AND8223/D

Predicting Thermal Runaway

Prepared by: Roger Paul Stout, PE
ON Semiconductor

ON Semiconductor®

http://onsemi.com

APPLICATION NOTE

Abstract

This monograph explains what thermal runaway is, some settings in which is it of concern (and some in which it is not), and how to predict it. An archetypal "power law" device will be used to develop certain closed−form relationships describing thermal runaway. "Perfect" thermal runaway will be defined to occur in a system whose operating point is located precisely at the point of runaway. Consequently, since real systems should be designed well away from this point, thermal runaway will be seen to have two natural manifestations. In the first, the thermal resistance of the cooling system is fixed, and ambient (or more generally, thermal ground) is increased until runaway occurs. In the second, thermal ground is fixed, and the thermal resistance of the cooling system is increased, again until runaway occurs. In many applications, although the former limit occurs at a higher temperature than the latter, it will be found to be a more realistic constraint on the system design, because the cooling system resistance will be far from the limit implied by the latter. However, an interesting and surprising aspect of the latter limit is the constant temperature offset between thermal ground and thermal runaway temperature, dependent only on the strength of the power law itself, and independent (as it turns out) of the thermal resistance of the cooling system.

Glossary of Symbols

I, I_o	current (A)
k, k_1, k_2	slope of system line in non−dimensional coordinate system
Q, Q_o, Q_i	device power dissipation (W)
q	non−dimensional power
T	temperature (°C)
T_i	any reference temperature (°C)
T_J	junction temperature (°C)
T_a, T_c	thermal ground, ambient or coldplate (°C)
T_R, T_{R1}, T_{R2}	thermal runaway temperature for particular system resistance (°C)
T_x	thermal ground for runaway based on θ_{Jx1} or (°C)
T_y	thermal ground for runaway based on θ_{Jx2} or (°C)
V, V_R, V_o	voltage (V)
z	non−dimensional junction temperature
θ_{Jx}, θ_{Jx1}, θ_{Jx2}	steady state thermal resistance of system (°C/W)
λ	"strength" of power law (°C)

INTRODUCTION AND BACKGROUND

Thermal runaway may occur in a semiconductor packaging application when the power dissipation of the device in question increases as a function of temperature. In more particular, it describes the situation when no nominally steady−state operating point of the device, under the influence of the specific thermal system, can be established. Ordinarily, of course, a device that dissipates a fixed amount of power can always achieve a steady state operating condition, though the specific junction temperature attained may fall beyond recommended limits. If the thermal system around the device is characterized as having a steady state thermal resistance, then this equilibrium (or steady state) condition may be described as follows:

$$T_J = Q \cdot \theta_{Jx} + T_x \qquad \text{(eq. 1)}$$

For the present purpose, it is crucial to use a fixed reference temperature in the model, that is, the temperature that provides the thermal "ground" for the system. Thus, in an air cooled system, we would speak of "x" as being the ambient air temperature, and have, for example, T_a. In a water cooled system where the device of interest is mounted securely to the water−cooled block, it would be more appropriate to refer to the "coldplate" temperature, T_c, as the reference (or if the system is very efficient and the thermal resistance from the "case" of the device to the coolant is negligible in comparison to the device's internal resistance from junction to case, perhaps T_c is simply the case temperature).

From Equation 1, it may be seen that a small perturbation in power will result in a small perturbation in the junction temperature. If the power level briefly rises above the equilibrium value and then returns, the equilibrium temperature will eventually be restored. Likewise, if the power falls and returns, the temperature will drop and then return. This is because we have defined the system to be capable (at steady state) of dissipating exactly the amount of power needed to achieve the original steady state junction temperature. Indeed, solving Equation 1 for power, given the other values, we have:

$$Q = \frac{T_J - T_X}{\theta_{JX}} \qquad \text{(eq. 2)}$$

Also note that the rate of change of system power dissipation, with respect to changes in junction temperature, is given by:

$$\frac{dQ}{dT} = \frac{1}{\theta_{JX}} \qquad \text{(eq. 3)}$$

From Equations 2 or 3, then, we see that for a small increase in T_J, the system (described by θ_{JX}) can dissipate slightly more power than the original Q.

Straight−Line Devices and Systems

Identification of the device/system operating point may be illustrated graphically. In Figure 1, we begin with a device that produces a fixed amount of power regardless of its temperature. With power as the vertical axis, and temperature as the horizontal axis, this means we have a "device" line that is horizontal. The "system" line is a straight line of slope $\frac{1}{\theta_{JX}}$ (Equation 3), intersecting the horizontal axis at thermal ground, T_X. Where the device line and system line cross is the nominal operating point (T_J, Q). Wherever the system line is higher than the device line, more power is leaving the system than is being introduced by the device, so the system cools. Conversely, wherever the device line is above the system line, the system heats up. Thus we see that to the right of the equilibrium operating point, more power can be dissipated by the system than the device is producing, and to the left, more power is coming in than may be dissipated. In either case, one might picture the resulting imbalance in power as a *restoring* force that causes the junction temperature to move toward the operating point.

But what if the power dissipation is a function of temperature? Clearly, as seen in Figure 2, if the power *falls* as temperature increases (and rises as temperature decreases), the system still will experience tendencies to restore the nominal equilibrium value in the face of small perturbations. Note that it doesn't matter how steeply the power falls with temperature, the system line will always remain above the device line for temperatures above the operating point, and below it to the left of the operating point. Hence all such systems are stable. (The possibility of oscillatory behavior is not excluded, but is beyond the scope of this discussion.)

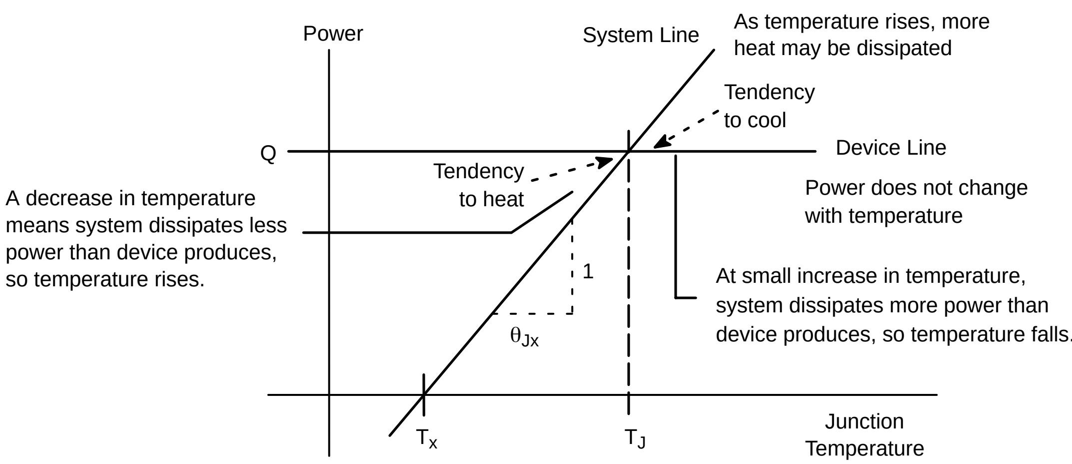

Figure 1. Operating Point of Thermal System with Temperature−Independent Power

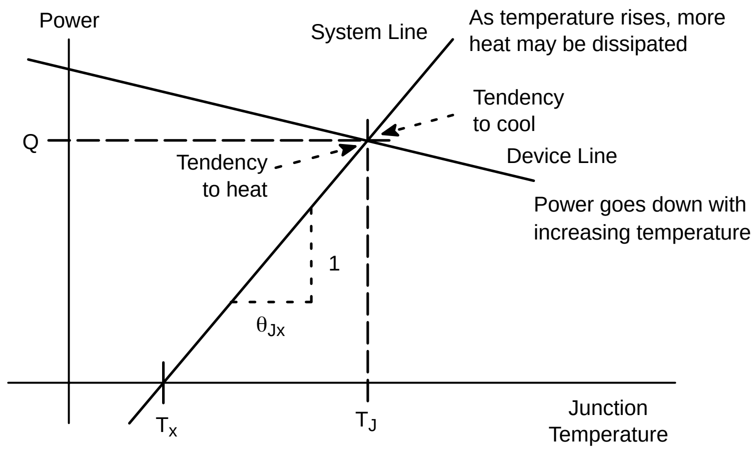

Figure 2. Operating Point of Thermal System where Power Decreases with Temperature

On the other hand, if the power increases with temperature, things get more interesting. As shown in Figure 3, power *may* increase with temperature without causing any particular difficulty. Indeed, if the device line has a constant slope, the device and system lines will always intersect at a unique point, and if (as shown in Figure 3) the slope of the device line is smaller than that of the system line, the intersection continues to represent a stable operating point. (If the lines happen to be parallel, then there is either no valid operating point at all, or the operating point is undefined. What happens in this system if the power or temperature makes a brief excursion away from equilibrium? Suppose the temperature goes up – according the device curve, it will now dissipate more power; but according to the system curve, the system can dissipate even *more* power. Hence (as with the previously considered situations), the system will tend to drive itself back towards the original operating point − likewise with small downward perturbations.

Now consider Figure 4. Here, the two straight lines still have a unique intersection, so it would seem that the system should have a valid operating point. Not so, unfortunately. Again, see what happens if there is a small perturbation in temperature. According to the system line, a small increase in temperature results in a small increase in the ability of the system to dissipate power. According to the device line, however, that same small perturbation in temperature results in an even larger increase in power dissipation. Hence the system cannot dissipate that much power, even at the increased temperature, so the temperature will rise even more, the dissipation will increase even more, and so on, until catastrophe results. (It may also be interesting to consider a negative perturbation: If the temperature drops slightly, the system line says the system is capable of dissipating less heat, but the device line says the device will produce even *less* power, so the system becomes a thermal “black hole” and sucks all the energy out of the universe! Obviously one or more of our simplistic assumptions in the model breaks down at some point.)

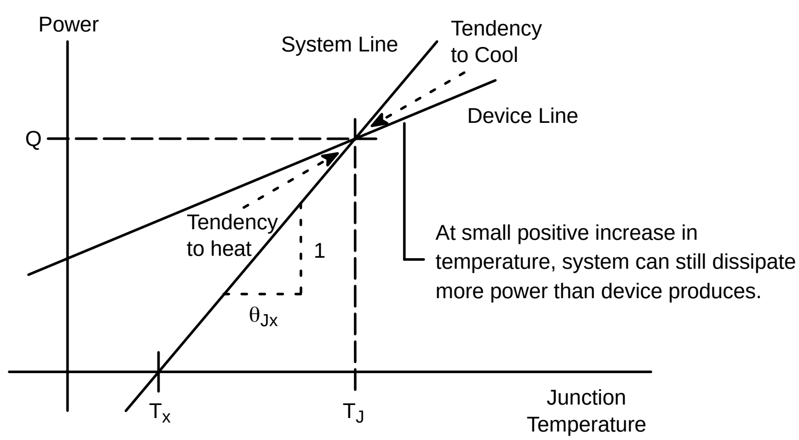

Figure 3. Operating Point of Thermal System where Power Increases with Temperature, Slopes Favorable

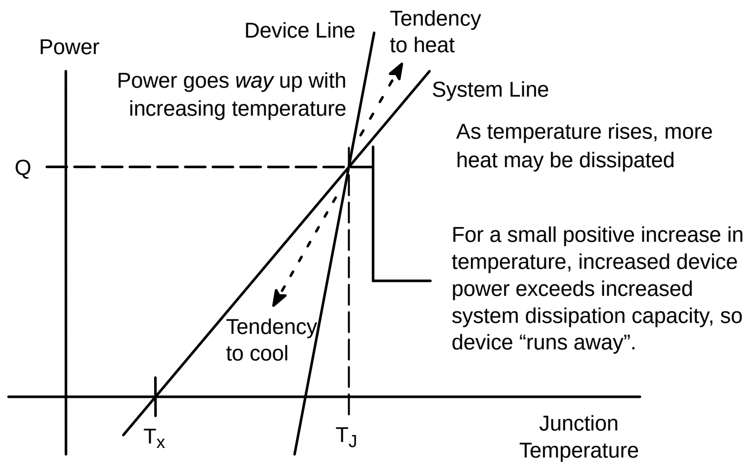

Figure 4. Operating Point of Thermal System where Power Increases with Temperature, Slopes Unfavorable

Figure 4, therefore, illustrates the essence of thermal runaway. And even though we’ve hypothesized “brief” perturbations, as if time had something to do with it, it is a concept essentially grounded in steady state system descriptions. Certainly time enters the picture if we are able to detect the onset of runaway and change the system quickly enough (for instance, turn off the power supply, changing the device line to a constant zero, independent of temperature).

Devices with Non−Linear Power Characteristics

Even more interesting observations may be made when we start looking at device lines that are not straight, as in Figures 5 through 9. In Figures 5 and 6, the device curve has a negative second derivative, meaning that although power does increase with temperature, its *slope* decreases with temperature. In Figure 5, although there are two intersections between the system and device lines, the lower “operating” point cannot be maintained. There the relative slopes are in the wrong relationship (as in Figure 4), resulting in *non*−restoring perturbations. In fact, if the device can be initialized to some temperature above the lower point, operation will “run away” to the upper point (which is stable, as in Figure 3). If it can’t be initialized to at least the lower intersection temperature, then it cannot be powered up at all.

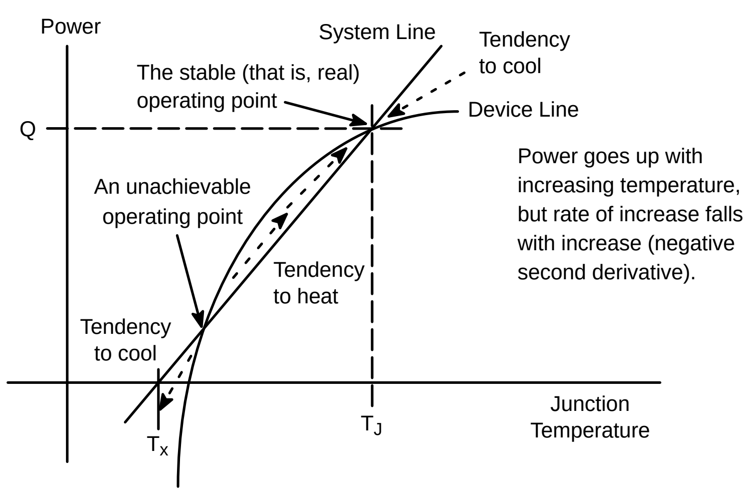

Figure 5. Operating Points of Thermal System when Device Line has Negative Second Derivative

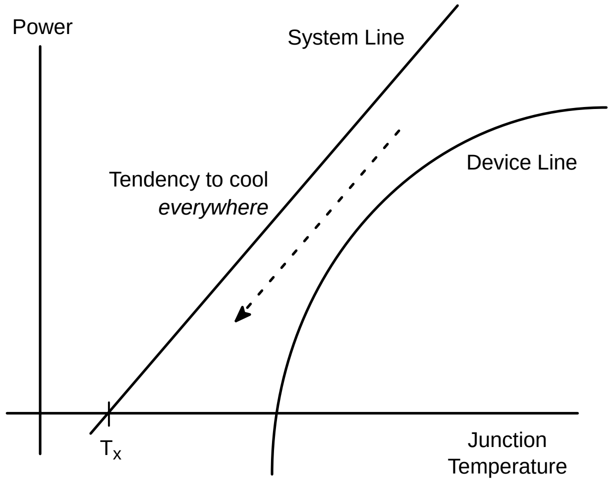

Figure 6. System with No Operating Point, Negative Second Derivative, Cannot be Powered Up

In Figure 6, the device line never reaches the system line at any point. With no intersections, there are no potential stable operating points, and indeed, this represents a system that can never be powered up, because no matter the temperature, the device cannot produce enough steady state power to “keep itself warm’’, so to speak, given the cooling system. If the slope of the cooling system is decreased (that is, thermal resistance increased), eventually a stable operating point may be established.

In Figures 7 through 9, we finally consider systems where the device line has a positive second derivative, meaning its *slope increases* with temperature. In Figure 7, as in Figure 5, there are two intersections between the device and system lines. In this case, however, only the lower operating point is stable. If junction temperature is perturbed above the upper intersection, it will run away and never return. Anywhere between the upper and lower points, it will “run away” back down to the lower point.

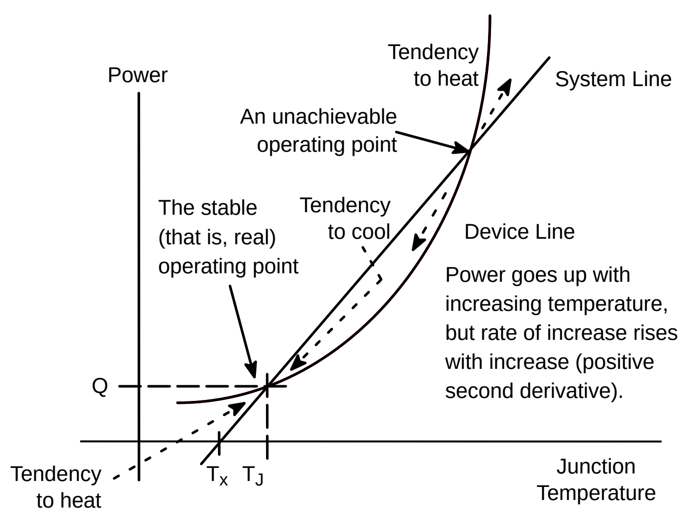

Figure 7. Operating Points of Thermal System when Device Line has Positive Second Derivative

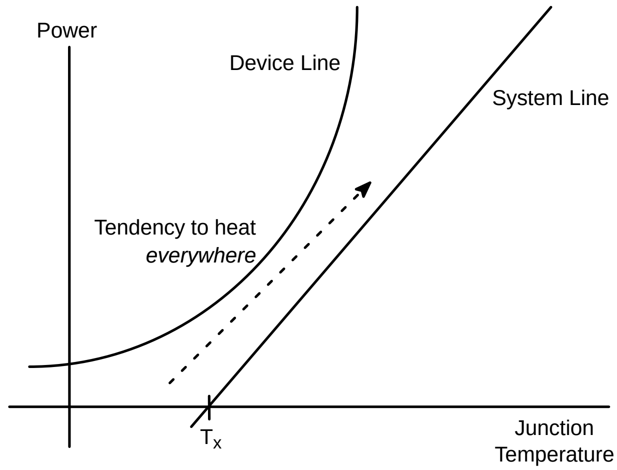

Figure 8. System with NO Operating Point, Overheats as Soon as Powered Up

In Figure 8, the device line (with a positive second derivative) never intersects the system line. Therefore, as in Figure 6, there are no steady state operating points. In contrast with Figure 6, however, any attempt to power up this system will result in thermal runaway, because at every temperature, the device produces more power than the cooling system can dissipate.

Finally, in Figure 9, we have the "perfect runaway" system. The device line (again with a positive second derivative) intersects the system line at exactly one point. For smooth curves, this must occur at tangency, that is, where the two curves have the same slope. Since the slopes are equal at the intersection point, there is not any tendency to heat or cool at that exact point. Strictly speaking, this is a point of neutral stability. And though *negative* perturbations will push the system back towards the neutral point, we can't expect this system to stay at equilibrium indefinitely. Even an infinitesimal perturbation upwards means there will be more power in than the system can dissipate, hence runaway will commence.

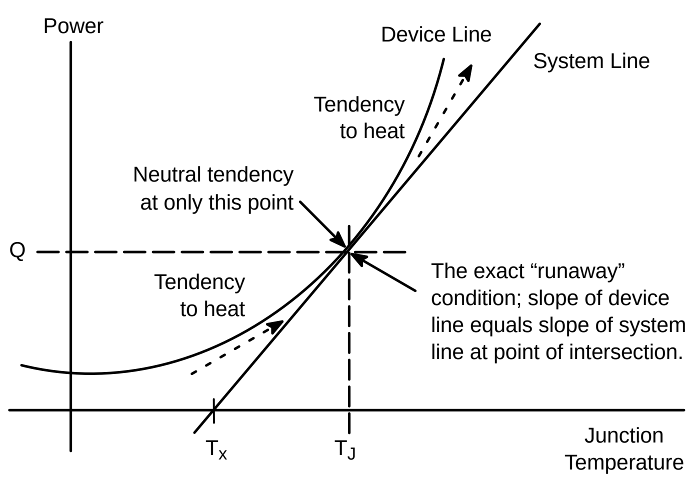

Figure 9. System with Exactly One "Runaway" Operating Point, Device has Positive Second Derivative

Same Device, Different Straight−Line Systems

Throughout the preceding discussion, we've focused on changes in the device characteristics in order to gain an understanding of the nature of the thermal runaway phenomenon itself. For these purposes, the system line has been fixed. By now, the significance of the relationships between the slopes of the device and system lines, and the relative position of one line above the other, should be clear. (It should also be clear that whether either or both lines are straight or curved is not of direct importance to the concept.)

In semiconductor applications, the problem will likely be turned around. It is likely that we will have a particular device, chosen primarily for the sake of its electrical properties, and we want to know what it will take to keep it operating at an acceptable temperature. If it turns out that, due to the thermal properties of the package, we simply cannot design an acceptable thermal system in which to place it, we may have to select another device with more favorable thermal characteristics. But restrictions on its thermal characteristics are generally secondary, and we won't know if they're significant until we try to design in the device we want. Let us therefore now turn our attention to systems where the device curve is fixed (in these examples having positive second derivative, which as we'll see later is true of "power−law" devices), and see what we can learn about the system characteristics with respect to thermal runaway. For simplicity, it will continue to be convenient to depict system lines as perfectly straight, characterized by a constant slope (the reciprocal of the system thermal resistance), and a specific reference temperature (thermal ground).

In Figure 10 we show one device line and three different system lines. System line A we'll call the "original" system. It is characterized by its slope, the reciprocal of its thermal resistance junction−to−ground, θ_{Jx1} , and its x−intercept, i.e., thermal ground, T_x. As depicted, it makes two intersections with the device line, but only the lower temperature intersection is a stable operating point (as discussed earlier). There is plenty of distance between the stable point and the unstable point, so we have some confidence that such a thermal system will work for this application.

However, there are two things of common interest in a thermal design. What if ambient (thermal ground) goes up? And what if the thermal resistance of the system increases, or is simply higher than planned? First consider the question of increasing ambient. All else being the same, system line B shows what happens as the existing thermal system (same slope) is shifted to the right – i.e., increasing ambient upwards to T_y until "perfect runaway" is reached and the device line is tangent to the system line. Clearly we cannot expect to safely operate our system if ambient approaches this critical temperature; that critical ambient temperature is associated with a particular "runaway" temperature, T_{R1}. On the other hand, system line C tells us that given a fixed ambient T_x, how much worse the system can be (i.e., bigger thermal resistance θ_{Jx2}), yet still avoid runaway. This gives us a completely different potential runaway temperature, T_{R2}. A surprising thing we will discover about power−law devices, is that the temperature *offset* between thermal ground and thermal runaway (for *any* system line tangent to the device curve), is fixed based on the power law, and is actually independent of the thermal system!

This is not to say, however, that the lower of these two potential runaway scenarios is more likely to occur. What should be clear, however, is that as the operating margin of the designed system is decreased (i.e., the placement of the stable actual operating point with respect to the two runaway conditions), the two runaway points converge, and the system will not be able to tolerate perturbations in either ambient *or* slope.

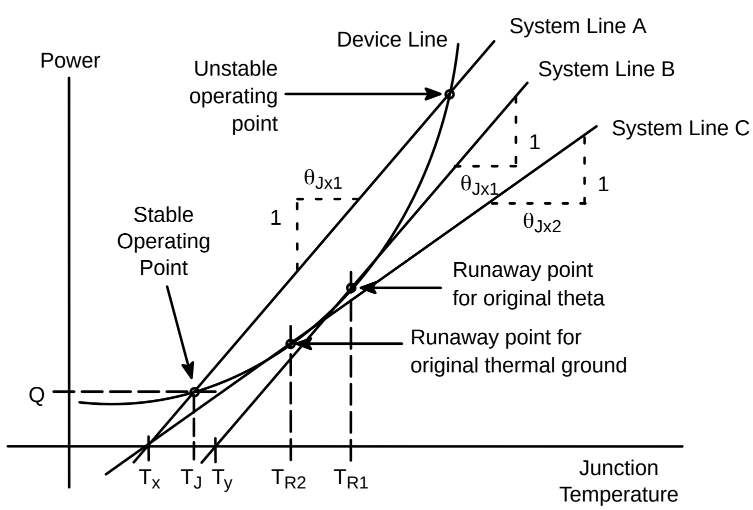

Figure 10. Generic Power Law Device and Generic Linear Cooling System

Power−Law Devices and Straight−Line Systems

Before proceeding, it is worthwhile to consider when a device power characteristic will be a power law at all. Typically, simple devices such as diodes and rectifiers, and even many transistors under certain conditions, have a current vs. temperature characteristic (at fixed voltages) that follows a power law. For instance, leakage current in diodes is often described using the rule of thumb that leakage doubles for every increase in temperature of 10°C. This particular power law could be expressed as:

$$I = I_O 2^{\frac{T}{10}} \qquad \text{(eq. 4)}$$

(Here, I_O would be the leakage current at 0°C. If the behavior is really a power law, obviously we can figure out what the leakage is at 0°C, given its value at any other temperature.) Equation 4 can be re−expressed in terms of the base of the natural logarithms, as in:

$$I = I_O e^{(\ln 2)\frac{T}{10}} = I_O e^{\frac{T}{\left(\frac{10}{\ln 2}\right)}} \qquad \text{(eq. 5)}$$

So we see that any power−law behavior, that is, behavior described by a geometric increase in the dependent variable (e.g., a factor of 2) for a linear increase in the independent variable (e.g., every 10°C), is just another exponential function, as in Equation 6:

$$I = I_O e^{\frac{T}{\lambda}} \qquad \text{(eq. 6)}$$

From this example, it may be seen that we can define the "strength" of the power law in terms of that parameter λ. If device behavior is governed by a power law, then knowing the value of the independent variable (in this case, current) at any two corresponding values of the independent variable (in this case temperature), we can say:

$$\lambda = \frac{T_1 - T_2}{\ln\left(I_1 / I_2\right)} \qquad \text{(eq. 7)}$$

So far in this example, we've only talked about current, and we need power. But given the premise that a diode under reverse bias (i.e., leakage mode) sees a constant voltage (reasonably assumed to be independent of the temperature of the particular diode), its power dissipation, therefore, is also a power law:

$$Q = V_R I_O e^{\frac{T}{\lambda}} = Q_O e^{\frac{T}{\lambda}} \qquad \text{(eq. 8)}$$

Note that under constant *current* (as opposed to constant voltage operation), the primary operating condition for many applications of diodes (e.g., flyback in an inductive circuit), the power law relationship does *not* hold with temperature. Indeed, it is typical for diodes that voltage goes down linearly with temperature at a fixed current. Therefore, diodes in constant current operation have linearly decreasing power with increasing temperature, and as in Figure 2 earlier, are essentially immune to thermal runaway concerns.

For the power law situation, however, we may calculate the slope of the device curve by taking its derivative, that is:

$$\frac{dQ}{dT} = \frac{Q_O}{\lambda} e^{\frac{T}{\lambda}}, \text{ also } \frac{d^2Q}{dT^2} = \frac{Q_O}{\lambda^2} e^{\frac{T}{\lambda}} \qquad \text{(eq. 9)}$$

Thus we see that the slope increases with increasing temperature, and all the previous discussions based on device curves with positive second derivatives apply. In particular, Figure 10 provides the setting for the subsequent mathematical development.

Perfect Runaway

We are now in a position to find the mathematical solution to "perfect runaway'' in a power−law device cooled by a linear thermal system. To begin, let's recap the two pertinent equations.

the linear system line:

$$Q = \frac{T - T_X}{\theta_{JX}} \qquad \text{(eq. 2)}$$

the power−law device line:

$$Q = Q_O e^{\frac{T}{\lambda}} \qquad \text{(eq. 8)}$$

In this system of two equations, we have two variables, the junction temperature, T, and the power dissipation, Q. We also have four independent parameters, the thermal ground reference, T_X, the thermal resistance of the cooling system, θ_{JX}, the reference power level Q_O (which, as expressed, would be the power dissipated by the device at 0°C), and the strength of the power−law function itself, λ. By finding an exact solution to this set of equations, that is, a pair of values (T, Q) that satisfies both equations, we will also create some relationships between the various parameters.

For clarity in solving the equations, it will be helpful to make a change of variables. Let us define:

$$z = \frac{T - T_X}{\lambda} \qquad \text{(eq. 10)}$$

With this new variable, we can rewrite our two equations as follows. From Equation 2, we have:

$$Q = \frac{T - T_X}{\theta_{JX}} = \frac{\lambda}{\theta_{JX}} \frac{T - T_X}{\lambda} \qquad \text{(eq. 11)}$$

that is:

$$Q = \frac{\lambda}{\theta_{JX}} z \qquad \text{(eq. 12)}$$

the linear system line, and from Equation 8 we have:

$$Q = Q_O e^{\frac{T}{\lambda}} = Q_O e^{\frac{T_X}{\lambda}} e^{\frac{T - T_X}{\lambda}} \qquad \text{(eq. 13)}$$

or:

$$Q = Q_O e^{\frac{T_X}{\lambda}} e^{z} \qquad \text{(eq. 14)}$$

the power−law device line (curve).

Equation 14 suggests that we might choose to define the non−dimensional power q as:

$$q = \left(\frac{1}{Q_O} e^{\frac{-T_X}{\lambda}}\right) Q \qquad \text{(eq. 15)}$$

thus, we can make a final restatement of the two governing equations as follows.

the device line:

$$q = e^z \qquad \text{(eq. 16)}$$

the system line:

$$q = kz \qquad \text{(eq. 17)}$$

where:

$$k = \frac{\lambda}{\theta_{JX} Q_O} e^{\frac{-T_X}{\lambda}} \qquad \text{(eq. 18)}$$

Eliminating q from between Equation 16 and Equation 17, we are left with a remarkably simple relationship defining points z that are the intersection points (i.e., the operating points) of our system.

$$kz = e^z \qquad \text{(eq. 19)}$$

Figure 11 is a graphical representation of our transformed problem, originally seen in Figure 10. It may be interpreted as the pure exponential function and its intersection with a straight line of slope k, passing through the origin of the (z, q) coordinate system.

Now for points representing "perfect runaway" of this system (i.e., the system line being tangent to the device−line), there is an exact solution. It begins by recognizing a unique property of the pure exponential function, namely, the slope of the function is equal to its value at every point. In other words, everywhere on the device line:

$$\frac{dq}{dz} = e^z = q \qquad \text{(eq. 20)}$$

Indeed, we may make an even more interesting observation for lines tangent to the pure exponential curve (see Figure 12). Since the slope may be defined as the "rise over the run," it must be the case that the "run" is exactly unity for *every* tangent extended to its z−intercept. That is, in general the distance between the z−coordinate of any point on the curve, and the corresponding z−intercept of the tangent through that point, is:

$$z_R - z_X = 1 \qquad \text{(eq. 21)}$$

Thus, for z_{R1}, the runaway point arising from a shift in the original ambient.

$$k_1 = e^{z_{R1}} \qquad \text{(eq. 22)}$$

hence:

$$z_{R1} = \ln k_1 \qquad \text{(eq. 23)}$$

So we find the maximum ambient causing this runaway (from Equation 21) to be:

$$z_{X1} = \ln k_1 - 1 \qquad \text{(eq. 24)}$$

For z_{R2}, the runaway point arising from an increase in system thermal resistance, since the original ambient is represented by $z = 0$, we now find from Equation 21 that:

$$z_{R2} = 1 \qquad \text{(eq. 25)}$$

hence:

$$k_2 = e \qquad \text{(eq. 26)}$$

We should now return to the original, dimensional, form of the problem, to fully understand the significance of Equations 21 through 26. First, converting Equation 21 back into the original coordinates, we find:

$$T_R = T_X + \lambda \qquad \text{(eq. 27)}$$

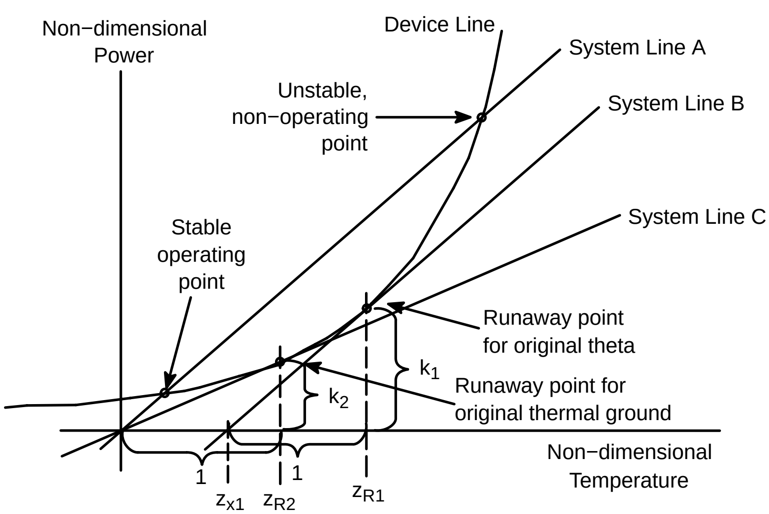

Figure 11. Solution of Device and System Equations in Non−dimensional Coordinate System

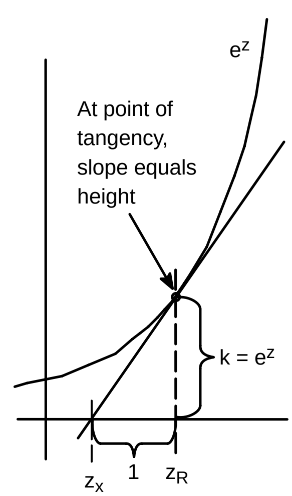

Figure 12. Auxiliary Plot of Tangent to e^z

This is the result previously expressed verbally, namely for every “perfect runaway” system line (i.e., the system line is tangent to the device line), the distance between thermal ground and the runaway point is actually fixed based solely on the power law strength, and doesn’t depend on the particular value of thermal ground or the system line’s slope! This delta is precisely the strength of the power law, λ.

From Equation 23, we find:

$$\frac{T_{R1} - T_x}{\lambda} = \ln\left(\frac{\lambda}{\theta_{Jx1} Q_o} e^{\frac{-T_x}{\lambda}}\right) \quad \text{(eq. 28)}$$

$$\frac{T_{R1}}{\lambda} - \frac{T_x}{\lambda} = \ln\left(\frac{\lambda}{\theta_{Jx1} Q_o}\right) - \frac{T_x}{\lambda} \quad \text{(eq. 29)}$$

$$T_{R1} = \lambda \ln\left(\frac{\lambda}{\theta_{Jx1} Q_o}\right) \quad \text{(eq. 30)}$$

This expression tells what the runaway temperature would be of a system with the same slope as the actual system, as ambient increases until runaway occurs. Clearly it depends not only on the system thermal resistance θ_{Jx}, but also on the power−law characteristics of the device (and not merely on λ, the strength, but also on the specific reference power level, Q_o). So again from Equation 27, we have:

$$T_{x1} = \lambda \ln\left(\frac{\lambda}{\theta_{Jx1} Q_o}\right) - \lambda \quad \text{(eq. 31)}$$

If instead we assume thermal ground is fixed, and want to know the largest thermal resistance for which the system will remain stable, Equation 26 tells us that:

$$\frac{\lambda}{\theta_{Jx2} Q_o} e^{\frac{-T_x}{\lambda}} = e, \text{ so}$$

$$\theta_{Jx2} = \frac{\lambda}{Q_o} e^{-\left(\frac{T_x}{\lambda} + 1\right)} \quad \text{(eq. 32)}$$

and this runaway temperature, from Equation 25, is the fixed λ above the original thermal ground, that is,

$$T_{R2} = T_x + \lambda \quad \text{(eq. 33)}$$

Power Laws Based on Power

Finally, even though a real device may not follow the particular rule of thumb quoted earlier, if it follows a power law at all (hence the strength can be calculated), these general results will apply. For completeness, let us also express the power−law strength directly in terms of power (rather than current). Instead of Equation 7, we might prefer, therefore:

$$\lambda = \frac{T_1 - T_2}{\ln\left(Q_1 / Q_2\right)} \quad \text{(eq. 34)}$$

It may be noted that any two reference temperatures and power levels may be used to calculate λ, but a zero temperature power level must be used in Equations 28, 31, and 32. If neither of those reference temperatures is zero, then either one may be used to calculate Q_o, once λ has been calculated, using:

$$Q_o = Q_i e^{\frac{-T_i}{\lambda}} \quad \text{(eq. 35)}$$

The Operating Point Itself

All these runaway limits themselves are obviously of some interest, but they remain hypothetical in the sense that if the system is designed to operate in a safe region, the operating point itself will be safely removed from these theoretical runaway points. As was indicated in the preceding development, the system may still experience restoring forces for brief power or temperature excursions taking the junction temperature all the way up to the upper intersection between the system and device lines, considerably beyond either runaway temperature T_{R1} or T_{R2}. The problem, of course, is that if the perturbation lasts for “very long” (somewhat vaguely defined), the system will actually be attempting to find a new steady state at this upper point, which once reached, as we have seen, is not stable. And if the perturbation is actually due to a permanent shift in either ambient or system resistance, then in fact the two intersection points, and the two disparate runaway points, will all be converging toward a single dangerous operating/runaway point somewhere in the middle.

So how much margin does one actually have in a specific design? To answer that, one must solve Equation 19 for the particular thermal ground and system resistance of interest. Due to its transcendental nature, Equation 19 has no general closed−form solution, so it must be solved graphically or numerically. Even so, some general statements can be made. First, it has already been observed that the critical value for perfect runaway, at the designed thermal ground, results when $k = e$ (Equation 26). This means that for any proposed combination of thermal ground and system resistance, if $k > e$, there are going to be two intersection points, the lower one being the stable operating point. On the other hand, if $k < e$, there will be no solution at all.

So the first step in finding the actual operating point is to compute k (Equation 18) and see if there *is* a solution. (It is probably easier to calculate the ratio k/e to see if it’s greater than unity, since you probably don’t remember e, and it’s built into your calculator or spreadsheet anyway!) If there is not a solution, there are three possible approaches:

(a) lower ambient: Equation 31 gives the maximum ambient tolerable (with no margin), given the device and the originally supposed system resistance;

(b) lower system thermal resistance: Equation 32 gives the maximum tolerable theta (with no margin) given the device and the originally supposed thermal ground;

(c) a different device must be used, and given that most devices are going to have a roughly similar power law strength, this means finding a device with a lower Q_o − almost certainly meaning one with a larger piece of silicon inside.

A Quantitative Example

Consider a particular rectifier in an SMB package. From its data sheet we find the following information:

Vr (V)	12	40
Tmax (°C)	125	125
Tref (°C)	75	75
Itmax (A)	8.50E−3	2.80E−2
Itref (A)	5.20E−4	1.70E−3

Assuming reverse bias operating mode, Equations 34 and 35 give us, for the two reverse voltages (note similarity in λ at two temperatures).

λ (°C)	17.9	17.8
Q_o (W)	9.4E−5	1.02E−3

The data sheet also gives a psi−JL value of 25°C/W, so let's suppose that in an actual application we can achieve a system thermal resistance of 100°C/W (in still air, and with other heat sources in the vicinity). Let's also suppose that we're going to have a worst−case ambient of 75°C. Given all these values, we can now compute all the remaining quantities for which we've previously derived expressions:

k/e (compare to unity)		10.6	0.97
given theta	T_x max (°C)	117.2	74.4
	T_{R1} (°C)	135.1	92.2
given ambient	θ_{JX2} max (°C/W)	1055	96.6
	T_{R2} (°C)	92.9	92.8

For the lower voltage application, we've got lots of margin. First, the k/e ratio is much larger than unity, so we know we've got a stable operating point. Second, given our choice of 100°C/W for θ_{JX1}, max ambient is comfortably high, and runaway temperature exceeds maximum rated junction temperature. Third, we're nowhere near the limiting system resistance (1055°C/W), so its associated runaway temperature of only 92.9°C is no concern.

On the other hand, at the higher voltage application, we have problems. First, the k/e ratio says there is no solution for the operating point. We could have concluded this indirectly by noting that our chosen theta exceeds the max theta calculated based on the given ambient, and that maximum ambient exceeds our chosen ambient, given theta. Unfortunately, we can't lower the maximum ambient, but upon further consideration, we believe that a theta of 80°C/W is quite realistic. Recomputing values, we now find that k/e has risen to 1.609, maximum ambient has risen to 83.5°C, and runaway temperature based on this theta will be 101.3°C. Both runaway temperatures are well below maximum rated temperature for the device, but we won't know our "real" margins without calculating the actual operating point, and its unstable upper partner. So, using the goal−seek feature of a spreadsheet program (or the iterative procedure outlined in the sidebar), we find the two mathematical solutions of Equation 19, for $k = 1.609e$, to be $z = 0.312$ and $z = 2.315$.

These translate into a stable operating point at 80.6°C (and 0.09 W), and its unstable partner at 116.3°C (0.69 W). So no matter how you look at it, even this 80°C/W system design will succumb to thermal runaway somewhat below the maximum operating temperature of the device. We have at least established some reasonable margins around the design point.

Solving Equation 19

Observe that Equation 19 may be rewritten ln (kz) = z, then posed as function y = z−ln (kz) to which Newton's method may be applied. This results in the fairly robust iterative formula shown at right. For k ranging between 2.72 and 1000, convergence is to a dozen significant digits in fewer than 10 iterations.

$$z_{i+1} = \frac{1 - \ln(kz_i)}{\frac{1}{z_i} - 1}$$

$$z_o = \frac{1}{k}$$ Converges to lower point

$$z_o = \ln(k)$$ Converges to upper point

A Seeming Paradox

Interestingly, Equation 30 shows that for a given device in a given thermal system, thermal runaway temperature does *not* depend directly on the value of thermal ground. It does, however, require defining the system thermal resistance to encompass the entire thermal path from junction down to thermal ground. A specific example will serve to illustrate what may appear to be a nonsensical consequence, but which in fact will emphasize this requirement, once understood.

Imagine the two thermal systems illustrated in Figure 13.

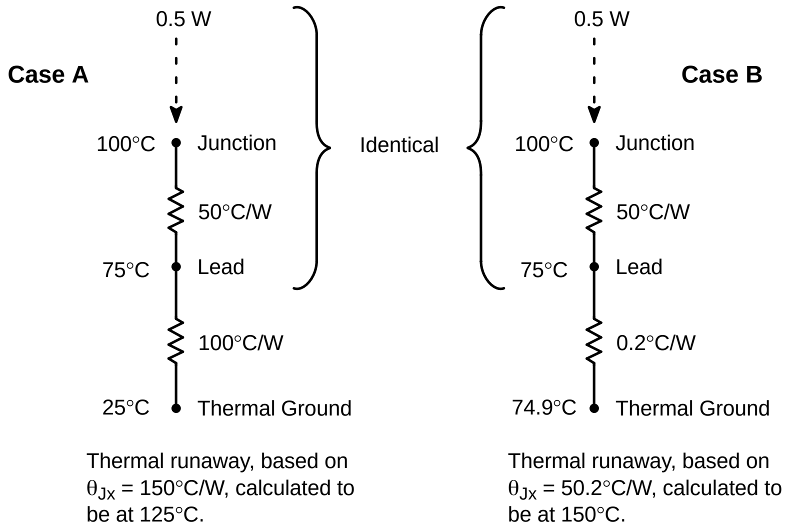

Figure 13. Two Applications of Another Device (*not* the device of the previous example!)

The point of these two system models is that the very same device, seeing exactly the same "local" conditions (that is, junction temperature, lead temperature, and power dissipation), has, according to Equation 30, two vastly different "runaway" temperatures (based on the two different system resistance values). Clearly the *only* difference in the two applications is the combination of thermal ground and total system thermal resistance. So the question which may present itself is, "If the device is experiencing the exact same boundary conditions (so it would seem), how could it possibly be sensitive to the thermal resistance beyond those boundaries, and thus have different runaway values?" Another way of putting it might be, given this simple model, "If we increase the power level in Case A until T_J reaches 125°C, why will runaway occur, whereas if we increase the power level in Case 2 until T_J reaches 125°C, will it remain stable?"

Paradox Lost

The answer lies in the very reason that the total system thermal resistance must be used in calculating runaway temperature. Consider first what happens in Case A if the power is increased by a small increment, say 0.1 W. Thermal ground, of course, remains fixed at 25°C (that's why it's thermal ground). T_L, on the other hand, rises by 10°C, and T_J rises by 15°C. Consider now Case B for the same small increment of power of 0.1 W. Again, thermal ground remains fixed (this time at 74.9°C). Now, however, T_L, rises by only 0.02°C, and T_J rises by only 5.02°C.

We haven't been explicit about the device power−law parameters (though they went into the runaway calculation, and into the very fact that the operating points of Case A and Case B are what they are). Indeed, changing the power must be viewed as a temporary disturbance to the system anyway, since you can't actually change the power at all and still remain in equilibrium – that's why the power law device has a specific operating point in a particular context.

But we can surely see this: the fact that a small perturbation in power pushed the temperature of Case A's T_J up by 15°C, whereas that same perturbation in power pushed up the temperature of Case B's T_J by only one−third that amount, suggests that Case A is much more sensitive to small perturbations. (This could have been turned around to show that it only takes about a third as much power in Case A to cause a similar increment in T_J as is required in Case B. So we'd say Case B is more "robust" to power fluctuations.) The argument then goes that because this is a power law device, small perturbations in Case A are more dangerous than in Case B. This may be translated into a lower runaway temperature.

Yet another way to see it is that if we'd designed our two systems to run at a very small margin below the runaway temperature of Case A, then (due to this differing sensitivity in T_J to power), a very small perturbation in Case A would send it into runaway, whereas the same small perturbation in Case B is just that, a small perturbation.

There is yet a third way to look at it. Consider how little the lead temperature changes in Case B, for the same increment in power as in Case A. Because the lead is so much closer to thermal ground in Case B, it makes sense that at a given operating point, power and temperature at the junction should be much more firmly "anchored." Potential runaway must therefore be farther away from the operating point, even though the operating point itself is the same in both applications.

CONCLUSION

This monograph has focused on power−law devices. In short, this is because we can make some precise statements about their behavior, even if in a somewhat idealized thermal system. Nevertheless, it should be clear that regardless of the actual shape of the device line, or even of the system line, the concepts driving a thermal design with respect to thermal runaway are generally applicable, even if they must be implemented graphically or through some other iterative method of calculation.

For a power−law device, Equation 27 was perhaps the most surprising result, namely that for a "perfect runaway" system (that is, one where the operating point is at exactly the thermal runaway point), the offset between thermal ground and runaway temperature depends *solely* on the power−law strength. However, we should never design a system to operate at precisely the runaway point, so although this result is an interesting mathematical fact, it is only of secondary importance. In fact, there are two different theoretical runaway points (one based on holding system resistance constant, and varying ambient, and the other based on holding ambient constant, and varying the system resistance). The latter is always the lower of the two, and is at the fixed power−law strength, λ, from the nominal thermal ground. (It is also surprising that for the 2x−per−decade rule of thumb, λ comes in at only 14.4°C! A cursory study of several actual data sheets shows that λ may be as high as 20−21°C for some real devices. In any event, this margin is quite small in absolute terms.) However, in many situations, there will be relatively much more margin in system resistance than in thermal ground, so the former runaway point will be the one of primary interest.

Thermal runaway has sometimes been stated exclusively in terms of the relative slopes of the device and system lines. That is, if:

$$\frac{dQ}{dT} > \frac{1}{\theta_{Jx}} \qquad \text{(eq. 36)}$$

then runaway will occur. It should be evident from this study that this is somewhat of an oversimplification. If both device and system lines are straight lines, it is surely true. However, if either or both lines are curved, then not just the relative slopes, but also their second derivatives, and the specific placement of the lines relative to each other, are all factors. More particularly, for power−law devices, there is not just a stable operating point and potentially two runaway points, but also an unstable higher operating point. It is actually this upper, unstable operating point that limits the junction temperature (and spells disaster if junction temperature exceeds this value), so long as the conditions do not persist for "too long." But what happens due to increases in ambient or system resistance is that the two "operating" points begin to converge, bringing the two runaway points inwards with them. All four points thus coalesce at a common value when conditions depart from the nominal for a sufficiently long period of time, and runaway then occurs. Clearly, therefore, the precise manner and duration by which the system is perturbed from the operating point, dictates the exact temperature at which runaway ultimately occurs.

It is also certain that one can never have a stable operating point where the condition of Equation 36 is violated locally (for instance, the upper point just referred to in power−law device operation). But, as we saw, in the case of devices with *negative* second derivatives, this only means that the *lower* of two theoretical operating points (i.e., solutions to the set of equations) cannot be maintained. The system may "run away" from the lower point, but not necessarily to oblivion, only to the upper, stable point.

Accounting for all these factors, it will be found that in many applications, thermal runaway is likely to occur at a temperature *far* below the maximum rated temperature of the device. In other words, if thermal runaway is a failure mode of the application, it may be the most stringent constraint on the thermal design.

PUBLICATION ORDERING INFORMATION

LITERATURE FULFILLMENT:
Literature Distribution Center for ON Semiconductor
P.O. Box 61312, Phoenix, Arizona 85082−1312 USA
Phone: 480−829−7710 or 800−344−3860 Toll Free USA/Canada
Fax: 480−829−7709 or 800−344−3867 Toll Free USA/Canada
Email: orderlit@onsemi.com

N. American Technical Support: 800−282−9855 Toll Free USA/Canada

Japan: ON Semiconductor, Japan Customer Focus Center
2−9−1 Kamimeguro, Meguro−ku, Tokyo, Japan 153−0051
Phone: 81−3−5773−3850

ON Semiconductor Website: http://onsemi.com

Order Literature: http://www.onsemi.com/litorder

For additional information, please contact your local Sales Representative.