Thermal Design For Power Electronics

by Dennis L Feucht

You don't see too many mechanical engineers designing electronic controls. Likewise, the tendency for electronics engineers to neglect the thermal design of what are largely electronic subsystems can lead to trouble. This article presents thermal design considerations for typical power electronics design. The whole subject will not fit into one article, but I will try to capture the essence of it, and include some useful details filtered out of the heat-transfer literature.

Conductive Heat Transfer

Heat transfer can occur by three mechanisms: conduction, convection, and radiation. For electronic components, conduction occurs through metal, ceramic, or other materials, and is characterized by a parameter that corresponds to electrical conductivity: thermal conductivity, k_θ. With a formula analogous to the geometric equation for electrical conductance, thermal conductance is:

$$G_\theta = \frac{k_\theta \cdot l}{A}$$

where, l is the thermal path length through the material of conductivity k_θ, and A is the cross-sectional area of the path. Thermal resistance, $R_\theta = 1/G_\theta$. For air, thermal conductivity,

$$k_\theta\,(\text{air}) \cong 2\ \text{W/mm}\cdot\text{K}$$

Thermal conductivities for some commonly used materials in electronics are given in the following table. Note that while copper is excellent, diamond is extraordinary. Once diamond-film thermal pad technology is worked out, it will shift the thermal "bottleneck" from the insulating pads to the heat sinks and device packages. Ideally, circuit boards would be made out of diamond, for the low thermal conductivity of fiberglass (FR4) etched circuit-boards (ECBs) is a major impediment to heat transfer for surface-mounted power devices. And stagnant air, where the heat eventually terminates its flow, is the worst conductor of all. It is therefore good that there is also convection!

Material	W/m·°C	Material	W/m·°C
Silicon	150	Solder	40
Diamond	1500	Copper	459
Aluminum	240	Silver	429
Ceramic (AlN)	150	Alumina (Al_2O_3)	21
FR4 ECB	0.2	Iron	80
G10 ECB	0.32	Steel,1008	44
Air	0.026	BeO	240
Ferrite	3.4	Manganin	23

The Ohm's Law of thermal circuits is:

$$\Delta T = R_\theta \cdot P$$

where, ΔT is the temperature difference across the thermal resistance with power, P, dissipated in it. This is often used in the form:

$$T = R\theta \cdot P + TA$$

where, TA is the ambient temperature, the temperature of the surrounding air, where the heat ultimately ends up.

For a power-transistor heatsink thermal design, the equation can be elaborated accordingly:

$$T_J = (R_{JC} + R_{CS} + R_{SA}) \cdot P + T_A$$

In series are thermal resistances from the semiconductor junction to the case (transistor package), case to insulating pad (if any), pad to heatsink, and heatsink to ambient. R_{JC} is given in the semiconductor device specifications, and R_{CS} in thermal-pad specifications.

Heatsink Convective Heat Transfer

The last thermal resistance, R_{SA}, from heatsink to ambient, is the hardest to determine. Heatsink manufacturers also give specifications for R_θ of heatsinks based on volumetric air flow and temperature. For extruded heatsinks length is also a parameter.

Extruded heatsinks come out of the die in the length dimension of the fins and are cut to a specified fin length. The length dimension must be mounted vertically for convection airflow through the fins. As air is heated by the sink its density decreases, and it rises through the air columns between the heatsink fins. The heat of the sink powers air flow; this is referred to as natural convection. When a fluid flow is constrained by a solid surface, whether it be the skin of an airplane wing, the inside of a pipe, or a heatsink fin surface, the drag at the surface is relatively large, and forms a slow-flow region near the surface called a boundary layer. Air along a typical heatsink fin forms a boundary layer of about 3 mm in thickness, with substantial airflow beginning at least 3 mm away from the surface. This places a lower limit on fin spacing for natural convection cooling. The rule of thumb for fin spacing for maximum convective heat transfer is about 0.3 the fin length. Optimal fin height (not length) to thickness is 100 for natural convection and 40 for forced convection, caused by an external air mover.

The geometric formula for thermal conductance for convective heat transfer is:

$$G_\theta = \frac{1}{h \cdot A}$$

where, h is the convection coefficient, in $W/mm^2 \cdot K$. For still air,

$$h \text{ (still air)} \cong 25\ \mu W/mm^2 \cdot K$$

with emphasis upon the "approximately" in the sign; this value is only a rough approximation and depends upon too much that varies from one particular instance to the next to regard it with any firmness of conviction.

The general equation for convective heat transfer is:

$$P = G_\theta \cdot \Delta T = h \cdot A \cdot \Delta T$$

The more general formula for h (from the 1960s transfer and rate processes book by Welty, Wicks, and Wilson of Oregon State University, page 341) is:

$$h \cong 1.646 \times 10^{-6} \cdot \left(\frac{\Delta T}{L}\right)^{1/4}$$

where, ΔT is in °C or °K, L in mm, and h in $W/mm^2 \cdot K$. This applies to a vertical plate of height L in air, with laminar (not turbulent) flow.

With the notorious difficulty of obtaining accurate thermal parameters, what is an engineer to do? Consider two options. The first is to use the heatsink data from the heatsink manufacturers. After a few thermal design exercises, one develops a semi-quantitative sense for how much heatsink, under the given conditions, is about right. The second approach is to use a finite-element-analysis (FEA) heat transfer program. This will put you solidly in the domain of the chemical and mechanical engineers. For unusual thermal conditions this might be advisable. But for almost all power electronics design heatsink catalogs contain sufficient information.

A newer form of heat transfer for electronics is the use of liquids such as water. Liquid cooling usually becomes relevant for systems handling over 10 kW of power and can reduce thermal resistance for the same volume by 50 times. If phase changes are involved, as occur in heat pipes, then the resistance can be reduced by 500 times -- clearly an attractive option for high-power design. However, despite the conductivity advantages the overhead of pipes and pumps precludes its use in smaller systems.

Radiation

Radiative heat transfer is usually negligible for semiconductor electronics because radiated heat varies by T^4, and is still in the flat part of the curve at temperatures semiconductors can handle. (This might change somewhat for SiC devices.) The radiated-power formula is:

$$P = \sigma \cdot A \cdot T^4$$

where, the Steffan-Boltzmann constant is $\sigma = 5.675 \times 10^{-15}$ W/mm^2·K^4 and temperature, T is in Kelvin. A is the radiating surface area. Radiated power amounts to a measly 136 μW/mm^2 at 100°C. A TO-220 package is roughly 100 mm^2, and at 150°C ≅ 423°K, then P ≅ 18 mW, rather negligible for such an overheated silicon device. We can forget about radiative heat transfer for silicon. Aluminum heat sinks can optionally be anodized black to improve radiation, but it will not help silicon electronics much. Suppose an SiC diode in a TO-220 package can operate at a maximum junction temperature of 500°C. Then its radiated power is up to 206 mW -- still not much to consider, but for a large extruded heatsink at this temperature, the anodizing would be worth considering.

Thermal Dynamics

When power transistors are pulsed, as is often the case nowadays, the thermal "circuit" dynamics enter into the design. By "thermal dynamics" I don't mean thermodynamics, which is a profound but misnamed science; it is really thermo-quasi-statics and is not related to dynamic heat transfer as such. Thermodynamics is based on states, like bubbles in a digital state diagram, and processes, like the directed arcs connecting the bubbles, largely irrespective of the processes that change the system from one state to another. Heat transfer is about how the (thermal) state changes occur and involves rates; thermodynamics does not. (You won't find the time variable much in thermodynamics books, except to calculate average power.)

Thermal "circuits" are not highly analogous to electronic circuits because they are distributed, and are more like delay lines than lumped-parameter circuits. Consequently the same linear dynamics equations and their parameters, such as time constants, do not show up except as approximations. (Thermal distortion in amplifier circuits is typically compensated with networks with several time constants, just as delay lines are.) We are back to approximations.

For pulsed power transistors with a duty ratio of D (that is, they are on and sourcing heat a fraction D of the time) semiconductor companies offer the approximate dynamic formula:

$$Z_\theta \cong D \cdot R_\theta \cdot (1 - e^{-T_s/\tau_\theta})$$

where, T_s is the switching period. The linearized thermal circuit model (using an electrical analog) is a parallel RC, with:

$$R = D \cdot R_\theta, \quad C = \tau_\theta / R_\theta$$

For a TO-220 package, $\tau_\theta \cong 50$ ms. The thermal impedance varies with D. Consequently, an interaction of note in the design of power circuits is that by minimizing D this minimizes heatsink (or fan) size and cost. Heat diffuses out from the junction into the case of a power device and if the thermally-conductive volume around the thermal source is depleted of heat (phonons), much more can be dumped into it for a little while. Analogously, the flow rate into an empty tank can be immense for a while without overflowing it. In addition, low D reduces power dissipation in the switch by $\sqrt{D}$, but increases the peak-to-average ratio of currents or voltages, causing device ratings to accommodate the peaks. The lesson is: if D can be reduced, it benefits the thermal design superlinearly.

Closure

Electronic design becomes briefly interdisciplinary when heat removal is considered. Most of the heat is generated by switches, with switching and conductive loss. Optimizing the device for minimum power loss and maximum power handling involves the electrical impedance at which power switching occurs. Conduction loss goes up superlinearly with MOSFET breakdown voltage:

$$r_{DS}(\text{on}) \propto \text{BV}_{\text{DSS}}^{2.5}$$

To minimize conduction loss:

$$P_c = \tilde{i}^2 \cdot r_{DS}(\text{on})$$

Minimization of rms current reduces it by the square, but for the same power, voltage must be increased and r_{DS}(on) increases by the 2.5 power. An optimum impedance is consequently a low impedance for the MOSFET, though not necessarily for the rest of the circuit.