal resistance, $R_{\theta\ j\text{-}c}$.*

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.

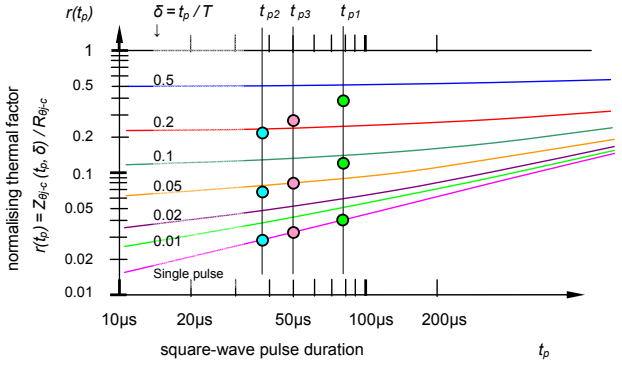


Figure 5.15. *Transient thermal impedance curves; normalised with respect to the steady state thermal resistance, $R_{\theta\ j\text{-}c}$.*

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 $T \to \infty$	$\delta = t_p/T$	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, $\delta = 0.5$

$$\Delta T_{j-c} = 50\text{W} \times r(t_{p1} = 75\mu\text{s}) \times 1\text{K/W} - 50\text{W} \times r(t_{p3} = 50\mu\text{s}) \times 1\text{K/W} + 100\text{W} \times r(t_{p2} = 37.5\mu\text{s}) \times 1\text{K/W}$$
$$= 50\text{W} \times 0.40 \times 1\text{K/W} - 50\text{W} \times 0.32 \times 1\text{K/W} + 100\text{W} \times 0.22 \times 1\text{K/W}$$
$$= 26\text{K}$$

For a 10% duty cycle, $\delta = 0.1$

$$\Delta T_{j-c} = 50\text{W} \times r(t_{p1} = 75\mu\text{s}) \times 1\text{K/W} - 50\text{W} \times r(t_{p3} = 50\mu\text{s}) \times 1\text{K/W} + 100\text{W} \times r(t_{p2} = 37.5\mu\text{s}) \times 1\text{K/W}$$
$$= 50\text{W} \times 0.12 \times 1\text{K/W} - 50\text{W} \times 0.08 \times 1\text{K/W} + 100\text{W} \times 0.066 \times 1\text{K/W}$$
$$= 8.6\text{K}$$

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 dependent on the duty cycle, which specifies the average power dissipation.

$$\bar{P}_d = \frac{\text{sawtooth area}}{T}$$
$$= \frac{½ \times 100\text{W} \times 100\mu\text{s}}{T} = \delta \times 50\,\text{W}$$

Thus the average case to junction temperature drop is

$$T_{j-c} = \bar{P}_d \times R_{\theta j-c}$$
$$= \delta \times 50\text{W} \times 1\text{K/W} = 50 \times \delta\,\text{K}$$

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

$$P_d = f_s \int_0^{1/f_s} v(t)\, i(t)\, dt \qquad \text{(W)} \qquad (5.43)$$

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

$$P_s = \tfrac{1}{6} V_s I_m \tau f_s \qquad \text{(W)} \qquad (5.44)$$

and for an inductive load, as derived in Chapter 6

$$P_s = ½ V_s I_m \tau f_s \qquad \text{(W)} \qquad (5.45)$$

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

$$P_\ell = I_\ell V_s (1-\delta) \qquad \text{(W)} \qquad (5.46)$$

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.

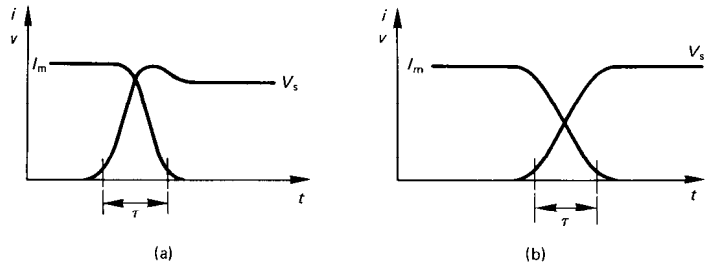


Figure 5.16. *Typical voltage and current at turn-off switching transition for: (a) an inductive load and (b) a resistive load. Current and voltage are interchanged at turn-on.*

5.6.3 Conduction power loss, P_c

The average conduction power loss under a steady-state current condition is given by

$$P_c = \delta I_{on} V_{on} \qquad \text{(W)} \qquad (5.47)$$

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 (3.16) and (4.12))

$$\begin{aligned} P_c &= \delta\, I^2_{d(rms)} R_{ds(on)} \\ &\simeq \delta\, I^2_{d(rms)} R_{ds(on)}(25°C)\left\{1+\frac{\alpha}{100}\right\}^{T_j-25°C} \qquad \text{(W)} \end{aligned} \qquad (5.48)$$

where α is the temperature coefficient of the on-state resistance, which is positive. A linear resistance approximation of equation (5.48) is accurate above 25°C if α is small, such that P_c can be approximated by

$$P_c \approx \delta\, I^2_{d(rms)} R_{ds(on)}(25°C)\left\{1+\alpha(T_j-25°C)\right\} \qquad \text{(W)} \qquad (5.49)$$

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

$$P_g = \delta\, I_G V_{GC} \qquad (5.50)$$

  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 (4.36):

$$P_G(R_{\text{int}}) = V_{gg} Q_T f_s \frac{R_{G\text{int}}}{R_{G\text{int}} + R_{G\text{ext}}} \qquad \text{(W)} \qquad (5.51)$$

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

$$J_R = V_R Q_R \qquad \text{(J)} \qquad (5.52)$$

where V_R is the reverse voltage applied to the diode just after turn-off. Losses are lower since the diode only supports voltage once peak reverse recovery has occurred (at the peak of the reverse current). The reverse recovery average power loss is given by

$$P_s \approx V_R Q_R f_s \qquad \text{(W)} \qquad (5.53)$$

The total average power loss is the algebraic sum of the steady-state conduction loss and the recovery loss.

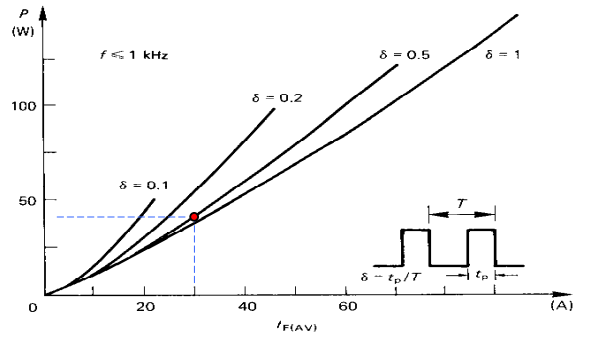


Figure 5.17. *Diode on-state energy loss at low frequency as a function of forward current for: (a) squarewave power pulses and (b) sinusoidal power pulses.*

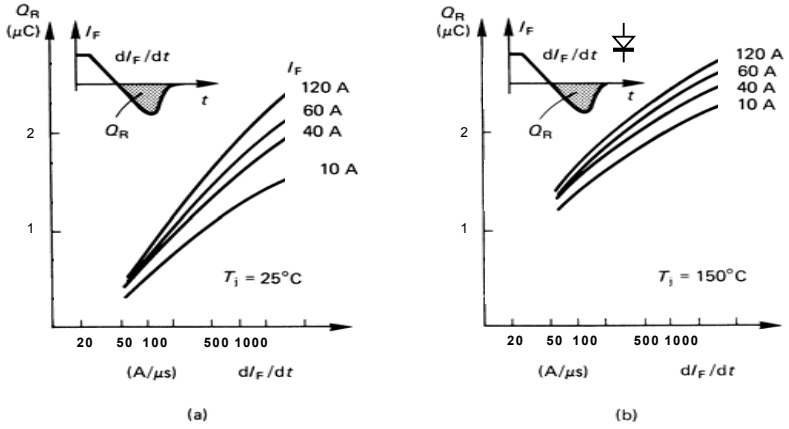
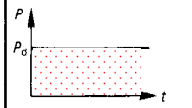
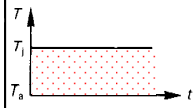
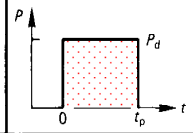
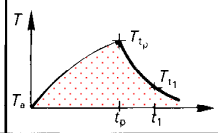
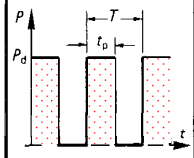
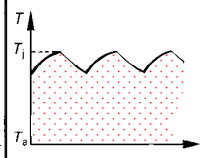
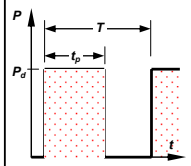
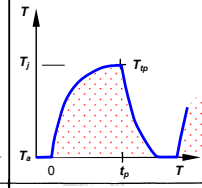
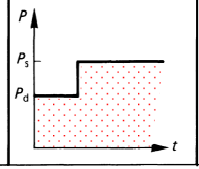
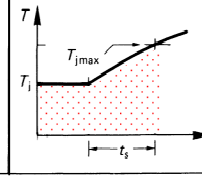


Figure 5.18. *Reverse recovery charge as a function of forward current and dI_F/dt at: (a) 25°C and (b) 150°C junction temperature.*

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 size 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 $I_{F(AV)} = 30$A for $\delta = 0.5$.
Minimum possible heat-sinking thermal resistance 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 $I_F = 60$A, 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 $\quad 150\text{V} = 40\text{V} + 45.2\text{A} \times (0.7\Omega + R_{\theta c\text{-a}})$ $\quad$ (in units of the electrical dual)

whence $\quad R_{\theta c\text{-a}} = 1.73$ 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

Externally, 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 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 interface 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

$$P_s = P_{s(\text{on})} + P_{s(\text{off})}$$
$$= \tfrac{1}{2} \times 100\text{V} \times 20\text{A} \times (1\mu\text{s} + 2\mu\text{s}) \times 10\text{ kHz} = 30\text{ W}$$

Total power losses P_d are 36W+30W = 66 W.

From

$$\hat{T}_j = T_a + P_d\left(R_{\theta j\text{-c}} + R_{\theta c\text{-s}} + R_{\theta s\text{-a}}\right)$$
$$125°\text{C} = 35°\text{C} + 66\text{W} \times \left(0.7 + 0.1 + R_{\theta s\text{-a}}\right)$$

Therefore $\quad R_{\theta c\text{-a}} = 0.56$ 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; rather the concept of thermal impedance is used.

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 $\quad T_c = 60°\text{C}$

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 $\quad R_{\theta c\text{-a}} = 10$ 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

$$P_d = P_o + I^2_{d(\text{rms})} R_{ds(\text{on})}(25°\text{C})\left\{1 + \alpha\,(T_j - 25°\text{C})\right\} \qquad \text{(W)} \qquad (5.55)$$

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

$$P_d = \frac{P_o + I^2_{d(\text{rms})} R_{ds(\text{on})}(25°\text{C})\left\{1 + \alpha\,(T_a - 25°\text{C})\right\}}{1 - I^2_{d(\text{rms})} R_{ds(\text{on})}(25°\text{C})\,\alpha\, R_{\theta j\text{-a}}} \qquad \text{(W)} \qquad (5.57)$$

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 yield

$$T_j = \frac{T_a + R_{\theta j\text{-a}} I^2_{d(\text{rms})} R_{ds(\text{on})}(25°\text{C})\{1 - 25\alpha\}}{1 - \alpha R_{\theta j\text{-a}} I^2_{d(\text{rms})} R_{ds(\text{on})}(25°\text{C})} \quad \text{(W)} \qquad (5.59)$$

Assuming typical α = 0.01/K and $R_{\theta\,\text{j-a}} = R_{\theta\,\text{j-c}} + R_{\theta\,\text{c-a}}$

$$T_j = \frac{35°\text{C} + 2 \times 5^2 \times 1\Omega \times (1 - 25 \times 0.01)}{1 - 0.01 \times 2 \times 5^2 \times 1\Omega} = 145°\text{C}$$

♣

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:

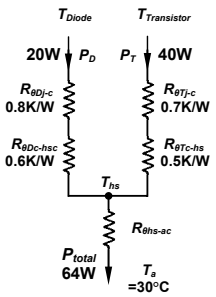
$$T_{Dj} - T_{hs} = 20\text{W} \times (0.8\text{K/W} + 0.6\text{K/W}) = 28°\text{C}$$
$$T_{Tj} - T_{hs} = 40\text{W} \times (0.7\text{K/W} + 0.5\text{K/W}) = 48°\text{C}$$

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

$$T_{Dj} - 36°\text{C} = 20\text{W} \times (0.8\text{K/W} + 0.6\text{K/W}) \quad \text{that is} \quad T_{Dj} = 64°\text{C}$$
$$T_{Tj} - 36°\text{C} = 40\text{W} \times (0.7\text{K/W} + 0.5\text{K/W}) \quad \text{that is} \quad T_{Tj} = 84°\text{C}$$

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 both P (since the on-state voltage is the same for both devices and the current is constant hence the same for both when each device conducts) 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

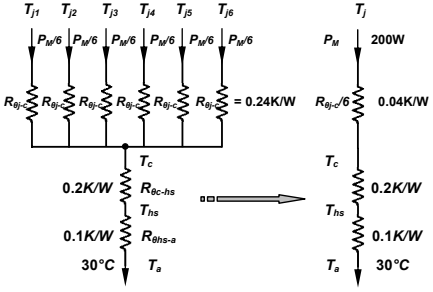
i. the diode junction temperature
ii. the current to double the rectifier lifetime (decrease junction temperature by 10°C)
iii. the heatsink to double the rectifier bridge lifetime (at 100A).

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

$$T_j - T_a = P_M \times \left(\tfrac{1}{6} R_{\theta j-c} + R_{\theta c-hs} + R_{\theta hs-a}\right)$$
$$T_j - 30°\text{C} = 200\text{W} \times \left(\tfrac{1}{6} \times 0.24\text{K/W} + 0.2\text{K/W} + 0.1\text{K/W}\right)$$
$$\Rightarrow \quad T_j = 98°\text{C}$$

ii. If the current is reduced so as to decrease the diode junction temperature by 10°C then

$$T_j - T_a = P_M \times \left(\tfrac{1}{6} R_{\theta j-c} + R_{\theta c-hs} + R_{\theta hs-a}\right)$$

$$88°C - 30°C = P_M \times \left(\tfrac{1}{6} \times 0.24K/W + 0.2K/W + 0.1K/W\right) \quad \Rightarrow \quad P_M = 170.6W$$

Assuming the diode on-state voltage drop is independent of current, that is remains 1V then

$$P_M = 2 \times I_o \times V_{Don}$$

$$170.6W = 2 \times I_o \times 1V \quad \Rightarrow \quad I_o = 85.3A$$

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

$$T_j - T_a = P_M \times \left(\tfrac{1}{6} R_{\theta j-c} + R_{\theta c-hs} + R_{\theta hs-a}\right)$$

$$88°C - 30°C = 200W \times \left(0.04K/W + 0.2K/W + R_{\theta hs-a}\right) \quad \Rightarrow \quad R_{\theta hs-a} = 0.5K/W$$

♣

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, using equation (5.4), 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.

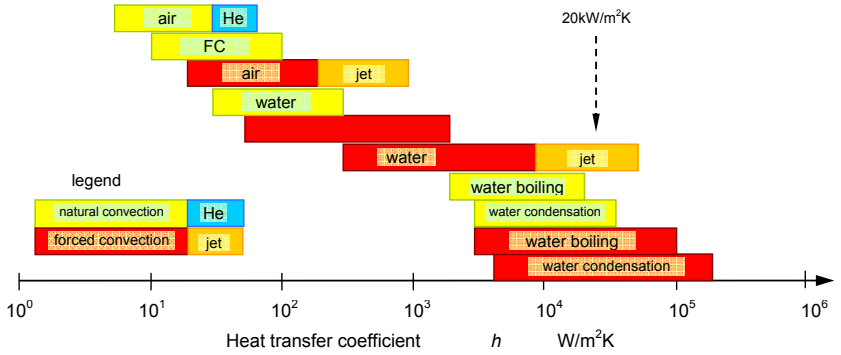


Figure 5.19. *Heat transfer coefficient h attainable with natural convection, single-phase liquid forced convection, and boiling for different coolants. (see figure 5.50)*

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.

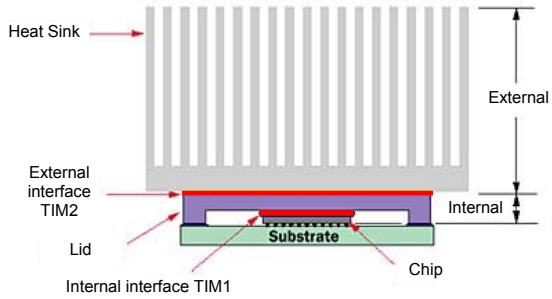


Figure 5.20. *Power semiconductor package with thermal conduction path to heat sink via TIMs.*

For high-power applications, the interface thermal resistance becomes an important constraint. Direct soldering (for example, reflow soldering) is often difficult, particularly 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 alternative 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, are much lighter, and have tuneable CTEs.
By employing 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 shows 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 to 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.

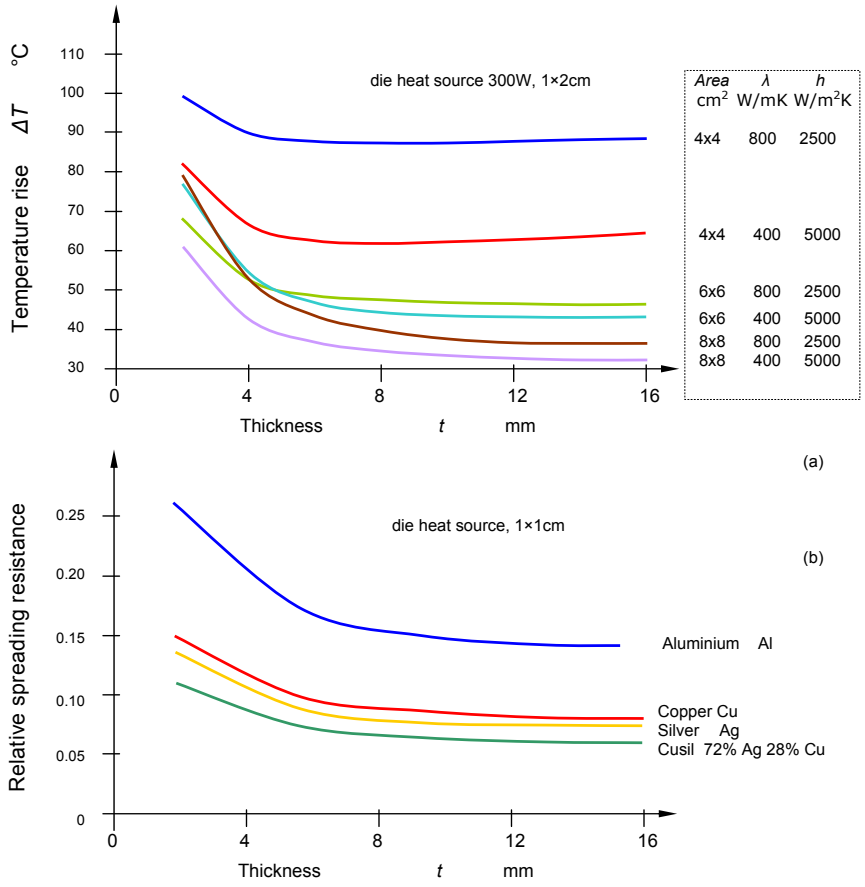


Figure 5.21. *Example of effect of thickness of heat spreader for various: (a) heat source areas, material thermal conductivities, and heat transfer coefficients and (b) material and resultant relative heat spreading resistance.*

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 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 heat sink type 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 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 shown 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 lower 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 for 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 slightly.
Unless the component is facing a hotter surface nearby, it is imperative to have the heat sink surfaces suitably 'black' 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; and
- Foam-fin – suitable for ducted airflow with a high pressure drop.

Heat sinks can be classified in terms of manufacturing methods and their final form shapes. Conventional heat sink manufacturing methods involve 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. 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 the requirement of small inter-fin spacing. These heat sinks are usually combined with a heat pipe to provide an effective thermal solution.

***Bonded/fabricated fins*:**
Bonded heat sinks are often built-up extrusions, typically manufactured by assembling extruded plates into slots or grooves on an extruded or machined heat sink base plate, and held in place by an interface, usually a two part thermosetting thermally conductive aluminium-filled 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. The bonding 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 using a punch. A common problem in forging is 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. This contact resistance is small, 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 swarf metal.

Swaged fins:
Individual fins are placed in a pre-grooved base, and then rollers swage 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 representing a single point on a fan characteristic curve.

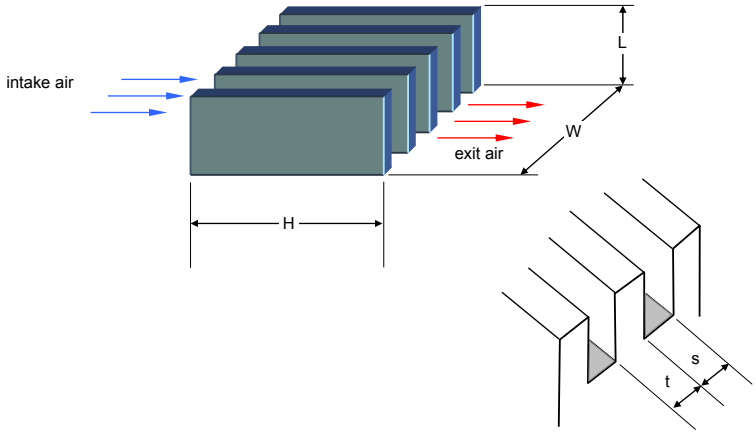


Figure 5.22. *Geometry of the plate fin heat sink analyzed.*

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 particularly small and relatively 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.

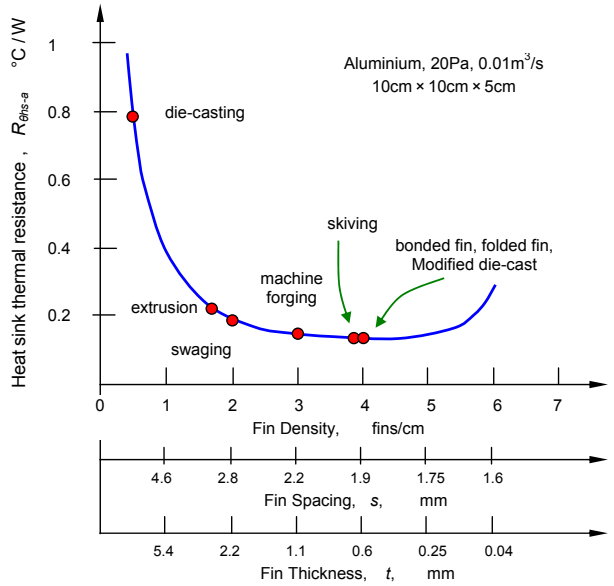


Figure 5.23. *Typical thermal characteristic plot for a heat sink: Thermal Resistance versus Fin Density, 0.1mx0.1m, 0.05m, Aluminium, 20Pa, 0.01m³/s.*

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	20:1	50:1	25:1	50:1
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.

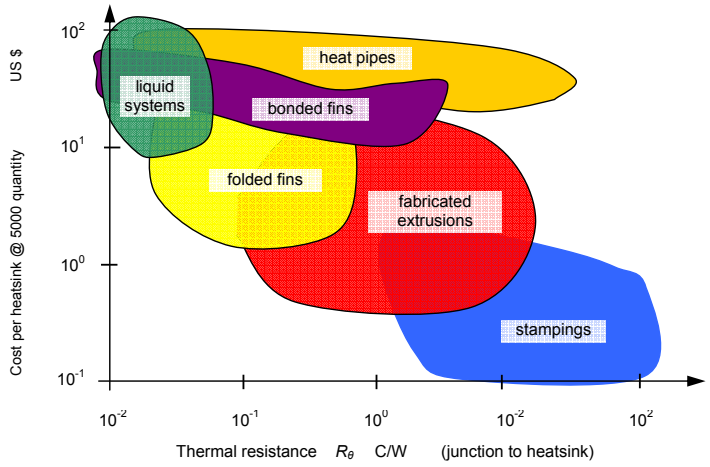


Figure 5.24. *Cost versus required thermal resistance.*

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.

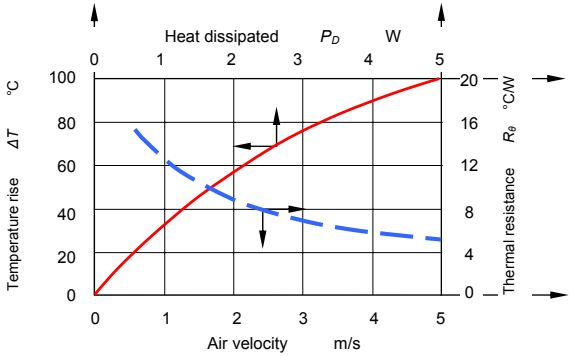


Figure 5.25. *Typical finned heatsink thermal performance graphs.*

The natural convection curves assume an optimal orientation of the heat sink with respect to 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 (thermoplastics) 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 composite plastic heat sinks (moulded base, folded fin) are 50% lighter than aluminium equivalents and advantageously are available in electrically insulative and electrical conductive grades.

5.11 Heatsink cooling enhancements

Other than 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 (Peltier devices)
  - 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.

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. The specific ratio - the ratio of the discharge pressure over the suction pressure - is used for defining fans (<1.11), blowers (1.11 to 1.20), and compressors (>1.20). *Axial fans* take and deliver air in an overall direction that is parallel to the fan blade axis (no change in direction) and can be designed to deliver a high flow rate, but 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 air changes direction twice (on entering and on leaving).

The most common *axial 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.

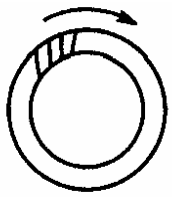
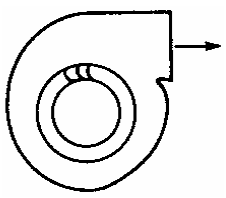
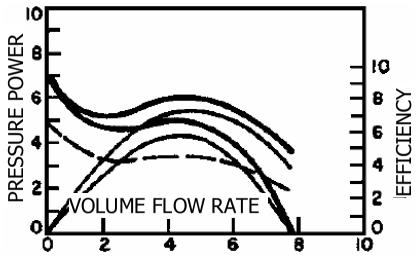
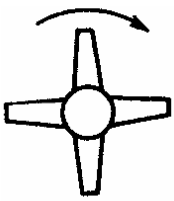
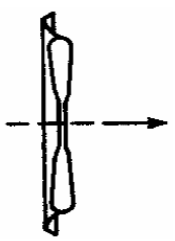
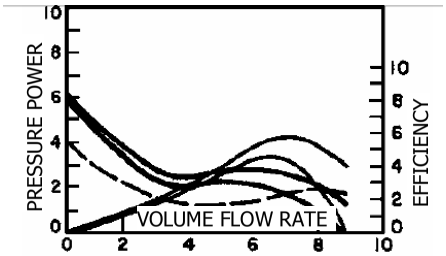
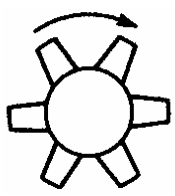
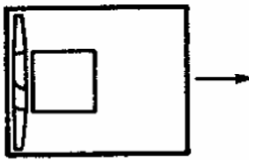
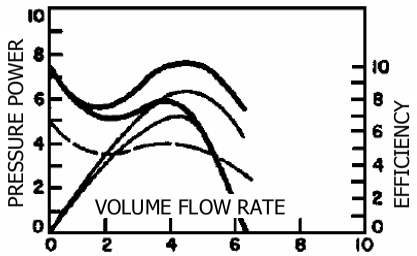
Centrifugal blowers may have a forward curved wheel, a backward curved wheel, airfoil-tubular, or be of the squirrel cage (radial) variety. They tend to be quiet: noise decreases with increased number of blades and have excellent pressure drop characteristics.

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.

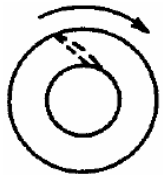
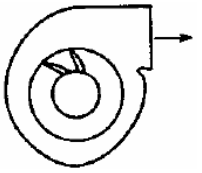
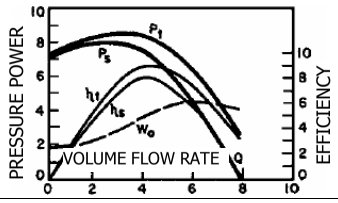
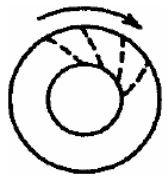
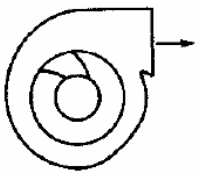
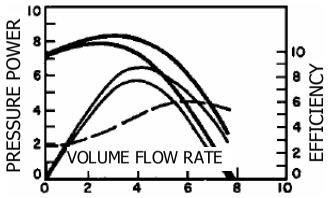
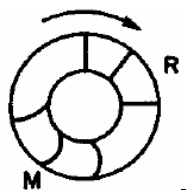
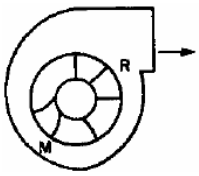
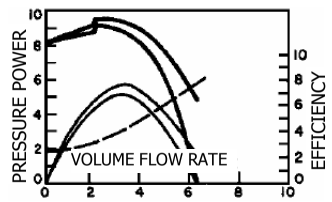
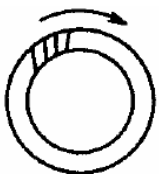
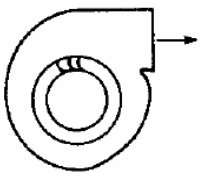
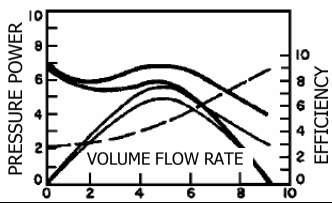
Table 5.16 Design, characteristics, and applications of axial fans and blowers

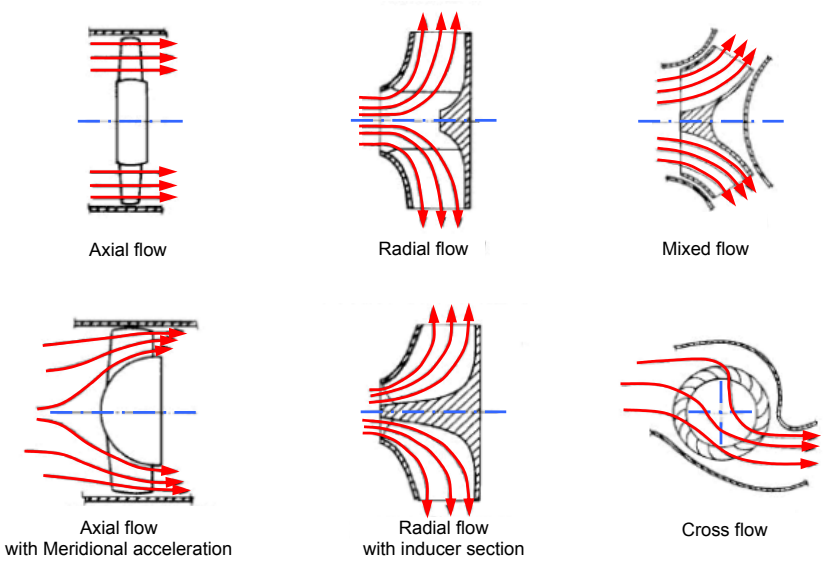
type		Impeller design	Housing design	Performance characteristics	applications
AXIAL FANS	PROPELLER	•Low efficiency •Limited to low pressure applications •Low cost impellers have at least two blades of constant thickness connected to a small hub •Primary energy transfer is by velocity pressure	•Simple circular ring, orifice plate or Venturi •Optimum design is small blade tip gap and forms smooth airfoil into wheel	• High flow rate but low efficiency and pressure capabilities • Maximum efficiency near free delivery (zero static pressure) • Discharge pattern circular and airstream swirls	•Low pressure, high volume air moving applications such as air circulation in a space or ventilation through a wall without ductwork •Some exhaust applications •Used for makeup air applications
	TUBE-AXIAL	•More efficient and develops higher static pressure than propeller fans •Usually 4 to 8 blades with airfoil or single thickness cross section •Hub less than transfer by velocity pressure	•Cylindrical tube with close clearance to blade tips	• High flow rate, medium pressure capabilities • Higher efficiency than propeller types • Performance curve dips before peak pressure • Discharge pattern circular and air stream rotates or swirls	•Low and medium pressure ducted HVAC* applications where air disturbance downstream is not critical •Some industrial applications such as drying ovens, paint spray booths, and fume exhausts *HVAC Heating Ventilation and Air Conditioning
	VANE-AXIAL	•Good blade design give medium to high pressure capability at good efficiency •Most efficient versions have airfoil blades •Blade may be fixed, adjustable or controllable pitch •Hub is usually less than half fan tip diameter	•Cylindrical tube with close clearance to blade tips •Guide vanes up or down stream from impeller increase pressure capability and efficiency	• High pressure characteristics with medium volume flow capabilities • Performance curve dips before peak pressure due to aerodynamic stall. Avoid operation in this region • Guide vanes correct circular motion imparted by wheel and improve fan pressure characteristics and efficiency	•General HVAC* systems with low, medium, to high pressure applications where straight through flow and compact installation are required •Good down-stream air distribution •Replaces tube-axial fans in industrial applications •More compact than centrifugal fans for same duty

BWW

type		Impeller design	Housing design	Performance characteristics
Axial fans	PROPELLER			
	TUBE-AXIAL			
	VANE-AXIAL			

type		Impeller design	Housing design	Performance characteristics	applications
CENTRIFUGAL FANS	AIRFOIL	• Highest efficiency of all centrifugal fan designs • 10 to 16 blades of airfoil contour curve backwards from rotation direction. Deep blade allow for efficient expansion within blade passage • Air leaves passage at velocity less than tip speed • For a given duty, has highest speed of centrifugal designs	• Scroll-type design for efficient conversion of velocity pressure to static pressure • Maximum efficiency requires close clearance and alignment between wheel and inlet	•Highest efficiencies at 50% to 60% of wide open volume. This volume has good pressure characteristics •Power reaches maximum near peak efficiency and decreases or self-limits toward free delivery	•General HVAC large systems •Large clean-air operations for significant energy savings
	BACKWARD INCLINED BACKWARD CURVED	• Efficiency slightly less than airfoil fan • 10 to 16 single thickness blades curved or inclined away from rotation direction • Efficient for same reasons as airfoil	• Uses same housing configuration as airfoil design	•Similar to airfoil fan, except peak efficiency slightly lower	•HVAC, as for airfoil types •Industrial applications where airfoil blade may corrode or erode due to environment
	RADIAL	• Higher pressure characteristics than airfoil and backward blade types • Curve has a break before pea pressure and fan should not be operated in this area • Power rises continually to free air delivery	• Scroll. Usually narrowest of all centrifugal designs • Because wheel design is less efficient, housing dimensions are not as critical as for airfoil and backward blade types	•Higher pressure characteristics than airfoil and backward blade fans •Pressure drop before peak pressure but this usually causes no problems	•Material handling in industrial plants •Rugged wheel is easy to repair. Wheel may be specially coated •Usual not for HVAC
	FORWARD CURVED	• Flatter pressure curve and lower efficiency than airfoil and backward blade types • Do not rate fan in pressure dip before peak pressure • Power rises continually to free air delivery. Motor selection must account for this	• Scroll similar to possibly identical to other centrifugal designs • Fit between wheel and inlet not as critical as for airfoil and backward blade types airfoil and backward blade types	•Pressure curve less steep than backward curved fans. Curve dips before reaching peak pressure •Highest efficiency after peak pressure at 40 to 50% of wide open volume •Rate fan to volume greater peak pressure •Account for power curve which rises continually towards free delivery, when selecting motor	•Primarily for low-pressure HVAC, such as residential furnaces, central station units and packaged air-conditioners

type		Impeller design	Housing design	Performance characteristics
CENTRIFUGAL FANS	AIRFOIL			
	BACKWARD INCLINED BACKWARD CURVED			
	RADIAL			
	FORWARD CURVED			



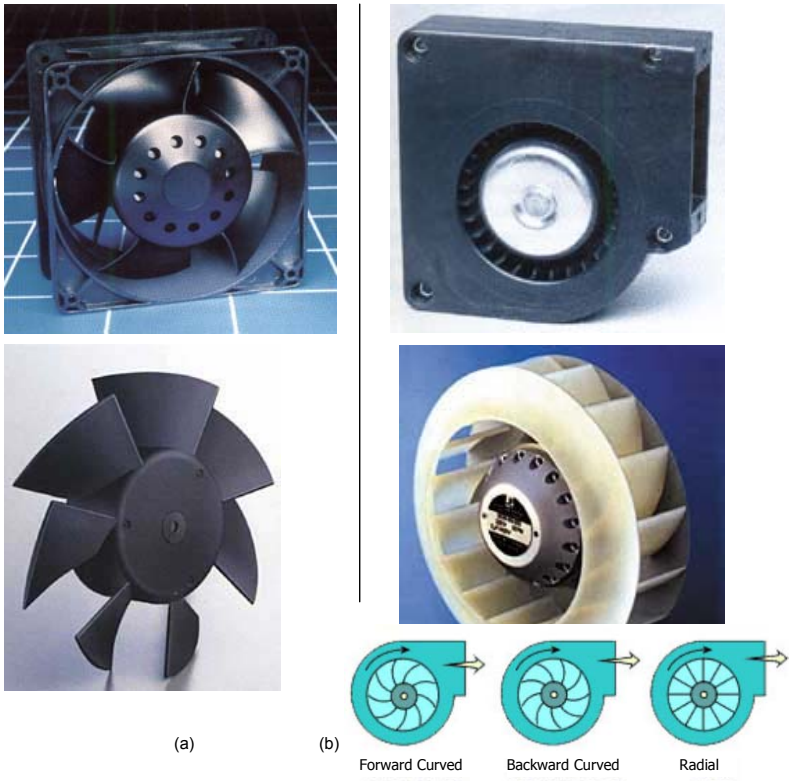


Figure 5.26. *Typical (a) impeller axial fan and (b) radial blower with backward curved blades.*

Pressure increase and flow rate in conjunction with the rotational velocity ω at the perimeter, the outer diameter D_o of the impeller, and the density ρ_f of the fluid medium result in the dimensionless parameters *pressure figure* Φ_f and *volume figure*, θ_f.

$$\phi_f = \frac{\Delta p}{\frac{1}{2}\rho_f \omega^2} \tag{5.62}$$

$$\theta_f = \frac{V}{\frac{1}{4}\pi D_o^2 \omega} \tag{5.63}$$

These 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 give a large increase in pressure and low flow rates
- Mixed-flow fans give medium pressure and medium flow rates
- Axial fans give 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 (the vertical axis). 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 maximum potential energy, hence produces a maximum possible pressure. At free delivery, point D, the fan is in the maximum kinetic energy condition. Although neither of these extreme conditions are likely to occur in practice, they are useful parameters when comparing fans. The fan stall region, B, is unstable, and should be avoided. The region C to D is the fan stable low-pressure area, where a preferred operating point near D improves efficiency and compensates for filter clogging.

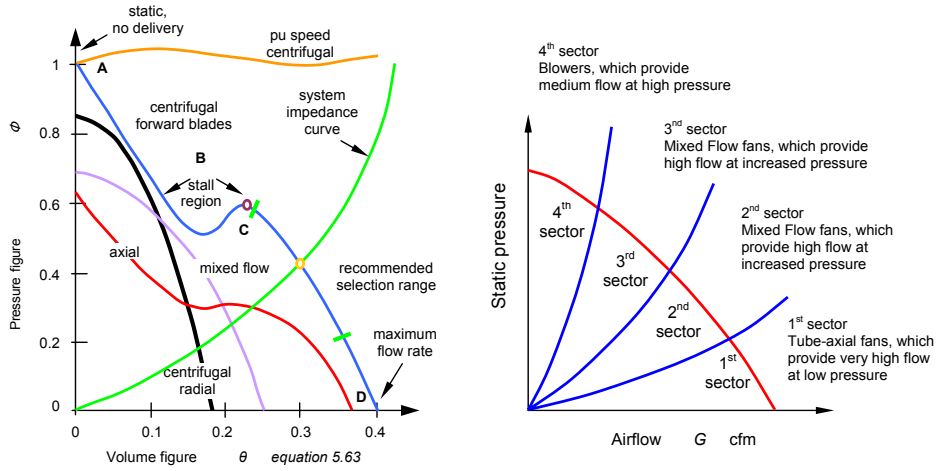


Figure 5.27. *Fan curves: (a) comparison of normalized curves for various fan designs and (b) typical fan characteristic regions.*

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. Considerations, such as available space and power, noise, reliability, and operating environment are other deciding factors.

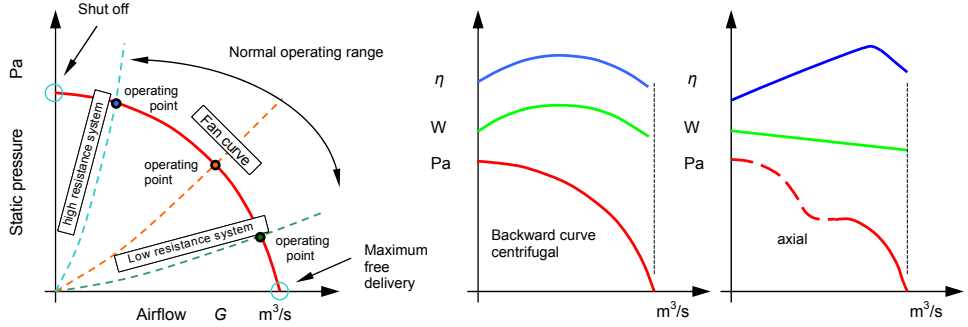


Figure 5.28. *Fan curve: (a) fan (backward curved centrifugal) system interaction; input power and efficiency (b) backward curved centrifugal and (c) axial fans.*

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 normally used and the sound pressure level obtained is expressed in dB(A).

$$SPL = 20\log\frac{P}{P_{Ref}} \qquad (5.64)$$

where P is pressure and
P_{ref} is the reference pressure

Since sound pressure level varies with distance to and direction of 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, and is unaffected by distance to the fan noise source.

$$PWL = 10\log\frac{W}{W_{Ref}} \qquad (5.65)$$

where W is the acoustic power of the source and
W_{ref} is the acoustic reference power

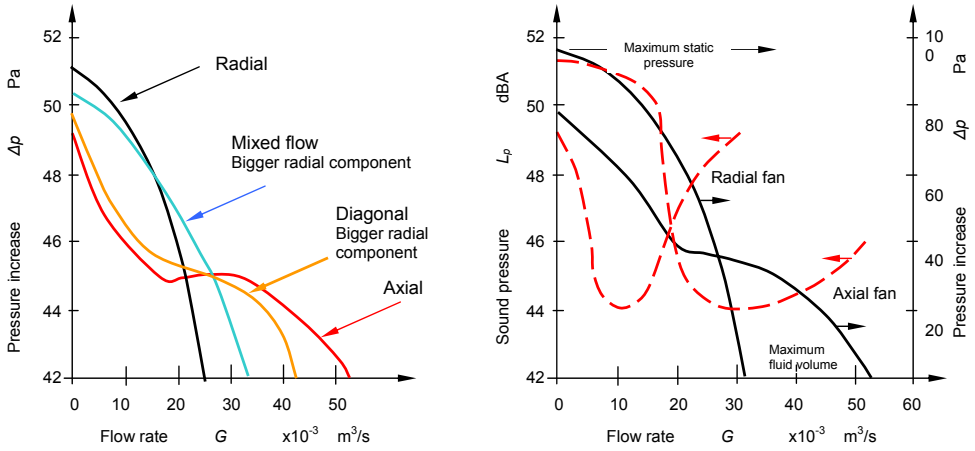


Figure 5.29. *Sound pressure level characteristics at 1m; comparison between axial and radial fans.*

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 a pressure increase or flow rate. Figure 5.29 depicts this characteristic for axial and radial fans, wherein the sound pressure level measured at a 1m distance from the fan intake side 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 5.17.
- *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 system dependent, 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 the enclosure interior surface 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:

$$P_D = m_f \times c_p \times \Delta T \qquad (5.66)$$

where P_D = amount of heat dissipated in the enclosure and transferred to the 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

$$m_f = \rho_f \times G \qquad (5.67)$$

where G = volumetric flow rate, m^3/s [1 cfm = 1.7m³/h]
ρ_f = air density, kg/m^3, which is altitude dependant, see Table 5.19

From equations (5.66) and (5.67), the required volumetric flow rate is calculated as

$$G = \frac{P_D}{\rho_\ell \times c_p \times \Delta T} \qquad (5.68)$$

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 superimposition 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 5.30. 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 ($n_q = 2$)

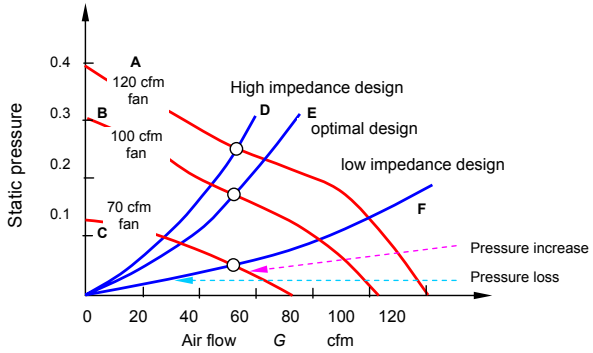


Figure 5.30. *Fan characteristics of three different fans.*

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 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, high 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 lower the flow gain 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.

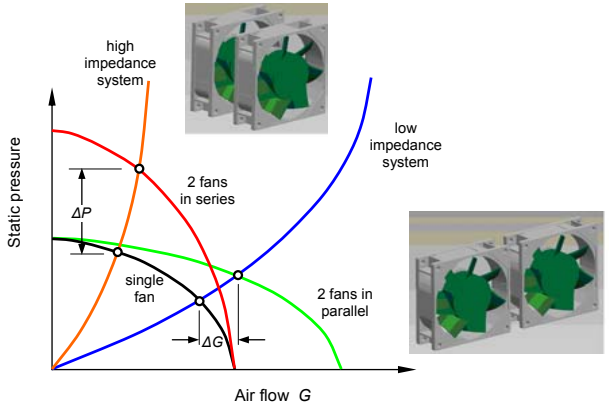


Figure 5.31. *The effects of multiple fans (series versus parallel fan operation) on system pressure and flow rate.*

5.12.2 The fan (affinity) 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:

$$\begin{aligned} \text{Volumetric Flow-rate: } & G = K_q N D^3 \\ \text{Mass Flow Rate: } & m_f = K_m \rho_i N D^3 \\ \text{Pressure: } & P = K_p \rho_i N^2 D^2 \\ \text{Power: } & HP = K_{HP} \rho_i N^3 D^5 \\ \text{sound: } & L_w = 3.71 \log_{10} V_t + 0.96 \log_{10} Q - 10.8 \quad \text{(imperial units)} \end{aligned} \tag{5.70}$$

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_i = 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

$$Re = \frac{\pi \rho_i N D^2}{C_R \nu} > 2 \times 10^6 \tag{5.71}$$

where C_R = correction factor
ν = absolute viscosity, Ns/m^2

From these relationships, it is possible to calculate fan performance at a different condition. Table 5.17 is a summary of the fan law equations in a dimensionless form, useful for fan analysis.

Table 5.17: Basic fan laws, scaling

Variable	Constants	Fan Laws
Speed (N)	Diameter (D) Density (ρ_i)	$G_2 = G_1 (N_2 / N_1)$ $P_2 = P_1 (N_2 / N_1)^2$ $HP_2 = HP_1 (N_2 / N_1)^3$
Diameter (D)	Speed (N) Density (ρ_i)	$G_2 = G_1 (D_2 / D_1)^3$ $P_2 = P_1 (D_2 / D_1)^2$ $HP_2 = HP_1 (D_2 / D_1)^5$
Density (ρ_i)	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_i)	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 15°C, when it dissipates 900W. A redesign results in the power dissipation increasing to 1200W. At 900W 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 $1m^3/s$.
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.

♣

Fan Efficiency

The overall fan efficiency, η_f, is the ratio between power transferred to the airflow (power output from the impeller) and the electrical input power used by the fan. The fan efficiency is in general independent of air density and can be expressed as:

$$\eta_f = \frac{P_f}{P_e} = \frac{\Delta p \times G}{P_e}$$

where
P_e is the electrical power input to the fan motor (W)
P_f is the fan power output calculated from volume flow (m^3/s) and pressure developed (Pa).
G = air volume delivered by the fan (m^3/s)
dp = static/total pressure (Pa)
The pressure could be total (mechanical) or static pressure and there can be a significant difference.

The equation for determining mechanical efficiency is:

$$\eta_{f\,-mechanical} = \frac{\Delta p\,(total\ pressure) \times G}{P_e} \tag{5.72}$$

The static efficiency equation is the same except that the outlet velocity pressure is not added to the fan static pressure

$$\eta_{f\,-static} = \frac{\Delta p\,(static\ pressure) \times G}{P_e} \tag{5.73}$$

Table 5.18 Fan peak efficiencies

	Fan type	Peak efficiency range, %
Centrifugal fan	Airfoil backward curved/inclined	79-83
	Modified radial	72-79
	Radial	69-75
	Pressure blower	58-68
	Forward curved	60-65
axial fan	Vane axial	78-85
	Tube axial	67-72
	Propeller	45-50

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.

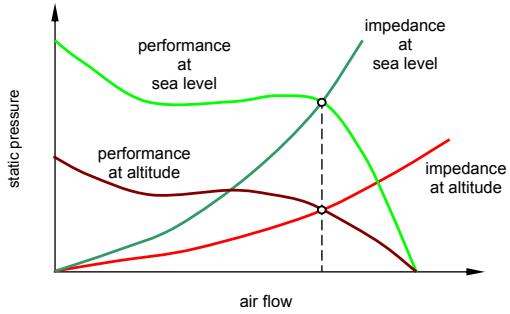


Figure 5.32. *Density effects on fan performance.*

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 5.19. 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 5.19: Air density ρ_ℓ and thermal resistance change with altitude, *z*

Altitude *z*	Density ρ_ℓ	Heatsink thermal resistance multiplier, p.u.		
m	kg / m^3	Fan-Cooled (General)	Fan-Cooled (High Power)	Naturally Cooled $\frac{R_z(z)}{R_{0\ metres}} = \frac{1}{1-5\times10^{-5}z}$
Sea Level	1.19	1.00	1.00	1.00
1,500	1.06	1.20	1.16	1.10
3,000	0.904	1.45	1.35	1.21
4,500	0.771	1.77	1.58	1.33
6,000	0.652	2.18	1.86	1.48
7,600	0.549			
9,100	0.458			

Enclosure cooling

In addition to selecting a fan, it is important to consider fan placement in or on 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 layouts 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 (hot) 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 area 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.

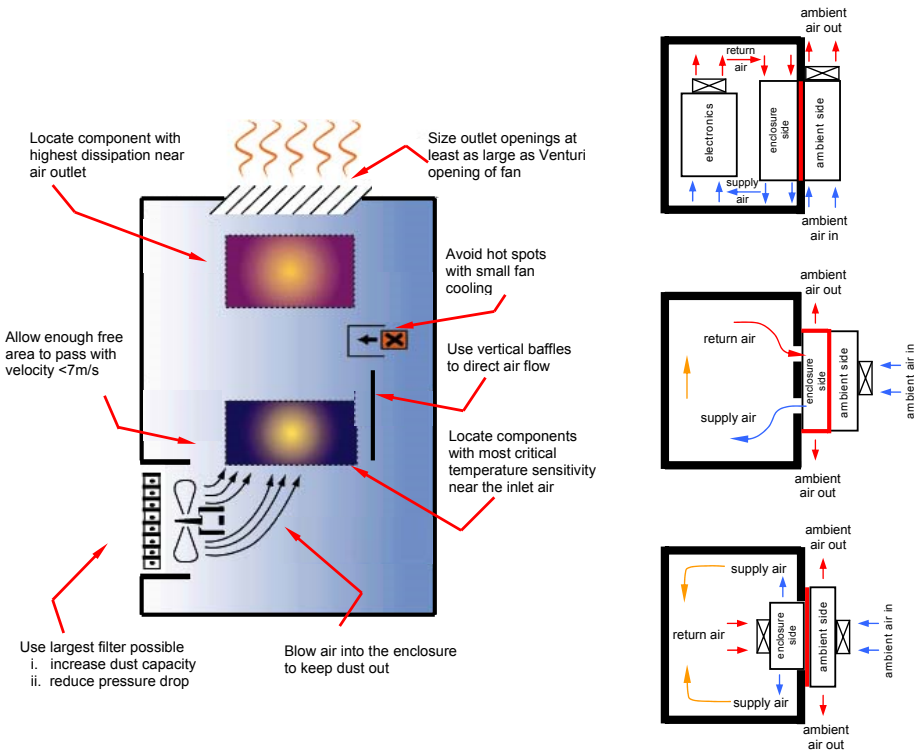


Figure 5.33. *Cabinet cooling, external mount versus through-mount, with vertical orientation.*

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

$$F(t) = 1 - e^{-\left(\frac{t}{\alpha}\right)^{\beta}} \tag{5.74}$$

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

$$H(t) = \frac{\beta_s}{\alpha_s} \times \left(\frac{t}{\alpha_s}\right)^{\beta_s - 1} \tag{5.75}$$

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 (5.74):

$$L_{10} = \alpha_s \times 0.10536^{1/ß_s} \qquad L_2 = \alpha_s \times 0.02020^{1/ß_s} \tag{5.76}$$

For example, given $\alpha_s = 100$k POH (power on hours) and $ß_s = 1.5$, $L_2 = 7,418$ 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

$$mttf = \alpha \times \Gamma\left(1 + \frac{1}{\beta}\right) \tag{5.77}$$

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 range from 25 to 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

$$\log L_{10} = -2.6 + \frac{k_T}{T_{brg}} - 0.301 \times S_{½} \tag{5.78}$$

where

$$S_{½} = S_G + S_N + S_P \tag{5.79}$$

$$S_N = 0.86 \times \frac{D_H \times N}{DN_L}$$

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 fan data sheet information.

$P = 960$g,
$C_r = 57$kg,
$D = 3$ 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 (5.79), 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

$S = S_G + S_N + S_P = 0 + 0.021 + 0.540 = 0.561$
$k_T = 2450/1.5^{(25-25)/10} = 2450/1.5^0 = 2450$, that is, assume no temperature derating.
From equation (5.78)

$$\log L_{10} = -2.6 + \frac{2450}{273.2°\text{C} + 42°\text{C}} - 0.301 \times 0.561$$
$$= -2.6 + 7.78 - 0.169 = 5.01$$

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 (5.76) gives

$$\alpha = L_{10}(0.10536)^{-1/2} = 246{,}460\,\text{hours}$$

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 $1.5^{(80-40)/10} = 5.1$.
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 (ball or sleeve) to use in the axial fan. The bearing type is a crucial factor in determining an axial fan's reliability. Table 5.20 outlines the relative features of the two bearing types when used in axial fans.

Table 5.20: 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 better 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 $h = 150\text{W/m}^2\text{K}$ can be reached with acceptable noise levels, which is about 1W/cm^2 for a 60°C temperature difference. Using 'macrojet' impingement, theoretically $900\text{W/m}^2\text{K}$ 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 (tiny wind currents) are moved by another electric field, thereby inducing secondary airflow. Cooling at a heat flux level of 40W/cm^2 (similar to water) is possible.

5.14 Liquid coolants for power electronics cooling

Although air-cooling continues to be the most widely used method for cooling 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 a 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 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 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 5.21 and Table 5.32X. Water, deionised water, glycol/water solutions, and dielectric fluids such as fluorocarbons and **p**oly**a**lpha**o**lefins PAO are the heat transfer fluids 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 meet 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. 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 5.21: Properties of different liquid coolant chemistries at 20°C

cooling chemistry	freezing point	flash point	viscosity	thermal conductivity	specific heat	density
			ν	λ	c_p	ρ_ℓ
	°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, aliphatic PAO-based fluids have replaced 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 and 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 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.

- ***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 of using facility or tap water is that it is readily available and inexpensive. However, facility water or tap water is likely to contain impurities, which cause corrosion in the liquid cooling loop and/or clog fluid channels. Therefore, use good quality water 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.
  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. If the tap water is relatively pure and meets recommended limits, a corrosion inhibitor should 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 an 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 becomes acidic when exposed to air. The carbon dioxide in air dissolves in the water, introducing ions, 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. Although ethylene glycol is the chemical used in automotive antifreeze, it 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, 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 to 30% EGW4. At this minimum concentration, the ethylene glycol also serves as a bactericide 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.

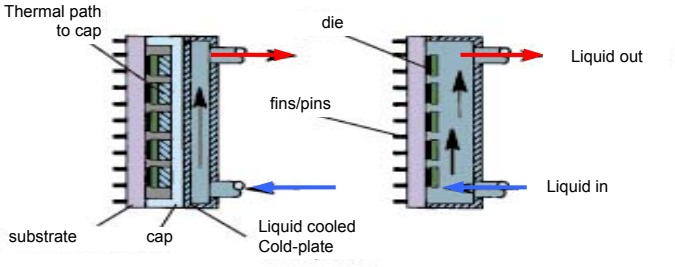


Figure 5.34. *Example of indirect and direct liquid immersion cooling for a disk package.*

5.16 Indirect liquid cooling

5.16.1 Heat pipes – indirect cooling

Heat pipes provide an indirect and passive (no moving parts) 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 (evaporator), 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 the 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 the adiabatic section to the 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 continuously circulating 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.

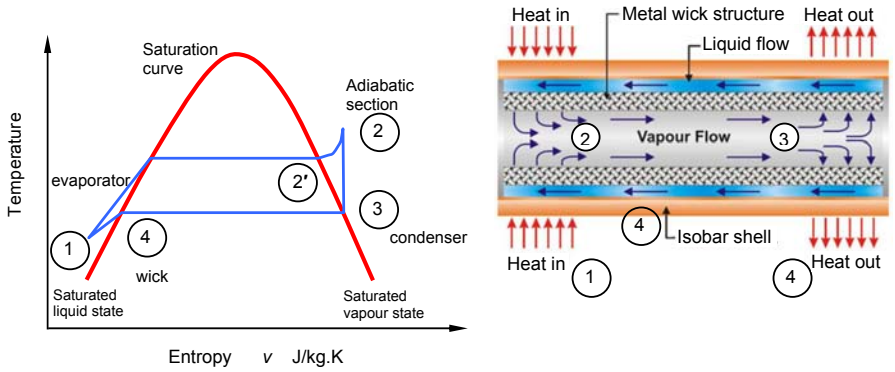


Figure 5.35. *Heat pipe thermodynamic T-v operational vapour power cycle.*

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.

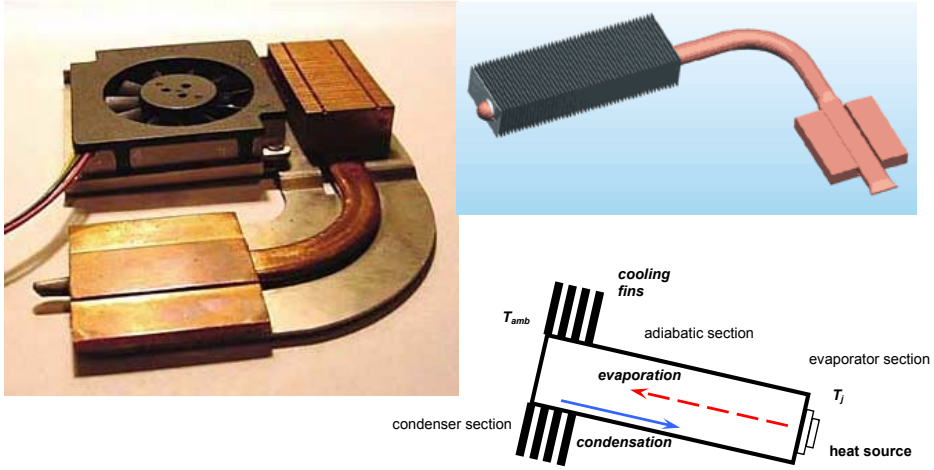


Figure 5.36. *Example of heat pipes used in a notebook application.*

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 (see figure 5.42):

- 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, and 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 creates 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 (capillary) structures available including (decreasing permeability and decreasing pore radius): grooves, screen, cables/fibred, and sintered powder metal, shown in figure 5.37 and summarized in Table 5.22. Specifically:

- *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 5.22: 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

Since grooved wicks have a large pore radius and a high permeability, 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 5.23 (see figure 5.42). 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 to 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 5.23: Heat pipe fluids, in increasing operating temperature range (see Table 5.29)

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	-200 to -80
Ammonia (ℓ)	- 78	- 33	0.295	2.95	Ni, Al, stainless steel	-70 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	-45 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	5 to 230
Toluene	- 95	110				50 to 200
Mercury/Mg	- 39	361	25.1 @ 360°C	181 @ 360°C	Stainless steel	190 to 600
Potassium	63	759	5.6 @ 750°C	181 @ 750°C	Stainless steel, Ni	400 to 800
Sodium	98	892	9.3 @ 850°C	224 @ 760°C	Stainless steel, Ni, Nb	500 to 1000
Lithium	179	1340	2 @ 1250°C	207 @ 1250°C	W, Ti, Niobium +1% Zirconium	900 to 1600
Silver	960	2212	4.1	413	Tantalum +5% Tungsten	1500 to 2200

The capillary or lifting height H is given by

$$H = \frac{2\gamma \cos\theta}{r_c g_f \rho_\ell} \qquad (5.80)$$

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 (axial power rating, APR) of a heat pipe are related to its cross-sectional area A_x and length ℓ according to equation (5.3)

$$P_D = k\frac{A_x}{\ell} \qquad \text{(W)} \qquad (5.81)$$

while the temperature difference ΔT between the hot and cold ends is

$$\Delta T = k' P_D \left(\frac{1}{A_e} + \frac{1}{A_c}\right) \qquad \text{(K)} \qquad (5.82)$$

where A_e and A_c are the effective evaporator and condenser surface areas.

As a heat pipe has no moving parts, its mttf is estimated to be over 100,000 hours of use. Improper bending and flattening of the pipe may cause leakage at 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.

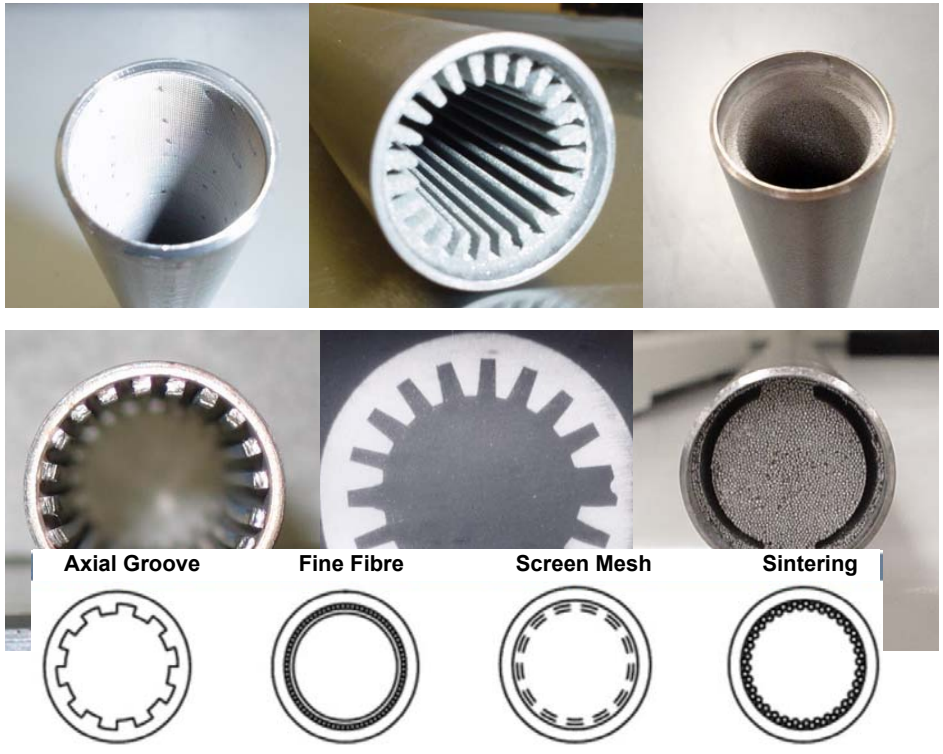


Figure 5.37. *Examples of heat pipes wicks.*

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 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. As in equation (5.81), 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 5.24, 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. This leads to evaporator dry-out, which prevents continuing the thermodynamic cycle 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 the five limitation boundaries.

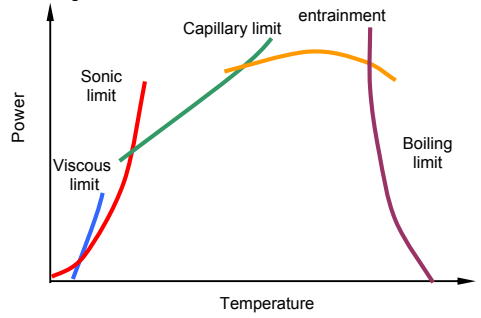


Figure 5.38. *Heat transport limitation boundaries of a heat pipe.*

For a heat pipe to function, the net capillary pressure difference between the evaporator (heat source) and condenser (heat sink) must be greater than the sum of all pressure losses occurring along the liquid and vapour flow paths. This relationship, termed the *capillary limitation*, and for correct operation:

$$\Delta P_{c\max} \geq \Delta P_{liquid} + \Delta P_{vapour} + \Delta P_g \qquad (5.83)$$

where ΔP_{cmax} is the maximum capillary pressure difference generated within the capillary wicking structure between the evaporator and condenser,
ΔP_{liquid} is the viscous pressure drops occurring in the liquid, to return the liquid from the condenser to the evaporator,
ΔP_{vapour} is the viscous pressure drops occurring in the vapour, to cause the vapour to flow from the evaporator to the condenser,

ΔP_g represents the hydrostatic pressure drop due to the gravitational head, $\rho_l g \ell \sin\varphi$.
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.

Table 5.24: 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

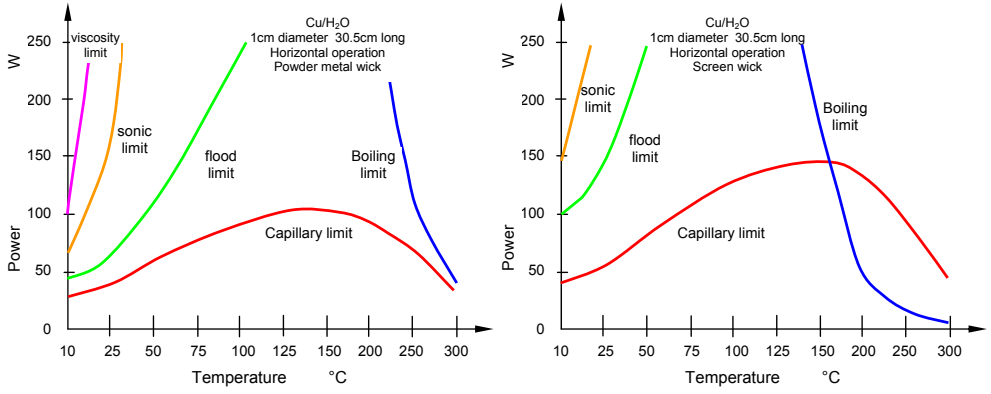


Figure 5.39. *Predicted heat pipe limitations, where the capillary limit is usually the limiting factor: (a) powder metal wick and (b) screen wick (viscosity limit on y-axis off-scale). Note temperature scale entrainment.*

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 this limit is set by the pumping capacity of the wick structure. As shown in Figure 5.40, the capillary limit is a function of the operating orientation and the type of wick structure.

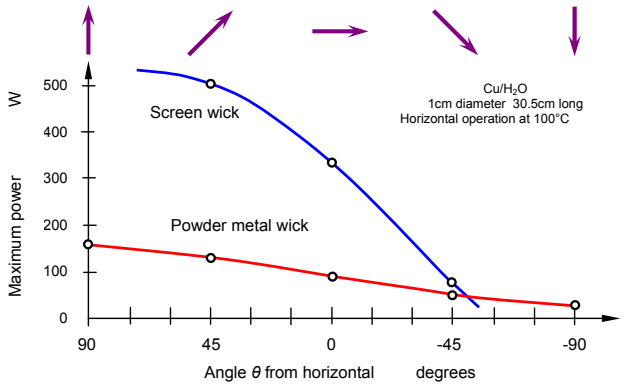


Figure 5.40. *Dependence of maximum stable heat flow on orientation, specifically capillary limits versus operating angle. Arrows indicate the direction of the heat flow.*

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:

$$\lambda_{eff} = \frac{P_T \times L_{eff}}{A_x \times \Delta T} = \frac{1}{R_{tl}} \times \frac{L_{eff}}{A_x} \qquad (5.84)$$

where: $L_{eff} = \frac{1}{2}(L_{evaporator} + L_{condenser}) + L_{adiabatic}$
λ_{eff} = 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.
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 5cm evaporator and 5cm condenser lengths.

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.

$$\Delta T = p_{evap} \times R_{evap} + p_{axial} \times R_{axial} + p_{cond} \times R_{cond}$$

$$\Delta T = 3.8\ \text{W/cm}^2 \times 0.2°\text{C/W/cm}^2 + 95.5\ \text{W/cm}^2 \times 0.02°\text{C/W/cm}^2 + 3.8\ \text{W/cm}^2 \times 0.2°\text{C/W/cm}^2$$

$$\Delta T = 3.4°\text{C}$$

♣

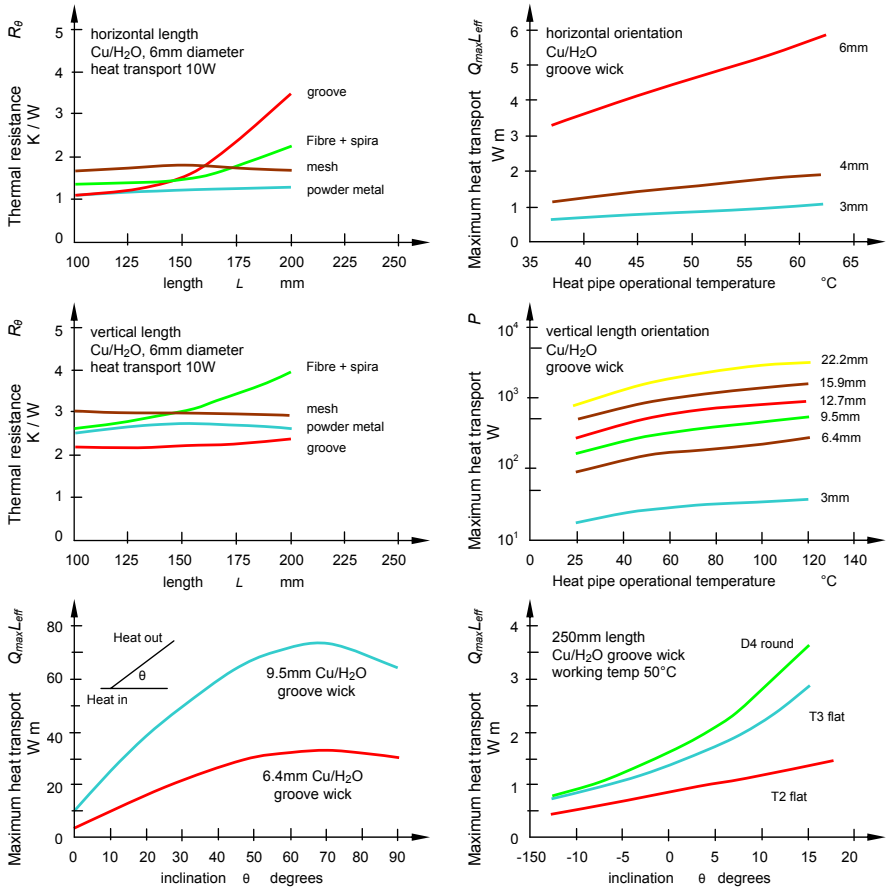


Figure 5.41. *Heat pipe parameter performance characteristics.*

Length and diameter affect on heat pipe performance

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 limits of the heat pipe. However, the operational temperature of the heat pipe also affects the sonic limit of the heat pipe. Figure 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 figure 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 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 mentioned, the groove heat pipe, with the lowest capillary limit, works best under gravity-assisted conditions where the evaporator is located below the condenser. Figure 5.41 shows the effect of gravity on groove wick heat pipes.

Performance affect of heat pipe flattening or bending

If a heat pipe is flattened or bent, the sonic and entrainment limits 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. Figure 5.41 shows the effect of flattening on a heat pipe.

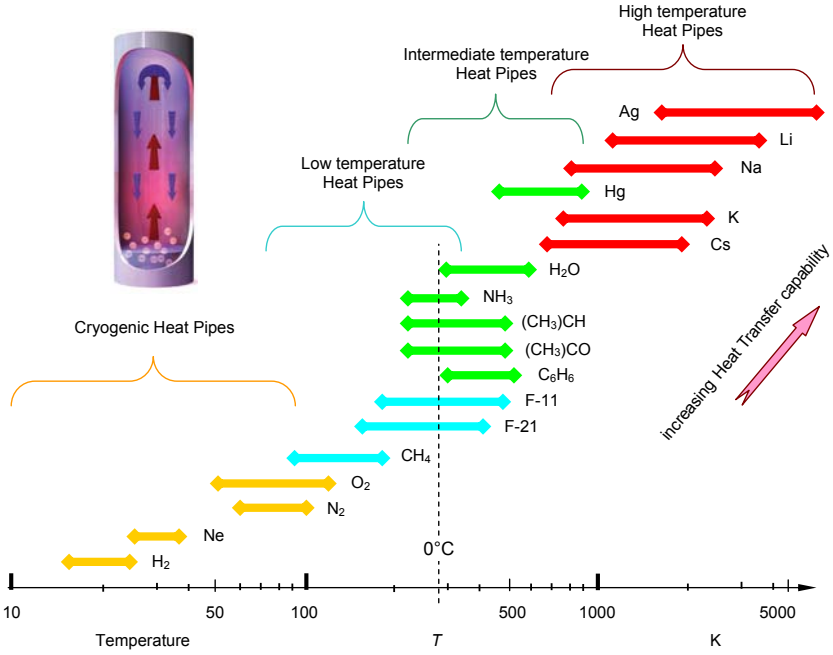


Figure 5.42. *Operating temperature ranges for various heat pipe fluids.*

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, and 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 an adjacent temperature range.

Low temperature heat pipe technology

Low order hydrocarbons are common low temperature heat pipe fluids, as specifically shown for ethane (in stainless steel) and ammonia (in an aluminium alloy) based constant conductance heat pipes in figure 5.43.

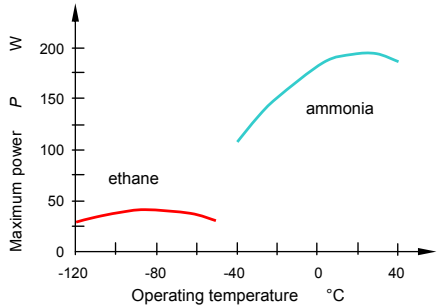


Figure 5.43. *Low temperature constant conductance heat pipe performance for Al/NH_3 and Al/C_2H_6 with slight adverse elevation, 12mm internal diameter, 60cm adiabatic section and 15cm evaporator/condenser.*

Intermediate temperature heat pipe technology

The intermediate temperature range extends from 450K 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 $TiC\ell_2F_2$ 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 the parts of Figure 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 being more desirable:

$$M_\ell = \frac{\rho_\ell \sigma \lambda}{\nu} \qquad (5.85)$$

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
ν = liquid viscosity, Pa

Figure 5.44 shows the Merit Number as a function of temperature for a number of fluids. Whilst 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 mid-region. However, Iodine has two potential problems: low liquid thermal conductivity and high corrosiveness.

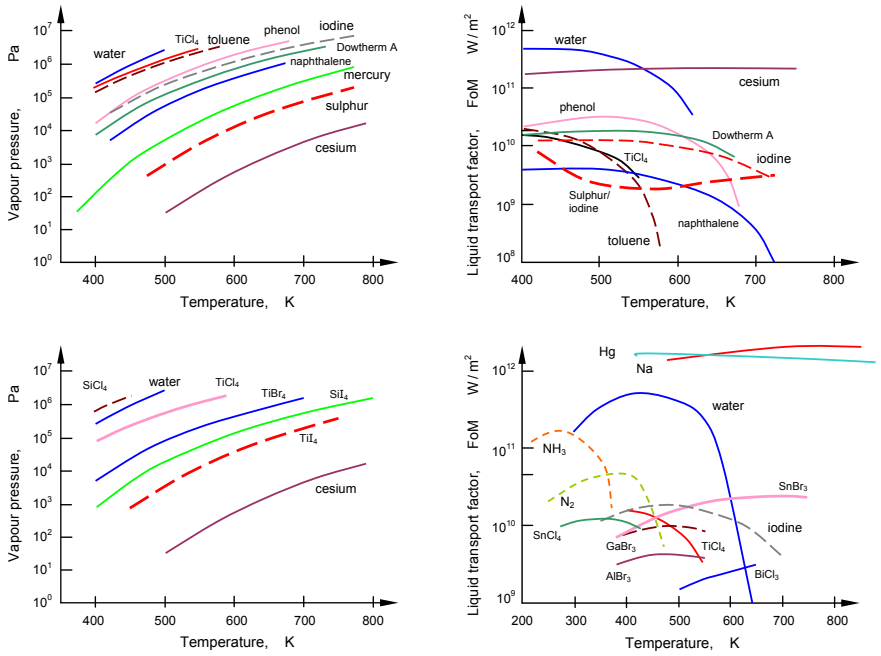


Figure 5.44. *Vapour pressure and figure of merit for intermediate temperature heat-pipe fluids, at a constant latent heat and normal boiling temperature.*

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.

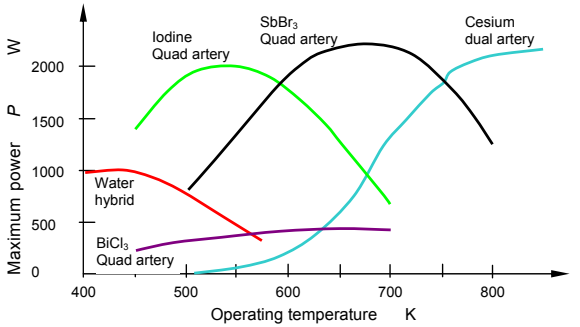


Figure 5.45. *Theoretical heat pipe powers for different fluids. Heat pipe dimensions: 2.54mm diameter, 220mm length.*

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 may be applicable.

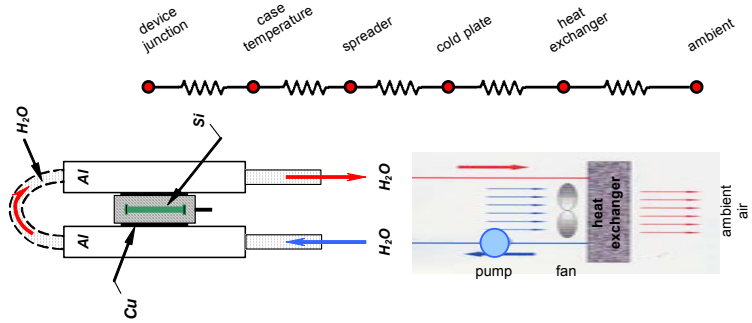


Figure 5.46. *Improved cooling with compact indirect cold-plate spreader water-cooling.*

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 fluid 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. 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.

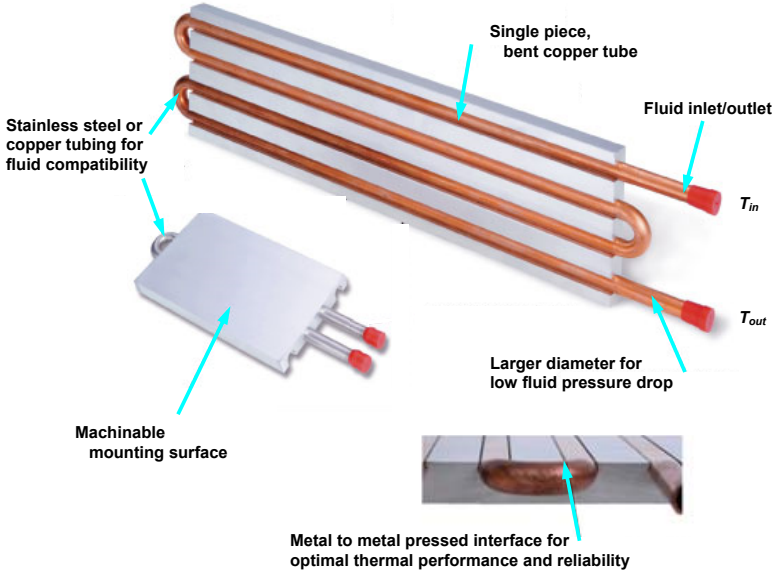


Figure 5.47. *Compact indirect cold-plate spreader water-cooler.*

Cold plates are used in closed systems where pumped coolants continuously cycle, conveying excess heat away from the devices being cooled. This heat is then dumped into the ambient air via a radiator heat exchanger or recirculating chiller, or to facility 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 transferring the conducted (dissipated) heat P_D to the coolant, a cold plate in steady-state must satisfy

$$Q_{flow} = P_D$$
$$mc_p(T_{out} - T_{in}) = P_D = hA(T_s - T_m)$$
$$\text{where } T_m = \tfrac{1}{2}(T_{out} + T_{in})$$

The heat transfer coefficient, from the heated surface, area A, at temperature T_s, is

$$h = \frac{\Delta T}{P_D} = \frac{Nu \times k}{D_H} \qquad (5.86)$$

where the Nusselt number is $Nu = 4.44$ for fully developed flow conditions.

$$\text{pressure drop}: \Delta P = \tfrac{1}{2}\frac{fL\rho V^2}{D_H} \quad (\text{Pa})$$
$$\text{pump power}: P_w = \Delta P \times G \quad (\text{W})$$
$$\text{heat exchanger length}: L = -\frac{m_f c_p}{\pi h D_H}\ell \text{n}\frac{\Delta T_{out}}{\Delta T_{in}} \quad (\text{m}) \qquad (5.87)$$
$$\text{outlet temperature}: T_{out} = T_s - (T_s - T_{in})e^{-\frac{1}{m_f c_p R_t}} \quad (\text{K})$$

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 increases with improving performance. The technologies to follow 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, ensuring good metal-to-metal contact and optimizing thermal transfer. Others secure the tubing to the channels with a performance-limiting thermal epoxy.
  Tubing material is predominantly either 5 mm or 10 mm out-side diameter copper or stainless steel. The larger diameter offers a lower pressure-drop. 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 in meeting 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.

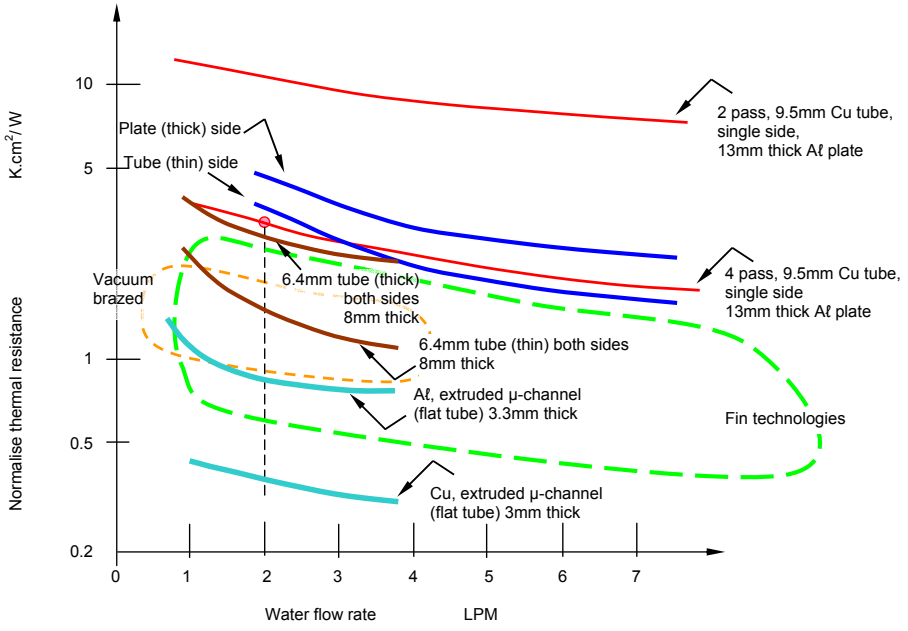


Fig. 5.48. *Performance comparison of various types of cold plates.*

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 5.21 and 5.25.) 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 5.25: 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.

$$T_{out} = T_{in} + \frac{P_d}{m \times c_p} = T_{in} + \frac{P_d}{\rho_\ell v \times c_p} \qquad (5.88)$$

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_ℓ = 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:

$$R_\theta = (T_{max} - T_{out})\frac{A}{P_d} \qquad (5.89)$$

where R_θ = thermal resistance, K/W per unit area
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 expressed 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, thence necessary plate thermal resistance.
Use $c_p = 4{,}184$ J/kgK and $\rho_f = 998$ kg/m^3

Solution

Given: 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 (5.88), the temperature change for 500W at a 2 lpm flow rate is less than 4°C. That is

$$T_{out} = T_{in} + \frac{P}{\rho v \times c_p}$$

$$= 20°C + \frac{500W}{998kg/m^3 \times \frac{2}{60} l/s \times 10^{-3} m^3/l \times 4{,}184 J/kgK} = 20°C + 3.6°C = 23.6°C$$

T_{out} is less than T_{max} so the required thermal resistance is determined by equation (5.89):

$$R_\theta = (T_{max} - T_{out})\frac{A}{P}$$

$$= (55°C - 23.6°C) \times \frac{50cm^2}{500W}$$

$$= 3.14°C.cm^2/W \text{ at } 2 \text{ lpm}$$

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.

♣

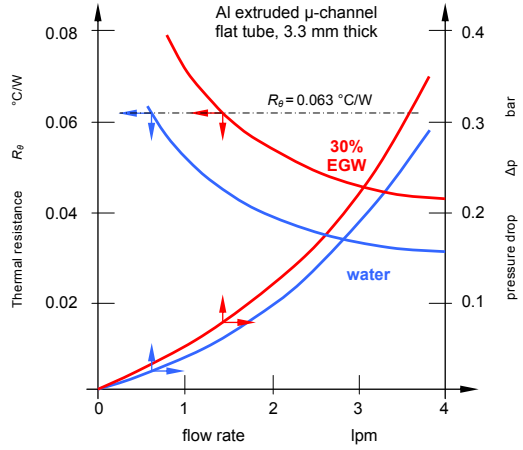


Figure 5.49. *Effects of flow rate and coolant variation on thermal resistance.*

In example 5.16 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 5.21. 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. (0.6 lpm and 0.024 bar if de-ionised water could be used).

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. With a device submerged in a pool of dielectric liquid (electrical non-conducting), 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 devices and its substrate surface 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.

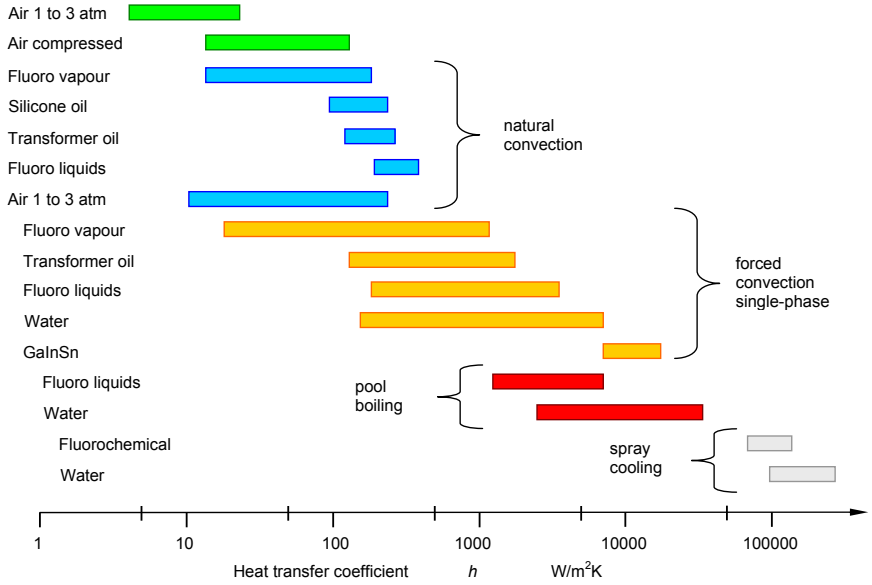


Figure 5.50. *Relative magnitude of heat transfer coefficients for various coolants and modes of convection. (see figure 5.19)*

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 5.26 are generally considered the most suitable liquids for direct immersion cooling, despite their poorer thermo-physical properties.

As shown in Table 5.26, 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 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 (heat of vaporization). 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 dielectric current will flow.

Table 5.26: 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	λ		W/m.K	0.0551	0.0545	0.057	0.613
Dynamic Viscosity	v	$\times 10^{-4}$	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 ambient 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.

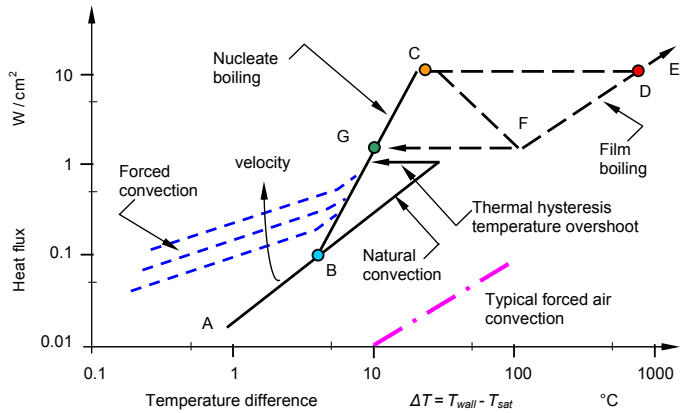


Figure 5.51: *Typical heat transfer regimes for immersion cooling with a fluorocarbon liquid.*

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

$$Q = C'_{sf} \times A \times (T_{wall} - T_{sat})^n \qquad (5.90)$$

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, which is poor. In most power electronic cooling applications, transition to film boiling will result in failure due to high junction 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 range 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. 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.

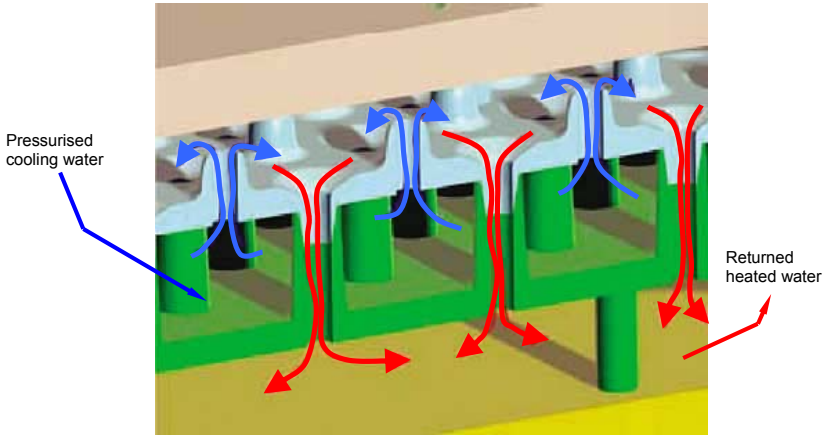


Figure 5.52. *Available multiple jet impingement liquid cooling.*

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 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 jet impingement. 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.

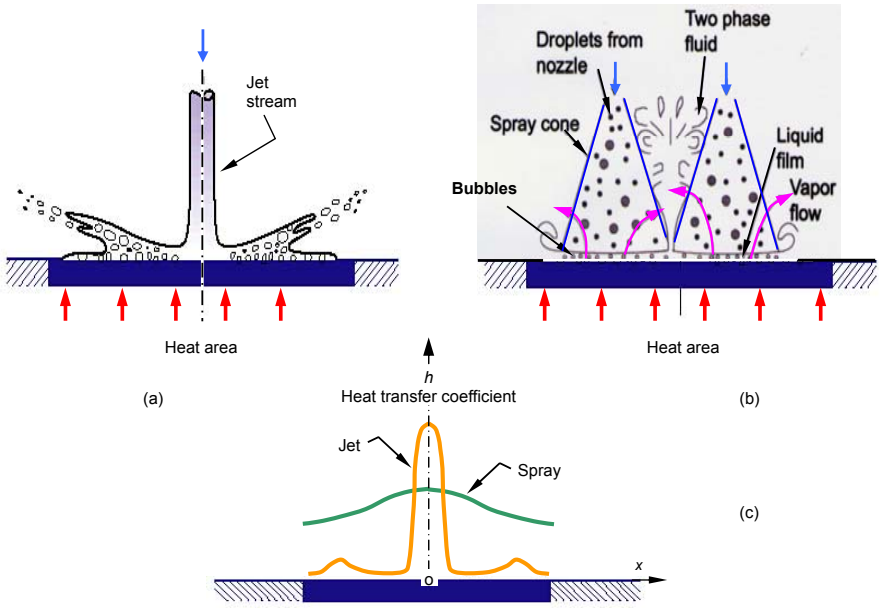


Figure 5.53. *Illustration of:*
(a) spray cooling; (b) jet impingement cooling; and (c) resultant heat transfer coefficient, h.

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, and 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.

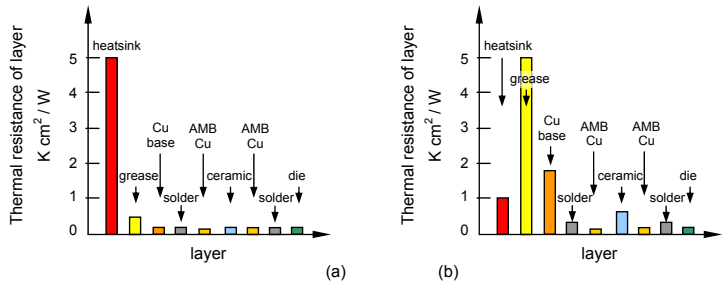


Figure 5.54. *Thermal resistivities of a typical power module with: (a) a conventional heatsink and (b) a microchannel heatsink, indicate that layers other than the heatsink can dominate the performance of microchannel designs.*

By using substrate level 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 with a 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.

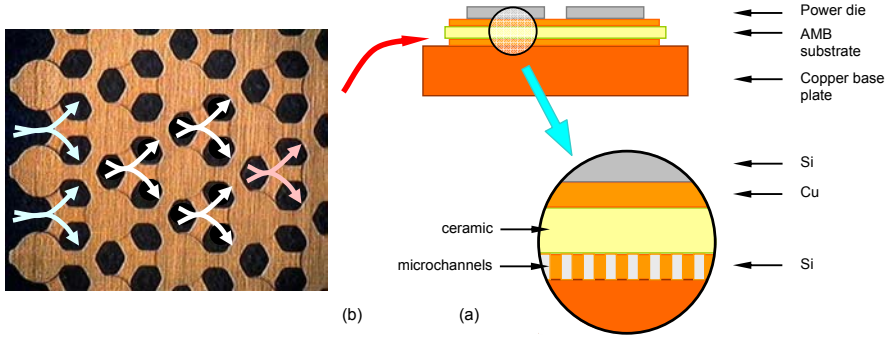


Figure 5.55. *Integrated microchannel heatsink design, where microchannels are: (a) fabricated directly into the bottom copper layer of the AMB substrate or (b) created between at least two thin (0.3mm) copper etched mesh layers, bonded by a melting process.*

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:

$$R_\theta = \frac{\Delta T}{P_d} = \frac{D_H}{k \times Nu} \qquad (5.91)$$

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
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:

$$\Delta P = \tfrac{1}{2}\rho v^2 f \frac{L}{D_H} \qquad (5.92)$$

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

$$Re = \frac{\rho_\ell \times v \times D_H}{\upsilon} \qquad (5.93)$$

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 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 lpm 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, due to 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 abates 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 a 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 lpm. 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 lpm at 30kPa.
It is 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 ½ lpm is possible. Local heat transfer coefficients may change appreciably over time leading to local temperature changes of 10 to 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 1000 W/cm^{2}K, 14 to 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, are 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 created 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 0.4 mlpm 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 5.27 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 lower thermal expansion coefficient compared to water and 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 (Lipowitz's alloy, eutectic alloy 50% Bi, 26.7% Pb, 13.3% Sn, and 10% Cd by weight) 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 coolant.

Table 5.27: Approximate thermal properties for liquid metals. Water is included for comparison

Metal	Melting point	Boiling point	Density	Specific heat	Conductivity λ	Viscosity ν	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 tables 5.26 and 5.27.
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 lpm 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 deficient (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.

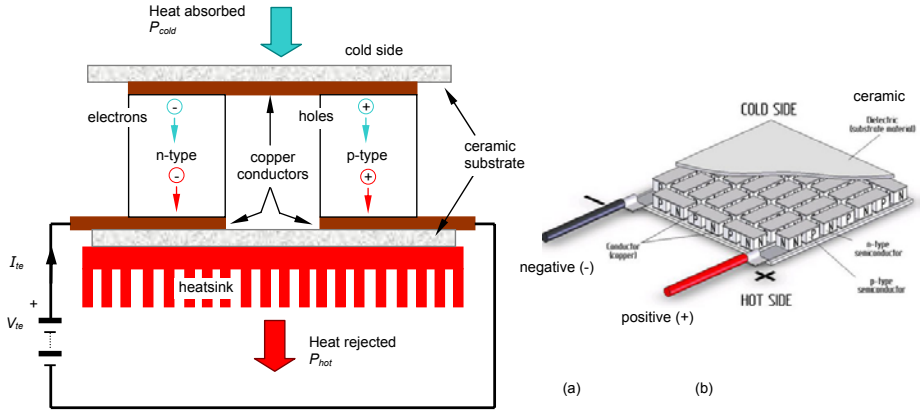


Figure 5.56. *The thermoelectric cooler: (a) principle and (b) module.*

Design for power electronic cooling 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, or the thermal load, in Watts, to be absorbed or removed at the cold surface of the TEC, P_{cold}.
The cold surface temperature is the desired temperature of the power electronic component 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 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 shows 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

$$CoP = \frac{\text{heat absorbed at the cold junction}}{\text{electrical dc input power}} = \frac{P_{cold}}{V_{te} \times I_{te}} = \frac{P_{cold}}{P_{tec}} \qquad (5.94)$$

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 allow 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.

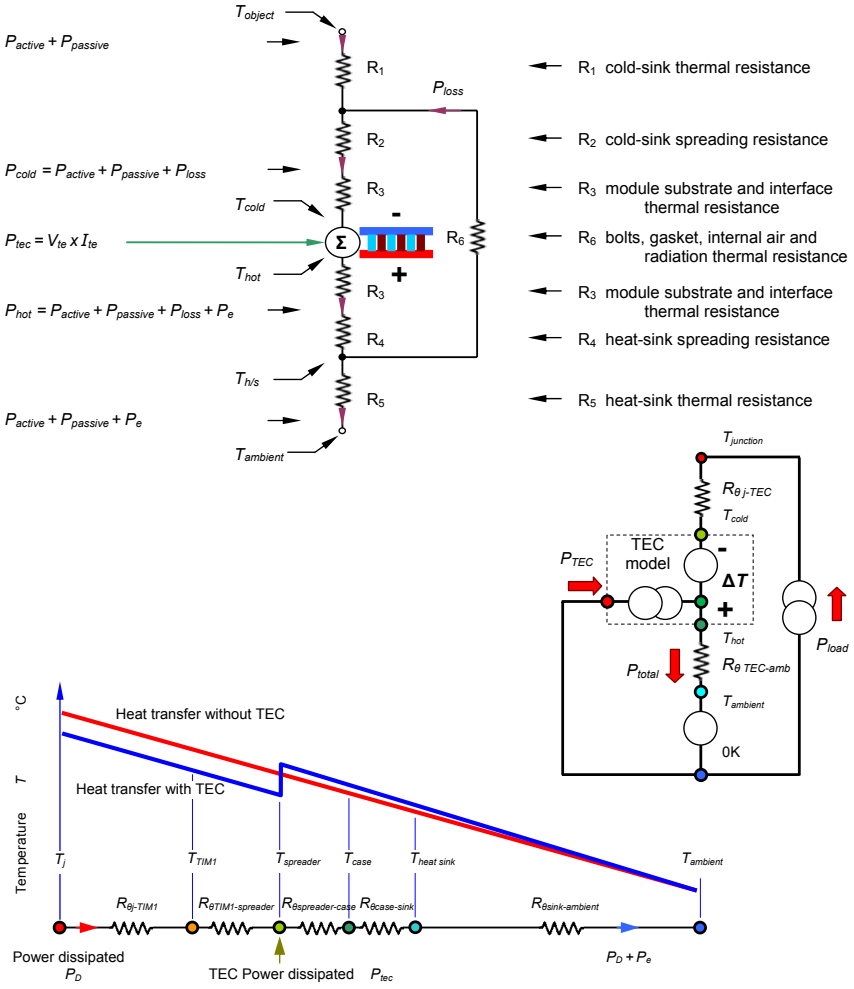


Figure 5.57. *The thermal resistance TEC cooling model.*

Example 5.17: ***Thermoelectric cooler design***

A semiconductor junction dissipating 100W has a steady-state thermal resistance to its case mounting of R_θ = 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} = 0W, determine:

i. the semiconductor junction temperature;
ii. the coefficient of performance CoP for the TE module; and
iii. heatsink requirement.

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 (5.94), the TE module coefficient of performance is

$$CoP = \frac{P_{cold}}{V_{te} \times I_{te}} = \frac{P_{cold}}{P_{tec}} = \frac{100}{74.7} = 1.39$$

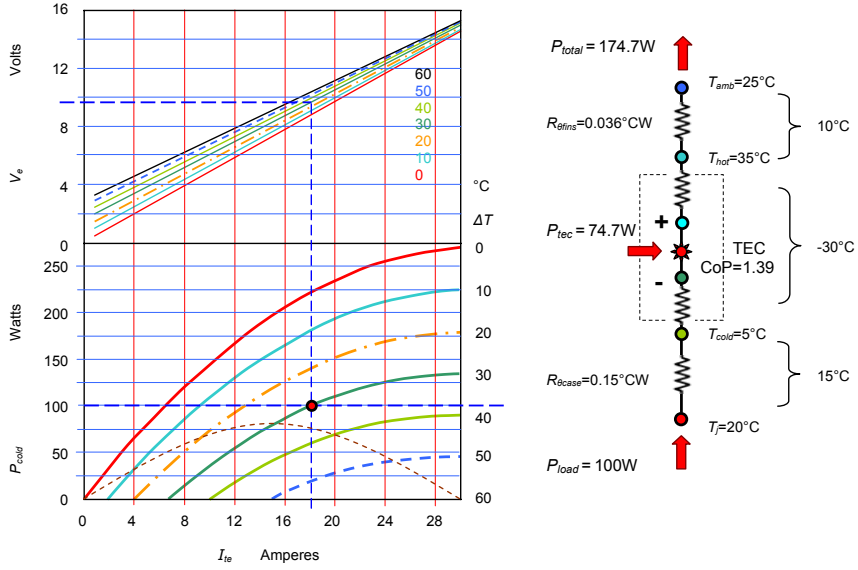


Figure 5.58. *Thermoelectric module characteristics,*
P_{max} = 270W, with 127 couples for a 35°C hot-side temperature T_{hot}, ΔT = 62°C, and I_{max} = 30A.

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

i. Using the TEC characteristics in figure 5.60, calculate the switch case temperature with and without a TEC, and the effective thermal resistance in each case.
ii. What is the maximum heatsink thermal resistance which when exceeded does not justify using TECs.

Solution

i. 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.

From an equivalent Ohms law calculation, the case temperature without a TEC is

$$T_c = T_{amb} + P_D \times \left(R_{\theta fins} + R_{\theta spread} + R'_{\theta TIM2} / A_{switch}\right)$$

$$T_c = 40°C + 120W \times (0.18K/W + 0.08K/W + 0.2cm^2K/W / 9cm^2) = 73.9°C.$$

This equates to a case to ambient thermal resistance of

$$R_{\theta c\text{-}amb} = \frac{(T_{case} - T_{amb})}{P} = \frac{(73.9°C - 40°C)}{120W} = 0.283°C/W.$$

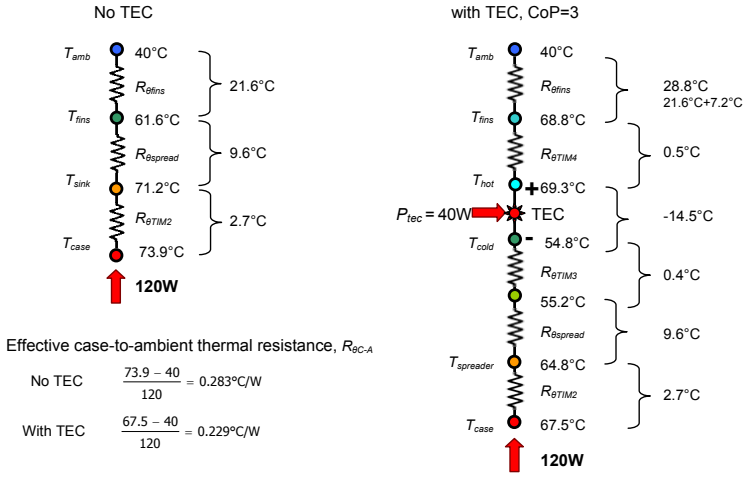


Figure 5.59. *Typical thermal resistor equivalent networks with and without a TEC.*

Although a TEC adds additional thermal interfaces, it interposes a negative ΔT in the thermal path. Also a TEC allows a simpler heat sink, one designed for a uniform heat input where spreading in the heat sink is not necessary.

With TECs, assume the four TECs operate 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 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, from equation (5.94), a *CoP* of 3 equates to a TEC input power of 40W and a optimal TEC ΔT of about -14.5°C from figure 5.60. This 40W must be dissipated by the heat sink in addition to the switch 120W heat load, thus raising the heat sink temperature by 40Wx0.18°C/W=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 (0.08°C/W assumed) 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.

$$T_c = T_{amb} + P_d \times \left(R_{\theta TIM2} + R_{\theta spread} + R_{\theta TIM3}\right) - T_{TEC} + \left(P_D + P_{TEC}\right) \times \left(R_{\theta TIM4} + R_{\theta fins}\right)$$

$$= 40°C + 120W \times (0.2 / 9 + 0.08 + 0.2 / 64) - 14.5°C + (120W + 40W) \times (0.2 / 64 + 0.18)$$

$$= 40°C + 12.7°C - 14.5°C + 29.3°C$$

$$T_c = 67.5°C$$

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 40W 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.

No gain in using a TEC approach results when the TEC thermal drop of 14.5°C is the same as the temperature rise – when the junction operating temperature is the same by both cooling methods.

ii. With a poor heat sink with a fin to ambient thermal resistance of 0.362°C/W instead of 0.18°C/W (with all other parameters maintained), the additional heat sink temperature rise due to the 40W TEC power increases from 7.2 to 14.5°C (0.362°C/W x 40W). The TEC gains become null and void.

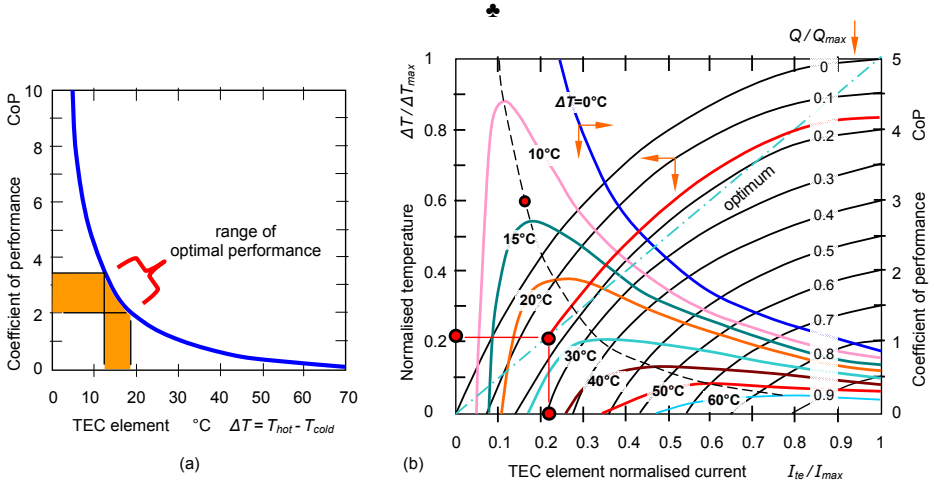


Figure 5.60. *Generic Bi_2Te_3 thermoelectric characteristics:*
(a) theoretical optimum (maximum) CoP vs ΔT and
(b) CoP as a function of current and generic thermoelectric performance curves.

Using examples 5.17 and 5.18 as a benchmark cases, 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 TEC negative ΔT. 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}.
- 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..
- 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 TE enhanced heat sink example 5.18, four TECs and a large spreader were utilized, where each TEC pumped 30W giving the required total of 120W. The generalized thermoelectric performance curve given in Figure 5.60b gives an indication of the type of TEC required. For a *CoP* of 3 operation, the optimum ΔT is 14.5°C (figure 5.60a). For a typical TEC with a ΔT_{max} of about 67°C, the ratio of $\Delta T/\Delta T_{max}$ is about 0.21. From Figure 5.60b, the curve for optimum (lowest power) operation, P/P_{max} is approximately 0.17. Therefore, for a 30W TEC heat load, the P_{max} for each TEC should be approximately 30W/0.17 = 176W.

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 120W/0.17=700W, which exceeds available limits of about 550W (for ΔT=0).

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.

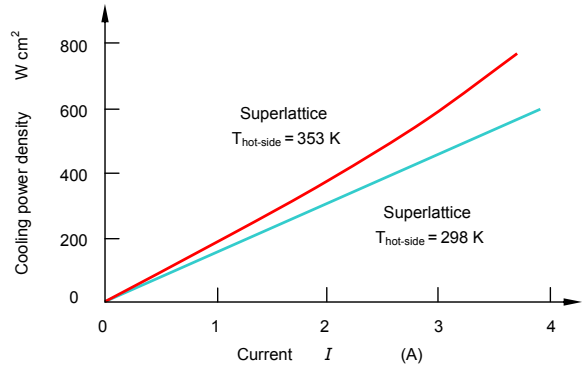


Figure 5.61. *Estimated power density for superlattice devices as a function of current.*

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 for electronics thermal management are 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 interval, 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). Another specification is duration index, *DI*, which is a comparison of time of the PCM at constant temperature during the phase change.

$$DI = \frac{h_r \times \rho}{\Delta T} \tag{5.95}$$

The list of candidates is restricted when issues such as material compatibility, electrical conductivity, toxicity, availability of thermal property data, and cost are considered.

Heat of fusion (solid to liquid)

Desirable characteristics of a solid-liquid PCM include high heat of fusion per volume (*DI*), congruent melting and freezing characteristics, high thermal conductivity, minimal supercooling, and low thermal expansion. Table 5.28 lists four categories of solid-liquid PCMs with melting temperatures in the range of 40° to 120°C, with water characteristics provide as a reference.
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 warning. 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.

Table 5.28: General solid-liquid PCM thermo-physical characteristics

property		Inorganic water	Organic paraffin	Organic non-paraffin	Inorganic salt hydrate	Inorganic metal eutectic
Phase change temperature	°C	0	37	2 - 164	7 - 117	< 0
Latent heat of fusion h_f	kJ / kg	334	230 - 290	120 - 240	170 - 340	30 - 90
Latent heat of vap h_{fv}	kJ / m^3 x 10^2	22.7	190 - 240	140 - 430	250 - 660	300 - 800
Density ρ_f	kg / m^3	1000	≈ 810	900 -1800	900 - 2200	≈ 8000
Thermal conductivity λ	W / mK	0.6	≈ 0.25	≈ 0.2	0.6 - 1.2	≈ 20
Duration Index *DI*	J/ m^{3}K	13.32	2.96	5.44		
Thermal expansion		low	high	moderate	low	low
Congruent melt		yes	yes	some do	most do not	yes
Supercool		yes	no	no	most do	no
Corrosion		low	low	some are	highly	some are
Toxicity		no	no	some are	highly	some are
Advantages			Simple to use No nucleating agent		Cheap Non-flammable	
Disadvantages			More expensive Some combustible		Careful preparation Need stabilising additives	

Table 5.29: Heat of vaporization at standard temperature and pressure (see Table 5.23)

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
Glycerine	290	974
Mercury	347	295

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 forces in the liquid state. The energy released when a vapour is condensed to a liquid is 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 a solid to liquid phase change, the volume change associated with a 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 5.29.

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/

http://pcmproducts.net/

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 5.30.

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

$$R_a = \frac{1}{n}\sum_{i=1}^{n}|y_i|$$

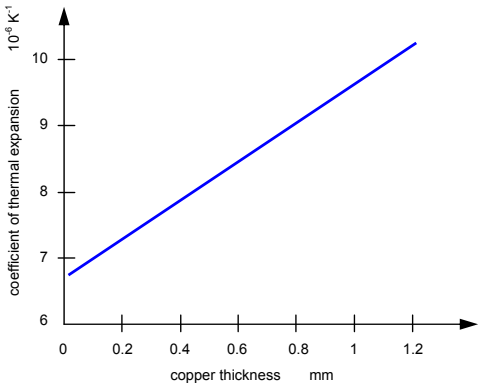
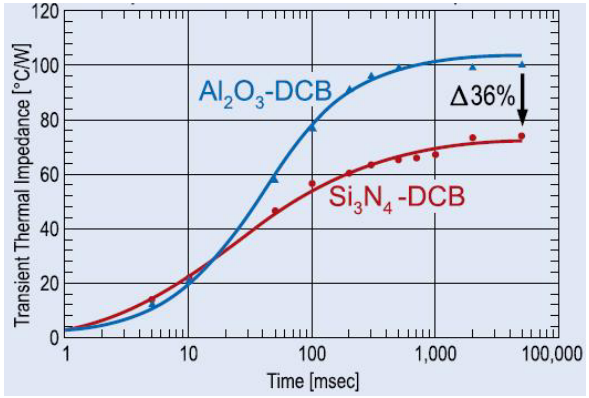


Figure 5.62. *Thermal expansion dependence on copper base plate thickness.*

Table 5.30: Substrates properties

Material		Aluminium nitride	Aluminium oxide	Beryllium oxide	Silicon carbide
symbol		$A\ell_3N_4$	$A\ell_2O_3$		
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 $A\ell_3N_4$ 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). Thermal impedance is compared in the following figure.



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, is shown in figure 5.62.

5.24 Appendix: Properties of substrate and module materials

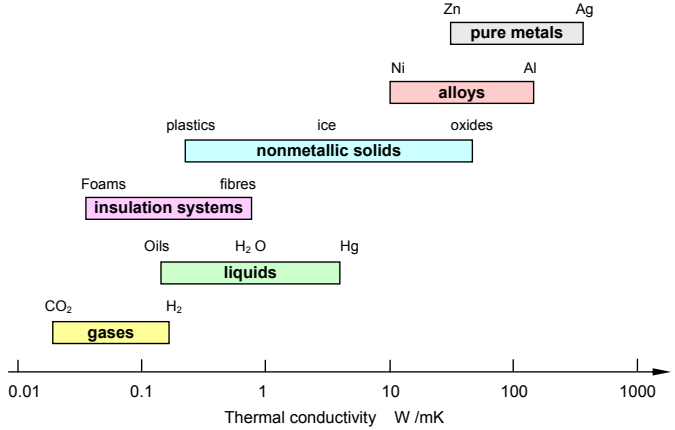


Table 5.31: Substrate characteristics

Material	Al_2O_3, BeO, AlN, Quartz, Silicon, Sapphire, Ferrite
Surface Finish	Al_2O_3 As Fired, 0.05µm maximum;
Polished	to 0.12µm
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

Table 5.32: 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 5.33: 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	$-1300+j1.3\times10^{14}$		220	23.8		metal

Table 5.34: 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	$A\ell SiC$	170		3000	7	
polyimide		0.2	1100	1400		
dielectric layer		0.3	1400	1120		
encapsulation		0.5		2000		
aluminium	$A\ell$	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				

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 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 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 that under a shallow incident angle, the emissivity of metals tends to one and insulators tend to zero.

A complication is, the emissivity of many materials of interest is strongly dependent on the wavelength. That is, 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 casings. 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 simulation software that support radiation heat transfer, assume the following simplifications:

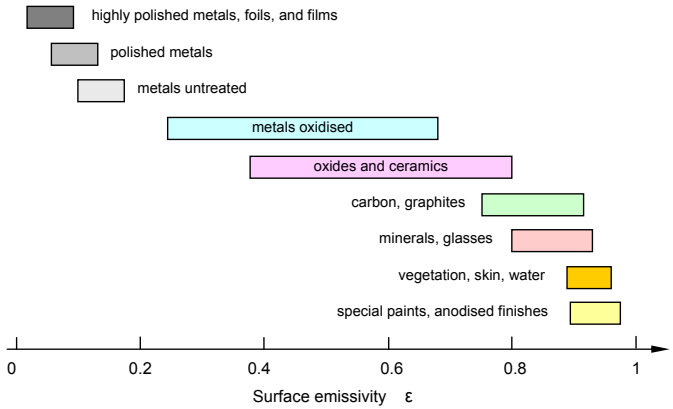
- surfaces diffuse, grey (sometimes specula [mirror-like]), opaque
- surfaces isothermal
- surfaces uniformly irradiated
- medium is transparent for all relevant wavelengths

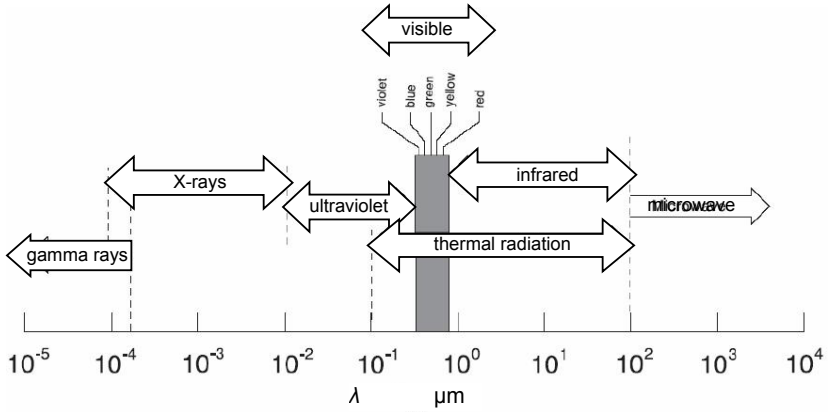
In summary, differences exist between experimental data and simulation results when radiation is a major factor, even when all other data are known to within 1%.

Table 5.35 shows typical material normal total emissivity values @ 20°C.

Table 5.35: 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 (5.96) and (5.97), as applicable, in conjunction with Table 5.36.

$$h = k_h \left(\frac{\Delta T}{D} \right)^{1/4} \tag{5.96}$$

$$h = k_h \left(\frac{\Delta T}{L} \right)^{1/4} \tag{5.97}$$

where ΔT is temperature difference, K
L is length, m
D is diameter, m

Table 5.36: Heat transfer coefficient constant, k_h

Geometry	Dependant geometrical parameter		constant k_h	equation
Vertical cylinder or plate	Vertical height *L*		1.42	(5.97)
Horizontal cylinder	Diameter *D*		1.32	(5.96)
Horizontal plate	$L = \frac{4 \times area}{perimeter}$	Upper surface hot	1.32	(5.97)
		Lower surface hot	0.59	
Vertical populated PCB	Vertical height *L*		2.44	(5.97)
Sphere	Diameter *D*		1.92	(5.96)

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 5.37 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 5.37: Ampcapacity

number of 6mm x 100mm 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, typically 300µm, are applied to the top and bottom surfaces of the isolation substrate material (0.25mm to 0.85mm thick, typically 0.5mm) 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 (about 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.

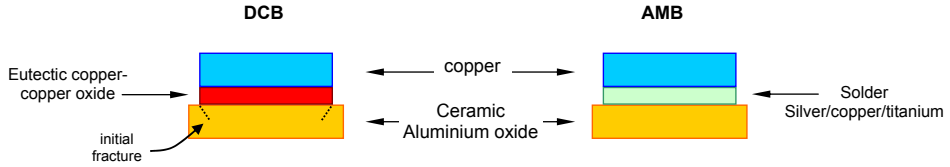


Figure 5.63. *Direct copper bonding, DCB and active metal brazing, AMB.*

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. AMB offers higher partial discharge levels than DCB.

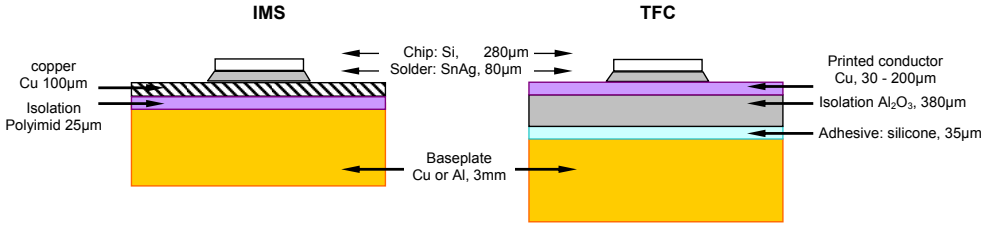
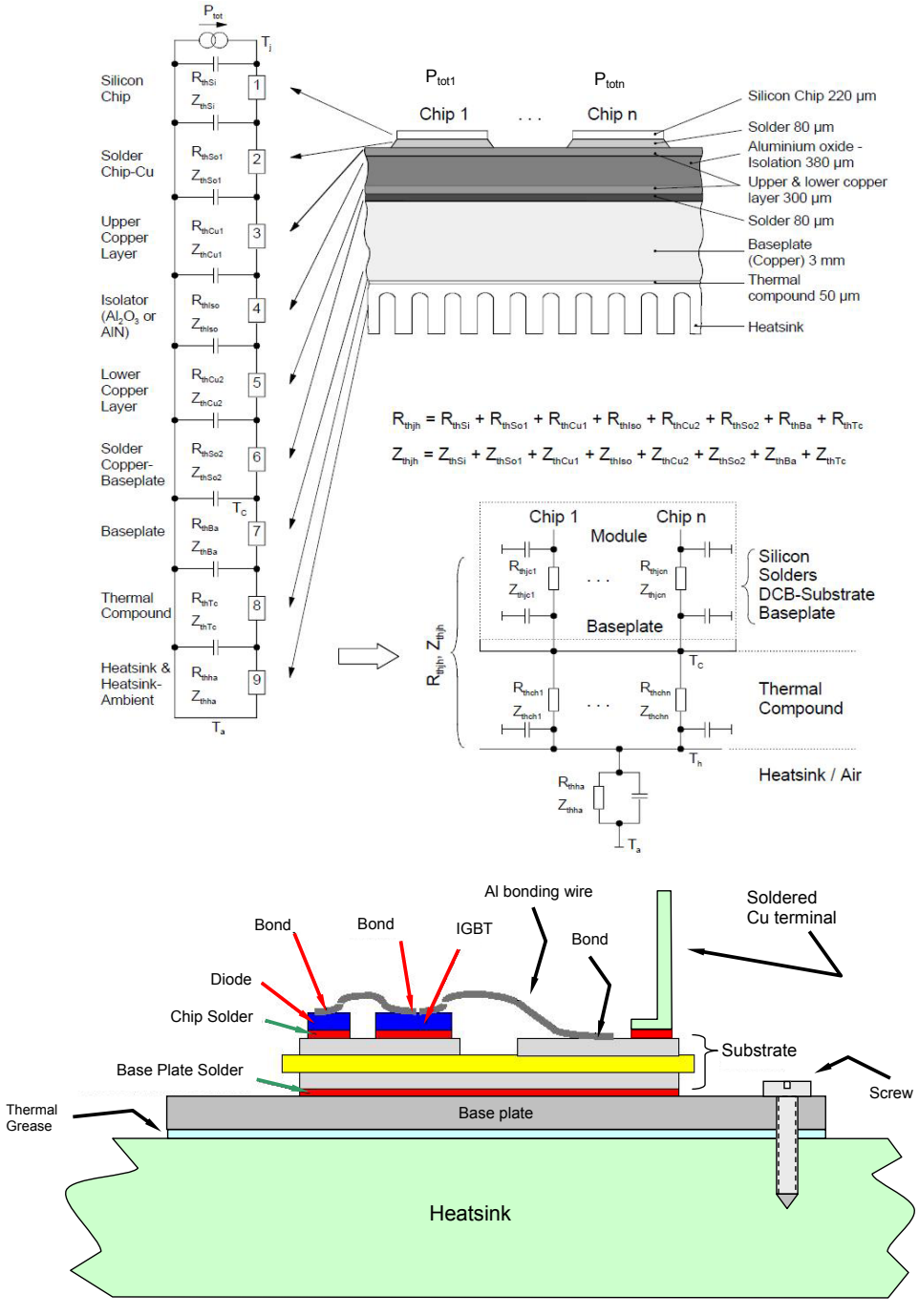


Figure 5.64. *Basic module structure of: (a) an IMS power module and (b) a TFC power module.*

*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 onto 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 onto 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.



$$R_{thjh} = R_{thSi} + R_{thSo1} + R_{thCu1} + R_{thIso} + R_{thCu2} + R_{thSo2} + R_{thBa} + R_{thTc}$$

$$Z_{thjh} = Z_{thSi} + Z_{thSo1} + Z_{thCu1} + Z_{thIso} + Z_{thCu2} + Z_{thSo2} + Z_{thBa} + Z_{thTc}$$

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 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 a metal matrix composite, MMC, base-plate allows operation to temperatures in excess of 400°C.

The baseplate of the power module can utilize 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 a power module.

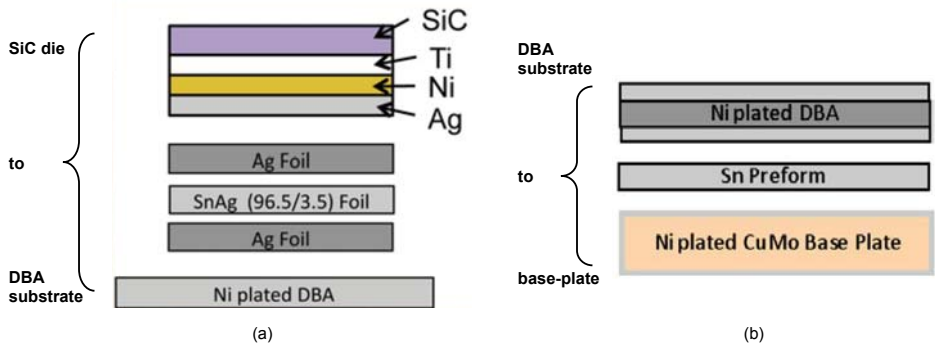


Figure 5.66. *Cross-section of various layers in the lead-free 400°C high temperature SiC power module: (a) Ag-Sn TLP die attach and (b) Ni-Sn TLP substrate attach.*

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

Table 5.38 Heat pipe working fluids

medium	Melting point	Boiling point @ atmospheric pressure	Useful range
	°C	°C	°C
Helium	-271	-261	-271 to -269
Nitrogen	-210	-196	-203 to -160
Ammonia	-78	-33	-60 to 100
Pentane	-130	28	-20 to 120
Acetone	-95	57	0 to 120
Methanol	-98	64	10 to 130
Flutec PP"	-50	76	10 to 160
Ethanol	-112	78	0 to 120
Heptane	-90	98	0 to 130
Water	0	100	0 to 150
Toluene	-95	110	30 to 200
Flutec PP9	-70	160	50 to 200
Thermex	12	257	0 to 225
Mercury	-39	361	150 to 350
Caesium	29	670	250 to 650
Potassium	62	774	500 to 1000
Sodium	98	892	600 to 1200
Lithium	179	1340	1000 to 1800
Silver	960	2212	1800 to 2300

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]

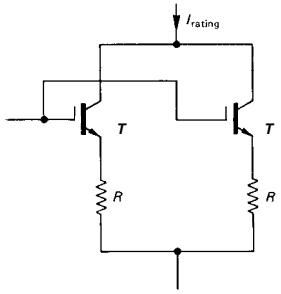


Figure 5.67. *Problem 5.2.*

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.

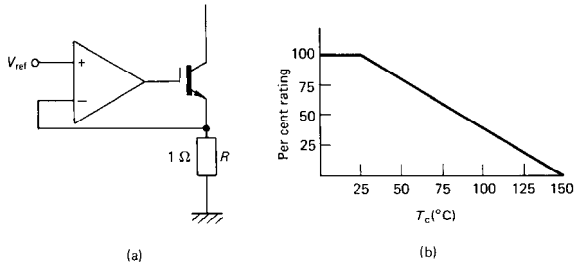


Figure 5.68. *Problem 5.3.*

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]

1	Explain the difference between fans, blowers, and compressors.
2	Which fan is used for moving flows against relatively low pressures? (a) radial fan (b) backward inclined fan (c) forward curved fan (d) axial fan
3	If efficiency is the main consideration, which fan is selected? (a) radial fan (b) backward inclined fan (c) forward curved fan (d) axial fan
4	Which fan is ideally suited for dusty conditions? (a) radial fan (b) backward inclined fan (c) forward curved fan (d) axial fan
5	System resistance refers to (a) static pressure (b) velocity pressure (c) total pressure (d) differential pressure
6	System resistance varies as (a) square of flow rate (b) cube of flow rate (c) proportional to the square root (d) directly with flow rate
7	The intersection of the system characteristic curve and the fan operating curve is called (a) design point (b) operating point (c) free flow point (d) shut off point
8	Varying the RPM of a fan by 10% varies the pressure by (a) 19% (b) 29% (c) 10% (d) no change
9	Varying the RPM of a fan by 10% varies the flow by (a) 27% (b) 20% (c) 30% (d) no change
10	Varying the RPM of a fan by 10% varies the power by (a) 27% (b) 37% (c) 10% (d) no change
11	Explain the factors which change the system resistance.
12	What are the affinity laws as applied to centrifugal fans?