AND8216/D

Minimizing Scatter in Experimental Data Sets

Packaging Technology Development

Prepared by: Roger Paul Stout, PE
ON Semiconductor

ON Semiconductor®

http://onsemi.com

APPLICATION NOTE

Abstract

Thermal transient data collected to characterize packaged semiconductor devices is often acquired from a number of individual samples that then must be averaged together. Yet each individual transient curve has a relatively large experimental uncertainty, which is usually a combination of unmeasurable initial transient thermal response (related to electrical transients that cannot be separated from the thermal signal), coupled with mounting variability not fairly attributable to the device itself being tested (but is rather due to differences in test board characteristics, solder thickness variation, etc.). When a limited range of the data is believed to accurately represent device−only dependence, a useful, simple, yet rigorous method has been developed to allow the analyst to minimize the scatter in the data set, while preserving the average value in the specified range. Application of this method to actual data is presented for purposes of illustration, and additional implications in the use of the method are discussed. Although this method was developed and is applied here in the context of thermal transient characterization, it is universal in potential application; the underlying mathematical "theorem" is completely general.

Key Words

variability, standard−deviation, thermal, transient

Glossary of Symbols

E	pooled variance of entire data set
T_{ij}	original, unshifted data values
$\overline{T}_i$	mean of the original, unshifted data values for device i
$\overline{\overline{T}}$	mean of the entire set of original, unshifted data values
theta	thermal resistance or normalized junction temperature, defined as (junction temperature − ambient)/power
x_i	general data variable for statistical formulas

Greek Symbols

ε_i	offset by which entire data set for a device is shifted
μ	mean value of a population
σ	standard deviation of a population
σ_j	standard deviation of population j
σ^2	variance of a population

Subscripts

i, j	data set indices and expressions
m	number of time steps over which data is correlated
n	number of devices for which data is correlated

INTRODUCTION

Accurate thermal transient measurements at very short time scales are a challenge for various reasons [1][2][3]. When a cooling curve technique is used for measurements, there is a glaring deficiency in the process. Purportedly identical devices will never start at the same normalized temperature (that is, steady state thermal resistance, or theta), mainly due to mounting variations between devices, random variations in ambient conditions, and to a much lesser extent, actual manufacturing variation within the devices.

© Semiconductor Components Industries, LLC, 2006

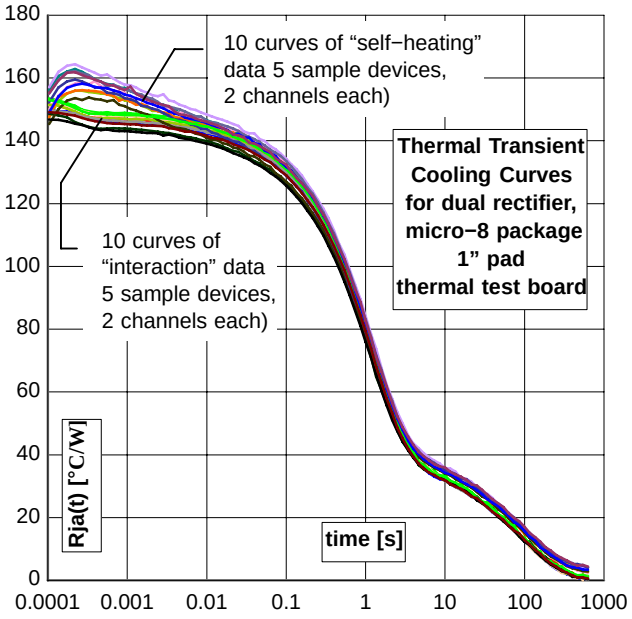

Figure 1. Typical Transient Cooling Data

Figure 1 illustrates a typical set of thermal transient cooling curves, obtained by testing five samples of a particular dual−rectifier device. Each of the two channels in turn, of each sample, was heated until steady state was reached. Then, power was turned off and both channels were measured as the entire device cooled back to ambient. The essential goal is to measure the overall thermal transient resistance of the device. It is somewhat complicated in this case by the fact that we are also interested in the indirect or “interaction” heating of each channel on the other, yet from the raw data in Figure 1, the difference between the channels (or even that there *is* a meaningful difference) is not entirely obvious. Ultimately we might use this data to generate a two−input thermal RC network representing the thermal performance of the device, enabling a customer to do dynamic modeling of the two junction temperatures when the device is heated asynchronously and asymmetrically. Before we do, however, we should prefer to have more confidence that the “signal” present in the data is worth the effort.

This monograph develops a general method of minimizing the scatter between an arbitrary number of data sets, while preserving the original mean values. It turns out, in fact, that a benefit of applying this technique to the specific case of thermal transient data, is identifying when there is a “flyer” in the group. Without this data analysis technique, due to the intrinsic variability in the measurement methodology, normal scatter often overwhelms the signal of internal variation.

Basic Idea

The basic idea is that one has an extended set of correlated data points for which the average values are presumed to have some statistical validity, but are believed to reflect, between the various individual subsets of data, some random offsets. For example, we have a set of transient cooling theta measurements for a particular device, which we want to statistically match up with a similar set of measurements for a supposedly identical device. Let us say we have a total of n devices, all presumed identical, and for each of these we have made transient cooling measurements over the same series of m time steps. The problem is that due to mounting variations, each device started at a somewhat different initial steady state theta, which is necessarily unknown. We’d like to see just how exactly these data sets can be made to overlay each other over those m points, simply by adjusting the relative position of the sets with respect to each other, yet not shifting the overall mean value of the sets.

Consider first a trivial case where the procedure results in a unsurprisingly trivial conclusion. Table 1 presents five “sets” of data, each set consisting of one theta measurement taken at the same time of 0.001 second.

How can these data sets be shifted so as to maintain this original average value, yet minimize the “scatter”? Pretty clearly, if set #1 shifts +2°C/W, set #2 by +1°C/W, set #3 stays put, set #4 shifts −1°C/W, and set #5 by −2°C/W, then all five sets will end up at a value of 3°C/W, and the standard deviation will be the minimum possible: 0°C/W!

Is the problem always this trivial? Certainly not. Table 2 extends the data “sets” of Table 1 with additional points for each of these five devices at a time value of 0.002 second. Now, in addition to the standard deviation for each individual time, we should compute the pooled standard deviation for these sets from the standard deviations of the individual time points, namely:

$$\sqrt{\frac{1}{2}(1.58^2 + 0.79^2)} = 1.25$$

Table 1. Hypothetical Cooling Data at 0.001 s

Time Seconds	Device 1 °C/W	Device 2 °C/W	Device 3 °C/W	Device 4 °C/W	Device 5 °C/W	Average °C/W	Std. Dev °C/W
0.001	1.0	2.0	3.0	4.0	5.0	3.0	1.58

Table 2. Hypothetical Cooling Data at 0.001 s and 0.002 s (pooled s. d. of 1.25)

Time Seconds	Device 1 °C/W	Device 2 °C/W	Device 3 °C/W	Device 4 °C/W	Device 5 °C/W	Average °C/W	Std. Dev °C/W
0.001	1.0	2.0	3.0	4.0	5.0	3.0	1.58
0.002	1.0	1.5	2.0	2.5	3.0	2.0	0.79

Table 3. Hypothetical Cooling Data with Initial Offsets (pooled s. d. of 0.56)

Time Seconds	Device 1 °C/W	Device 2 °C/W	Device 3 °C/W	Device 4 °C/W	Device 5 °C/W	Average °C/W	Std. Dev °C/W
0.001	3.0	3.0	3.0	3.0	3.0	3.0	0.00
0.002	3.0	2.5	2.0	1.5	1.0	2.0	0.79

Table 4. Hypothetical Cooling Data with Minimum−Scatter Offsets (pooled s. d. of 0.40)

Time Seconds	Device 1 °C/W	Device 2 °C/W	Device 3 °C/W	Device 4 °C/W	Device 5 °C/W	Average °C/W	Std. Dev °C/W
0.001	2.5	2.75	3.0	3.25	3.5	3.0	0.40
0.002	2.5	2.25	2.0	1.75	1.5	2.0	0.40

Intuitively, we see now that these devices didn't simply start at different steady state thetas, they appear to have cooled at different rates. No amount of shifting entire data sets up or down can change this basic character. Nevertheless, our "rule" or premise in this data−shifting algorithm is that all points for a specific device have to shift by the same amount, so if we were to apply those obvious shifts from the first trivial scenario to this larger set of points, we'd obtain the values shown in Table 3. And now the pooled standard deviation is 0.56 − definitely lower. But is *this* the best that can be done?

What if we use shifts of 1.5, 0.75, 0, −0.75, and −1.5, respectively? Table 4 results, with a pooled standard deviation of 0.40, smaller yet. Indeed, using the approach that will be developed in this monograph, one can show that this pooled standard deviation is the minimum possible value for this given data, and that (without changing the averages) these offsets *uniquely* yield this value. Further, it will be demonstrated that there is always such a set of offsets.

In this hypothetical introductory problem, we can't necessarily convince ourselves that there was much point in adjusting this data, given our "intuitive" observation that they don't seem even to represent the same cooling trends. However, in a real set of transient cooling curves which span several orders of magnitude of time, and a wide range of transient theta from a heated equilibrium to a final ambient equilibrium (Figure 1), a great deal can be learned about the similarities (or differences) between the individual device characteristics by utilizing the technique outlined herein, and by restricting our attention to different subsets of the time data for different purposes.

The Method

With the foregoing illustration in mind, let us now get mathematically precise. First, we are going to deal with m sets (the time steps above) of n correlated data points each. We shall designate this data as:

$$T_{ij}$$

where i denotes the device and j denotes the time step. For each device i, we will shift its entire set of points by some common offset ϵ_i. So a little more exactly, we are interested in discovering the following shifted sets of values:

$$T_{ij} + \epsilon_i$$

Now recall the definitions for the mean and variance of a population of x_i data values:

$$\text{mean } \mu = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \text{variance } \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

substituting for μ into σ^2, and with a little juggling, yields the standard alternative form:

$$\sigma^2 = \frac{1}{n}\left[\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2\right]$$

Thus we can write the variance of the j^{th} set of shifted values (i.e., over the n devices at a common time step j) as:

$$\sigma_j^2 = \frac{1}{n}\left[\sum_{i=1}^{n}(T_{ij} + \epsilon_i)^2 - \frac{1}{n}\left(\sum_{i=1}^{n} T_{ij} + \epsilon_i\right)^2\right] \quad \text{(eq. 1)}$$

and we can write the pooled variance of the entire set of data over all the specified time steps as:

$$E = \frac{1}{m}\sum_{j=1}^{m}\sigma_j^2 \qquad \text{(eq. 2)}$$
$$= \frac{1}{m}\sum_{j=1}^{m}\frac{1}{n}\left[\sum_{i=1}^{n}(T_{ij}+\epsilon_i)^2 - \frac{1}{n}\left(\sum_{i=1}^{n}T_{ij}+\epsilon_i\right)^2\right]$$

Our goal is to minimize the pooled variance with respect to the adjustable offsets ε_i, which means that we must take the derivatives of (2) with respect to each of them.

Because of the various summations and non−linear terms, execution of the differentiation process is not entirely obvious or straightforward. To ensure that the correct results are obtained, it is useful first to expand the terms of (2) somewhat, as follows:

$$E = \frac{1}{m}\sum_{j=1}^{m}\frac{1}{n}\left[\sum_{i=1}^{n}(T_{ij}^2 + 2T_{ij}\epsilon_i + \epsilon_i^2) - \frac{1}{n}\left(\sum_{i=1}^{n}T_{ij}+\sum_{i=1}^{n}\epsilon_i\right)^2\right]$$
$$= \frac{1}{m}\sum_{j=1}^{m}\frac{1}{n}\left\{\sum_{i=1}^{n}(T_{ij}^2 + 2T_{ij}\epsilon_i + \epsilon_i^2)\right. \qquad \text{(eq. 3)}$$
$$\left. - \frac{1}{n}\left[\left(\sum_{i=1}^{n}T_{ij}\right)^2 + 2\sum_{i=1}^{n}T_{ij}\sum_{i=1}^{n}\epsilon_i + \left(\sum_{i=1}^{n}\epsilon_i\right)^2\right]\right\}$$

Differentiation begins simply by pushing the differential operator through the outer summation:

$$\frac{\partial E}{\partial \epsilon_i} = \frac{1}{m}\sum_{j=1}^{m}\frac{1}{n}\frac{\partial}{\partial \epsilon_i}\left\{\sum_{i=1}^{n}(T_{ij}^2 + 2T_{ij}\epsilon_i + \epsilon_i^2)\right. \qquad \text{(eq. 4)}$$
$$\left. - \frac{1}{n}\left[\left(\sum_{i=1}^{n}T_{ij}\right)^2 + 2\sum_{i=1}^{n}T_{ij}\sum_{i=1}^{n}\epsilon_i + \left(\sum_{i=1}^{n}\epsilon_i\right)^2\right]\right\}$$

The next step, and probably most crucial to understand clearly, is to realize that the derivative of the first inner summation will, in fact, eliminate the summation itself, leaving only two individual terms in ε_i − likewise the second summation vanishes. This is because if the summations are explicitly expanded (i.e., we list separate terms for each of $\varepsilon_1, \varepsilon_2, \varepsilon_3$, etc.), only the specific terms for index i result in a contribution to its specific derivative.

To illustrate:

$$\frac{\partial}{\partial \epsilon_i}\sum_{i=1}^{n}2T_{ij}\epsilon_i = \frac{\partial}{\partial \epsilon_i}(2T_{1j}\epsilon_i + \ldots + 2T_{ij}\epsilon_i + \ldots + 2T_{nj}\epsilon_n) \qquad \text{(eq. 5)}$$
$$= (0 + \ldots + 2T_{ij} + \ldots + 0) = 2T_{ij}$$

similarly:

$$\frac{\partial}{\partial \epsilon_i}\sum_{i=1}^{n}\epsilon_i^2 = \frac{\partial}{\partial \epsilon_i}(\epsilon_1^2 + \ldots + \epsilon_i^2 + \ldots + \epsilon_n^2) = 2\epsilon_i \qquad \text{(eq. 6)}$$

Likewise, for the derivative of the product of the two summations (the first being constant):

$$\frac{\partial}{\partial \epsilon_i}\left(\sum_{i=1}^{n}T_{ij}\sum_{i=1}^{n}\epsilon_i\right) = \sum_{i=1}^{n}T_{ij}\cdot\frac{\partial}{\partial \epsilon_i}\left(\sum_{i=1}^{n}\epsilon_i\right) \qquad \text{(eq. 7)}$$
$$= \sum_{i=1}^{n}T_{ij}\cdot\frac{\partial}{\partial \epsilon_i}(\epsilon_1 + \ldots + \epsilon_i + \ldots + \epsilon_n)$$
$$= \sum_{i=1}^{n}T_{ij}$$

and, using the chain rule and following similar reasoning,

$$\frac{\partial}{\partial \epsilon_i}\left(\sum_{i=1}^{n}\epsilon_i\right)^2 = 2\left(\sum_{i=1}^{n}\epsilon_i\right)\frac{\partial}{\partial \epsilon_i}\left(\sum_{i=1}^{n}\epsilon_i\right) = 2\sum_{i=1}^{n}\epsilon_i \qquad \text{(eq. 8)}$$

Note the significant difference between (6) and (8)!

Finally, all the other terms (consisting of expressions involving purely T_{ij}) go away directly, simply because they are constants with respect to the ε_i, leaving the final result:

$$\frac{\partial E}{\partial \epsilon_i} = \frac{1}{m}\sum_{j=1}^{m}\frac{1}{n}\left\{2T_{ij} + 2\epsilon_i - \frac{1}{n}\left[2\sum_{i=1}^{n}T_{ij} + 2\sum_{i=1}^{n}\epsilon_i\right]\right\} \qquad \text{(eq. 9)}$$
$$= \frac{1}{m}\frac{1}{n}2\sum_{j=1}^{m}\left\{\epsilon_i - \frac{1}{n}\sum_{i=1}^{n}\epsilon_i + T_{ij} - \frac{1}{n}\sum_{i=1}^{n}T_{ij}\right\}$$

When we then set these derivatives to zero, we can multiply out the initial constants and end up with the following system of equations in ε_i:

$$0 = \sum_{j=1}^{m}\left\{\epsilon_i - \frac{1}{n}\sum_{i=1}^{n}\epsilon_i + T_{ij} - \frac{1}{n}\sum_{i=1}^{n}T_{ij}\right\} \qquad \text{(eq. 10)}$$

Noting that the outer summation over m simply creates m multiples of the ε_i terms, we have:

$$0 = m\left(\epsilon_i - \frac{1}{n}\sum_{i=1}^{n}\epsilon_i\right) + \sum_{j=1}^{m}\left(T_{ij} - \frac{1}{n}\sum_{i=1}^{n}T_{ij}\right) \qquad \text{(eq. 11)}$$

or, moving the ε_i to the left side of the equation, dividing through by m, and recombining the T_{ij} terms slightly:

$$\epsilon_i - \frac{1}{n}\sum_{i=1}^{n}\epsilon_i = \frac{1}{m}\frac{1}{n}\sum_{j=1}^{m}\sum_{i=1}^{n}T_{ij} - \frac{1}{m}\sum_{j=1}^{m}T_{ij} \qquad \text{(eq. 12)}$$

A careful inspection of the two terms now on the right leads to the realization that the first (double) summation is exactly the overall average of the entire original set of unshifted values (that is, a *meta−average* over all devices and over all time steps). The second term is the average of the set of unshifted values for device i alone (over those same time steps of interest). Thus, with hopefully obvious symbolism, we could write:

$$\epsilon_i - \frac{1}{n}\sum_{i=1}^{n}\epsilon_i = \overline{\overline{T}} - \overline{T}_i \qquad \text{(eq. 13)}$$

a rather simple statement of the solution considering what we've gone through to it! Incredibly, it gets even simpler (though we'll have to backtrack a little before all implications have been fully studied).

Recall our original stipulation that the averages of the individual data sets do not move, in spite of the offsets. Stated mathematically, we thus require:

$$\frac{1}{m}\sum_{j=1}^{m}\frac{1}{n}\left[\sum_{i=1}^{n}(T_{ij}+\epsilon_i)\right]=\frac{1}{m}\sum_{j=1}^{m}\frac{1}{n}\left[\sum_{i=1}^{n}T_{ij}\right] \quad \text{(eq. 14)}$$

This is equivalent to the statement that the sums of the offsets must be zero (which can be seen by separating out the T_{ij}'s from the ε_i's, and subtracting the common terms from both sides of the equation), viz.:

$$\sum_{i=1}^{n}\epsilon_i = 0 \quad \text{(eq. 15)}$$

Plug that into Equation 13, and presto!, we have this final expression:

$$\epsilon_i = \overline{\overline{T}} - \overline{T}_i \quad \text{(eq. 16)}$$

This is a fully determinate set of equations for each of the ε_i , completely symmetric in i (as we should have expected), and they aren't even coupled to each other! Perhaps unbelievable, considering the completely general version of the solution back in (13). After all, we had i coupled equations in exactly i unknowns, and our intuition should tell us that if all the offsets are free to shift, there should be no unique solution which yields the minimum scatter. To put it another way, if you did find some set of offsets which resulted in the minimum scatter, and then you changed all the offsets together by some common amount, although the sets of shifted values would certainly move, the scatter *between* them would not change compared to what it was before the common shift. If this is so, how could it be that we so easily came up with (16), which gives no hint of indeterminacy?

To understand completely, we must revisit (13) in light of some fundamental principles of linear algebra. One of these is expressed as Theorem 3.1 (ref [4]):

> *If one system of equations can be obtained from another system by a sequence of elementary operations, then the two systems are equivalent.*

Further, one of the *elementary operations* that can be performed (in accordance with this theorem) is:

> *The i^{th} equation is replaced by the sum of the i^{th} and p times the j^{th} equation ($i \neq j$).*

By applying this elementary operation repeatedly to any one of the original equations, successively adding in to it all of the other equations (each with a multiplier of 1), we can create the following "secondary" operation:

> *The i^{th} equation is replaced by the sum of all the original equations.*

So consider what happens when we apply this operation to our original set. In fact, let's back up to (12), rather than (13), so as to skip the (merely convenient) intermediate definitions of the average values. We again remind ourselves that the summation of a constant argument over n times, reduces to simply n times the argument; therefore once an argument has been summed over a dependent index, re−summing it over that same index is equivalent to summing a constant argument. Thus:

$$\sum_{i=1}^{n}\left(\epsilon_i-\frac{1}{n}\sum_{i=1}^{n}\frac{1}{n}\epsilon_i\right)=\sum_{i=1}^{n}\left(\frac{1}{m}\frac{1}{n}\sum_{j=1}^{m}\sum_{i=1}^{n}T_{ij}-\frac{1}{m}-\sum_{j=1}^{m}T_{ij}\right)$$

$$\sum_{i=1}^{n}\epsilon_i-\frac{1}{n}\sum_{i=1}^{n}\sum_{i=1}^{n}\epsilon_i=\frac{1}{m}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{i=1}^{n}T_{ij}-\frac{1}{m}\sum_{i=1}^{n}\sum_{j=1}^{m}T_{ij}$$

$$\sum_{i=1}^{n}\epsilon_i-\frac{1}{n}\cdot n\sum_{i=1}^{n}\epsilon_i=\frac{1}{m}\frac{1}{n}\cdot n\sum_{j=1}^{m}\sum_{i=1}^{n}T_{ij}-\frac{1}{m}\sum_{i=1}^{n}\sum_{j=1}^{m}T_{ij}$$

$$\sum_{i=1}^{n}\epsilon_i-\sum_{i=1}^{n}\epsilon_i=\frac{1}{m}\left(\sum_{j=1}^{m}\sum_{i=1}^{n}T_{ij}-\frac{1}{m}\sum_{i=1}^{n}\sum_{j=1}^{m}T_{ij}\right)$$

$$0=\frac{1}{m}(0)$$

$$0=0 \quad \text{(eq. 17)}$$

Since elementary operations on our set of equations can be used to generate an identity, we can therefore state that the original set of equations was indeterminate. And our constraint (15), namely, that the sum of the offsets equal zero, can therefore be viewed as nothing more than supplying sufficient, linearly independent information, to turn the indeterminate set of $i-1$ equations (in i unknowns), into the fully determinate set derived as (16).

The "Toy" Problem

We are now ready to return to the hypothetical data presented in Tables 1 and 2. In Table 1, m is but 1. Now (16) confirms what was already obvious, to wit: for minimum scatter, each device measurement needs to be shifted by the difference between its raw value, and the average of the entire set of devices. That's exactly what we did, and of course, the resulting scatter was identically zero. Indeed, this will be the case anytime there is no variation over time, regardless of how large m is. For the more complicated case of the Table 2 data, however, we could see that we didn't get the optimal scatter simply by using the same offsets as were trivially obvious for the data of Table 1. In Table 5, it is now evident that the offsets that were used to generate the shifted data of Table 4 are those arising from (16); therefore they produce the optimal, minimum scatter for this data.

Applying these shifts returns us immediately to the results of Table 4. Note that the "average" shift is zero, which of course also means that the sum of the shifts is zero, as it should be.

There is one remaining point to be made. Recall that (16) was derived beginning with the definition of *population* variance. It will be left as an exercise for the reader to prove that the same optimal shifts arise if the formula for the *sample* variance had been used, where $n-1$ appears in the denominator of the definition, rather than n.

Table 5. Averages and Resulting Optimal Offsets Using (16)

Time	Device 1 °C/W	Device 2 °C/W	Device 3 °C/W	Device 4 °C/W	Device 5 °C/W	Average °C/W
0.001	1.0	2.0	3.0	4.0	5.0	3.0
0.002	1.0	1.5	2.0	2.5	3.0	2.0
Avg	1.0	1.75	2.5	3.25	4.0	2.5
Shift	1.5	0.75	0	−0.75	−1.5	0.0

Application to Real Data

We now apply the scatter minimization technique so succinctly expressed as (16) to the real data originally shown in Figure 1. There are twenty individual curves. Ten represent "heated" channels (five of which are channel #1 on each test device, and five of which are channel #2 on each test device). Ten curves are the "unheated" channels, likewise split between the two channels of each of five test devices. Because the device is physically and geometrically symmetric between the two channels, through the following discussion we will in general always group together the five curves of channel 1 with the corresponding five curves of channel 2, based on their being heated or unheated. In some cases, we will average together all twenty curves of data.

What we should first like to do is simply minimize the scatter over our entire data set. Equation 16 says, essentially:

(a) compute the average value of each cooling curve; (b) compute the average of the averages, and then (c) shift each curve up or down according to how much its average deviates from the meta−average. Figure 2 results.

The data immediately becomes much more clear. The technique has obviously eliminated some sort of systematic "error" which causes one set of data to drift with respect to another. Even now, without having explicitly separated the heated channel data from the unheated channel data, we see a systematic separation of the curves. Before we deal with this issue, however, note the overall pooled standard deviation for the entire set of twenty curves: taking the standard deviation individually at each point in time, then pooling over the entire time history, we obtain 1.74°C/W. After we have minimized the scatter globally, this figure has been reduced to 0.98°C/W.

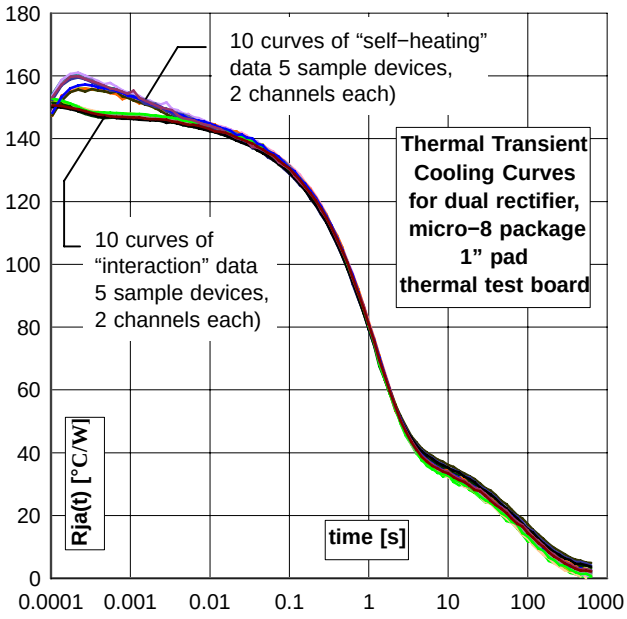

Figure 2. Scatter Minimized Over 0.0001−630 s

Next, we would like address the problem of the heated vs. unheated channels, and see just how different they really are from each other. We begin by making the assumption that all our devices are internally identical, and therefore “force” the minimum scatter very early in the data streams. We also recognize that the heated and unheated channels should *not* behave identically, thus we shall keep two separate meta−averages, one for the heated, and one for the unheated channel. From, say, 0.0005−0.001 s, there should be in theory only die, and possibly to a limited extent, die attach, material property contributions to the device thermal characteristics. In this data, we actually have five points along each curve in this time interval. If we calculate the pooled standard deviation of each subset of ten curves over those five time points, we obtain 2.19 and 1.85°C/W for the heated and unheated sets respectively. (It is also of interest to note the pooled standard deviations for the heated and unheated subsets over the entire time range from 0.0001−630 s, were 1.88 and 1.59. In other words, the data over the “short time” range is superficially somewhat more “noisy” than the overall data sets.)

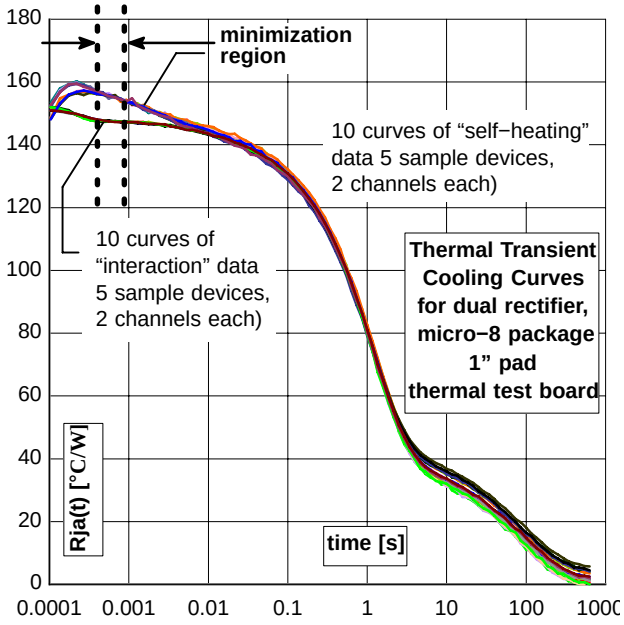

Figure 3. Scatter Minimized Over 0.0005−0.001 s

If we now apply Equation 16 to the five points of data of die timescale interest, we obtain the adjusted graph shown in Figure 3. The minimized pooled standard deviations for the heated and unheated curves have dropped dramatically to 0.32 and 0.12°C/W. The difference between the heated and unheated channels is now very clearly much more significant than the basic “noise” in the measurements might have suggested. Although we could have simply averaged the data sets and compared the averages, we would have had no assurance that the separation between the averages was statistically significant. (Indeed, the pooled standard deviations would have told us it was not.)

On the other hand, we have traded off clarity of short−time behavior for obfuscation of the long−time data. Let us therefore apply the error minimization technique to the data over the range of, say, 10 through 630 s, where our physical intuition (and the gross appearance of the plotted data) suggests that all five test boards (and both heated and unheated channels) ought to share identical behavior. Here again, all twenty curves are treated as a single group (rather than two individual sets of ten curves), and Figure 4 results. The pooled standard deviation over the time period of interest is only 0.31°C/W. Thus, the composite noise of the ten curves is actually modestly better than that of the individual channels in the first half millisecond.

Observing that the original unminimized scatter over the 10−630 s range was only 0.62 (considerably less than that of the entire data sets from 0.0001−630 s), we now have clear evidence that the devices behave very much as identically to each other during the short time scale, as do the test boards to each other at the long time scale. Yet the overall standard deviation of the entire data sets is much poorer than at either limited end range. Our conclusion, then, is that something between the “die” level and the “board” level is different between the various individual curves. We could not see this at the outset, because the variations that evidently are due to some intermediate timescale differences are compounded into the short−time data merely through the cooling curve experimental methodology.

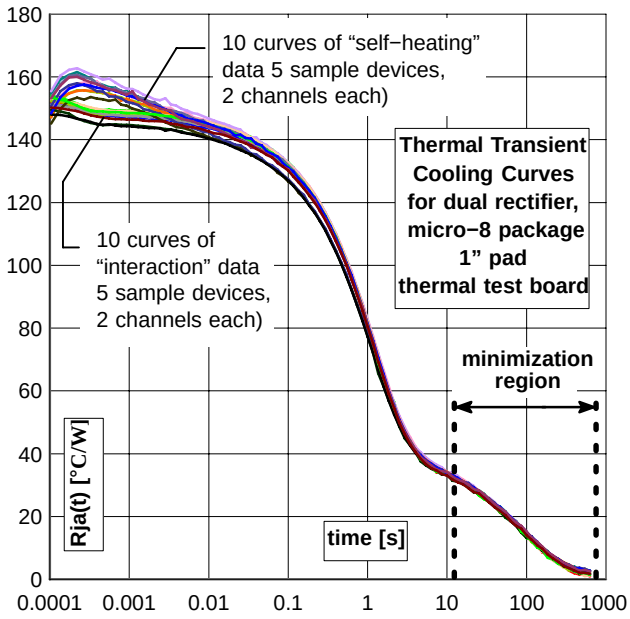

Figure 4. Scatter Minimized Over 10−630 s

Can we learn any more? So far, we believe from theoretical considerations that there should be a real difference in the heated vs. unheated families of curves at the short time, yet by the time steady state is reached, there should be no difference. The ability to minimize the scatter at the two extreme ends of the curves confirms this.

However, the differences due to package internal variation should not persist as late at 10 s into the cooling curves. Inspection of the curves, especially of Figure 3, suggests that it may be worthwhile to use the error minimization technique over the time scale of 0.1 − 0.2 s. Figure 5 is what results, again treating all twenty curves as a single group. This shows very clearly that if 0.1 s is used as a common baseline for all devices (irrespective of channel), the device−to−device variation does not really manifest itself until about 2 s. This is the timescale at which board mounting details contribute to the thermal performance of the system.

In Figure 6, we zoom in on the Figure 5 graph between 0.1 and 10 seconds, and observe that the four curves from sample #4 of the test group (two heated curves, one for each channel, and two unheated curves, one for each channel) split off from the others. If we were so inclined, we could then pull sample #4 from the batch and study it for physical explanations for the difference. In any case, even though the scatter was minimized only up through 0.2 s, the entire family of curves for all five samples is extremely tight out through about 2 s.

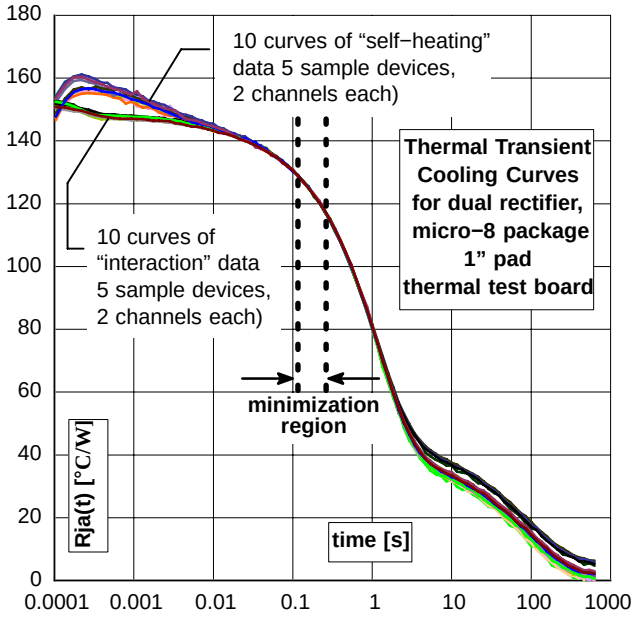

Figure 5. Scatter Minimized Over 0.1−0.2 s

Similarly, moving back to times earlier than 0.1 s, Figure 5 shows that the unheated channels seem to be more tightly grouped than the heated channels. Since we have already observed that the intrinsic scatter in these groups is very similar, this suggests that there is possibly a die−attach source of variation within the samples, which would influence the heated channels more than the unheated channels. Even more detailed inspection of the time scale prior to 0.0005 s yields evidence that there is a subtle difference between channel 1 and channel 2 of the samples; since only die material properties can affect thermal performance at this early in the cooling curve, we then focus our attention (if we are concerned), on design differences between the two channels, or possibly in the thermal cooling test hardware. Perhaps, for instance, there is a small difference in inductance in the test leads between the two channels.

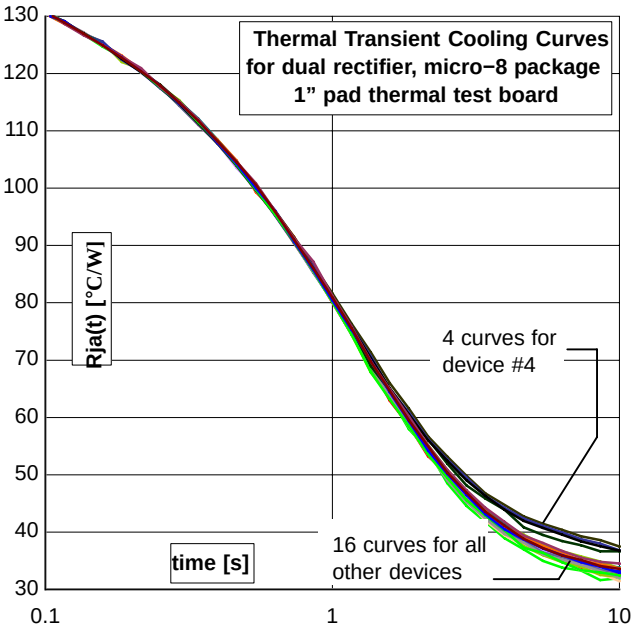

Figure 6. Board Mounting Timescale (0.1−10s of Figure 5)

Further Discussion

It may seem at this point that all we have done in this demonstration of the scatter minimization technique, is arbitrarily select different time periods over which to apply the mathematics developed at the outset. Certainly the “toy” problem had no particular rationale behind it other than the simple exposition of the method. In the actual semiconductor device cooling data used as the real−life illustration, however, the time scales selected were driven ultimately by the real physics behind the data. Fundamental considerations of thermal diffusion time for the materials present in the package, tell us that certain interactions are possible, and others are not possible. For instance, based on the thermal diffusivity of silicon and the known thickness of the die, we can be confident that local thermal disturbances at the surface of the device cannot possibly propagate to the back surface of the die (where it interfaces with the next major component of the package, the leadframe), at a time scale earlier than a couple of milliseconds. Similarly, based on the diffusivity of the leadframe material and the lead lengths, we can be certain that disturbances cannot be propagated out the leads earlier than a few tens of milliseconds. We also find, based on molding compound properties and its overlying thickness, that the heat conduction upwards from the silicon surface (and indeed, outwards from every metal−to−compound interface internal

to the package) is very slow in comparison, hence the influence of the exterior convecting surfaces of the case does not come into play so early in the transient curve. On the other hand, we can see that separate junction regions on the surface of the silicon *will* respond to each other on approximately that time scale.

Looking at it from the other direction, we can argue that since thermal disturbances *can* propagate through the die and leadframe on the order of tens of milliseconds or faster, then at time scales an order of magnitude longer, the package must be at local internal thermal equilibrium. Therefore any macroscopic changes in the temperature of the system (even though measured at the junction) must be driven by the external environment, beginning at the interface between the leads and the board, and continuing on to incorporate the board properties and convection effects. These domains also lend themselves to rough estimates of when they may and may not be significant. For instance, though it was not dwelt on in this analysis, the distinct "knee" in the cooling curves at about 5 s corresponds well with the estimated thermal diffusion time of the 1" square copper heat−spreader plated on to the thermal test circuit board. Likewise, other rough estimates of thermal time constants of system components (lumped RC analysis, for example, see [5]) may also provide important independent guidance for the selection of time scales of interest.

Finally, when one compares the data of one test sample vs. another, or vs. the average of the entire suite of data, knowing the time scales at which different effects may come into play permits one to identify whether the sample has a specific defect and should perhaps not be included in the average. (The so called "delta−V_{be}" tests for die−attach integrity, well known in the semiconductor manufacturing business, are a good illustration of the diagnostic power of associating certain time scales with particular causes.) Naturally, whether a specific sample *is required* to be included in an average has more to do with the purpose behind the data collection in the first place, and clearly is outside the scope of this monograph.

What must be understood here is that the scatter minimization technique cannot magically sift out totally random experimental variation, and as with most any statistical tool, it may be abused and improperly applied. Coupled with independent knowledge of what *should* be happening in the system, however, one can intelligently sort out the likely sources of the variation, and thus optimally extract the information intrinsically present, although obscured, in the actual data.

SUMMARY AND CONCLUSION

The scatter−minimization technique developed in this paper has been demonstrated to yield useful information about thermal transient experimental results that would otherwise remain completely obscured by ordinary measurement variability. For instance, package to package differences may be brought out clearly (whether internal to the package, or, as was the case in the data used for illustration here, external to the package at the board−mounting interface). Even more significantly, because thermal transient tests intrinsically provide information about different physical components of the thermal system, from test equipment, to die attach, to board mounting, to board structure, the error minimization technique allows the experimenter to probe the raw data for very specific information about the sources of the variation.

ACKNOWLEDGMENTS

The author would like to acknowledge the dedicated efforts of the staff of the ON Semiconductor Thermal Characterization Lab, specifically Ron Lawson, without whom we would have no test equipment with which to perform high quality thermal transient cooling tests, nor data to analyze.

REFERENCES

[1] J.W. Worman, "Sub−millisecond Thermal Impedance and Steady State Thermal Resistance," 15th Annual SEMI−THERM™ Symposium, IEEE, 1999.

[2] JEDEC Standards No. 24−3, 24−4, 51−1, Electronic Industries Alliance, 1990.

[3] MIL−STD−883E, Method 1012.1, U.S. Department of Defense, 31 December 1996.

[4] Ben Nobel, *Applied Linear Algebra*, Prentice Hall, 1969.

[5] J.P. Holman, *Heat Transfer*, McGraw−Hill, 1969.

ON Semiconductor and ON are registered trademarks of Semiconductor Components Industries, LLC (SCILLC). SCILLC reserves the right to make changes without further notice to any products herein. SCILLC makes no warranty, representation or guarantee regarding the suitability of its products for any particular purpose, nor does SCILLC assume any liability arising out of the application or use of any product or circuit, and specifically disclaims any and all liability, including without limitation special, consequential or incidental damages. “Typical” parameters which may be provided in SCILLC data sheets and/or specifications can and do vary in different applications and actual performance may vary over time. All operating parameters, including “Typicals” must be validated for each customer application by customer’s technical experts. SCILLC does not convey any license under its patent rights nor the rights of others. SCILLC products are not designed, intended, or authorized for use as components in systems intended for surgical implant into the body, or other applications intended to support or sustain life, or for any other application in which the failure of the SCILLC product could create a situation where personal injury or death may occur. Should Buyer purchase or use SCILLC products for any such unintended or unauthorized application, Buyer shall indemnify and hold SCILLC and its officers, employees, subsidiaries, affiliates, and distributors harmless against all claims, costs, damages, and expenses, and reasonable attorney fees arising out of, directly or indirectly, any claim of personal injury or death associated with such unintended or unauthorized use, even if such claim alleges that SCILLC was negligent regarding the design or manufacture of the part. SCILLC is an Equal Opportunity/Affirmative Action Employer. This literature is subject to all applicable copyright laws and is not for resale in any manner.

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.