Memory Thermal Management 101

Overview

With the continuing industry trends towards smaller, faster, and higher power memories, thermal management is becoming increasingly important. Not only are device sizes shrinking, but the PC boards they mount onto are also shrinking. Placing devices closer and closer together helps lower overall system size and cost and improves electrical performance, but increases the “power density”, which can lead to higher device temperatures and, thus, to lower component reliability.

This paper describes GSI Technology’s use of standardized thermal resistance data, methods for assessing a device’s critical thermal parameters in a system application, and how thermal parameters are related to reliability.

Junction Temperature—Definitions and Uses

Junction temperature (T_J) is the term used to describe the temperature of the die inside a semiconductor product package. It is typically higher than the temperature of the surrounding package (“case”) and the device’s exterior environment. T_J is typically expressed in ^{0}C.

Two T_J values are in most device datasheets. The first, and more conservative, value is found in the **Recommended Operating Conditions** section. The Junction Temperature specification in the Recommended Operating Conditions section describes the range of die temperatures over which the device is guaranteed to meet all datasheet performance specifications and demonstrate no worse than a specified failure rate throughout normal operating life.

Operating Temperature

Parameter	Symbol	Min.	Typ.	Max.	Unit
Junction Temperature (Commercial Range Versions)	T_J	0	25	85	°C
Junction Temperature (Industrial Range Versions)*	T_J	–40	25	100	°C

Note:
* The part numbers of Industrial Temperature Range versions end with the character “I”. Unless otherwise noted, all performance specifications quoted are evaluated for worst case in the temperature range marked on the device.

Another T_J value can be found in the **Absolute Maximum Ratings** section of the datasheet and is used when conducting device reliability stresses.

Absolute Maximum Ratings
(All voltages reference to V_{SS})

Symbol	Description	Value	Unit
V_{DD}	Voltage on V_{DD} Pins	–0.5 to 2.9	V
V_{DDQ}	Voltage in V_{DDQ} Pins	–0.5 to V_{DD}	V
V_{REF}	Voltage in V_{REF} Pins	–0.5 to V_{DDQ}	V
$V_{I/O}$	Voltage on I/O Pins	–0.5 to V_{DDQ} +0.5 (≤ 2.9 V max.)	V
V_{IN}	Voltage on Other Input Pins	–0.5 to V_{DDQ} +0.5 (≤ 2.9 V max.)	V
I_{IN}	Input Current on Any Pin	+/–100	mA dc
I_{OUT}	Output Current on Any I/O Pin	+/–100	mA dc
T_J	Maximum Junction Temperature	125	°C
T_{STG}	Storage Temperature	–55 to 125	°C

Note:
Permanent damage to the device may occur if the Absolute Maximum Ratings are exceeded. Operation should be restricted to Recommended Operating Conditions. Exposure to conditions exceeding the Recommended Operating Conditions, for an extended period of time, may affect reliability of this component.

It is important to understand that the two specifications of T_J are used for distinctly different purposes. A memory device operating outside of its Recommended Operating Conditions range may be functional for a short period of time. But it will exhibit increased risk of reliability-related failures with accumulated usage**.**

What is Thermal Resistance and How to Find the Specifications

A Thermal Resistance value is analogous to an Electrical Resistance value. It indicates how well or how poorly the thermal path in question can conduct heat energy. Objects or materials with high Thermal Resistance are insulators—they tend to block heat flow. Conversely, materials or objects with low Thermal Resistance, such as a piece of copper, conduct heat very well. A Thermal Resistance value indicates the temperature difference between a die junction and a given reference point (i.e., the case top, the PC board or ambient air) for each unit of power dissipated at the die surface.

Thermal resistance values, along with the power dissipation (P_D), package case temperature (T_C), and adjacent board temperature (T_B) are used to predict the junction temperature (T_J) of a device in a system environment.

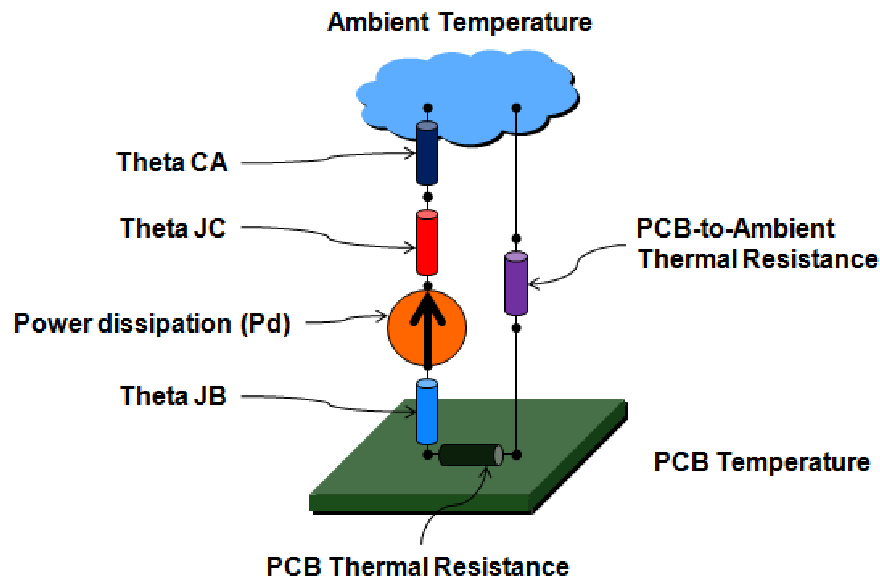
Thermal resistance values are expressed in ^{0}C/Watt. They can be found in most product datasheets. See below for an example.

Thermal Impedance

Package	Test PCB Substrate	θ JA (C°/W) Airflow = 0 m/s	θ JA (C°/W) Airflow = 1 m/s	θ JA (C°/W) Airflow = 2 m/s	θ JB (C°/W)	θ JC (C°/W)
165 BGA	4-layer	22.300	18.572	17.349	9.292	2.310

Notes:
1. Thermal Impedance data is based on a number of of samples from mulitple lots and should be viewed as a typical number.
2. Please refer to JEDEC standard JESD51-6.
3. The characteristics of the test fixture PCB influence reported thermal characteristics of the device. Be advised that a good thermal path to the PCB can result in cooling or heating of the RAM depending on PCB temperature.

The Main Thermal Metrics



θJA (Thermal resistance—Junction to Ambient): This value describes the ability of a package to conduct heat from the die to the surrounding air and is measured using a method intended to minimize the effects of other thermal pathways to the surrounding environment.. It is defined as the difference between T_J and T_A per watt of dissipated power.

$$\text{Theta } J_A = \frac{T_J - T_A}{P_D}$$	$T_J = T_A + (\text{Theta } J_A * P_D)$

It is important to note that θJ_A values are only expected to be valid in a specific JEDEC-specified test environment—not in an actual application where other thermal pathways are likely to dominate die temperature. (In other words, θJ_A cannot be used to predict die temperature in an actual use environment.)

θJ_A is typically measured with laminar airflow over the device at speeds of of 0, 1, and 2 meters/second. Air turbulence can dramatically alter the effectiveness of airflow cooling in real systems and detract further from the usefulness of datasheet θJ_A values for simplistic thermal analysis efforts.

θJ_C (Thermal resistance—Junction to Case): This value measures the ability of a package to conduct heat from the die to the "top" surface of the package. (The "top" of the package is generally understood to be the opposite side of the package from the interconnect plane.) It is defined as the difference between T_J and T_C per watt of dissipated power. θJ_C is very useful because it can be used reliably regardless of what is happening in the environment around the device. It describes a characteristic of the package itself.

$$\text{Theta } J_C = \frac{T_J - T_C}{P_D}$$	$T_J = T_C + (\text{Theta } J_C * P_D)$

θJ_B (Thermal resistance—Junction to Board): The most important pathway for heat to leave a package is through the board unless a heat sink is attached to the device. θJ_C captures the resistance of the junction-to-case pathway, but as the θJ_A thermal resistance number suggests, the Case-to-Ambient thermal path for most devices is quite poor. Obviously the Junction-to-Ambient thermal path is Junction-to-Case-to-Ambient. So, following the resistance analogy suggested earlier, θJ_A is the sum of θJ_C and θC_A, where θC_A is the Case-to-Ambient thermal resistance. Since the Junction-to-Case resistance is generally rather low, it follows that most of the Junction-to-Ambient resistance is the Case-

to-Ambient component. So, unless a heat sink is attached to the RAM, as a rule, the best (lowest resistance) thermal path away from the RAM is the connection to the PC board. The thermal resistance path to the PC board is called the Junction-to-Board Resistance (or θJ_B). This value measures the ability of the package to conduct heat away from the die and into the PCB. However, it also describes the path heat in the board takes to the die.

One of the key objectives of thermal analysis of a design is to determine whether the RAM will be heating the board or whether the board, due to the heat introduced by adjacent devices, is heating the RAM.

$$Theta\ J_B = \frac{T_J - T_B}{P_D}$$	$$T_J = T_B + (Theta\ J_B * P_D)$$

In most cases, calculating T_J from θJ_B is a more realistic analysis of a device's thermal performance.

Calculating Power and Junction Temperature

Let's review an example using GSI's 2Mb x 36 SigmaQuad-II Burst of 4 SRAM, operating at 333 MHz, in a 165-lead, 13 x 15 mm LBGA package (GSI part number **GS8662Q36BD-333**).

First, let's estimate the power dissipation (**P_D**) in a commercial temperature range (0^0C–70^0C) application. P_D should include core power and I/O switching power.

Core power is a straightforward formula using a nominal supply voltage (V_{DD}) and operating current at the target frequency (IDD). GSI specifies a worst-case IDD value for each speed bin. The specifications increase at a roughly linear rate vs. clock frequency.

Supply Voltage (V_{DD} nominal) =	**1.8 V**
Operating Current (IDD) @ 333 MHz =	**1100 mA**
Core power = V_{DD} * IDD @ =	**1.98W**

I/O switching power is a function of the core frequency (F), capacitive loading (CL), voltage swing (V), the number of I/Os switching simultaneously (N), plus a "data rate factor" which estimates the ratio of I/O toggling frequency to clock frequency (D).

Capacitive loading is a function of the memory I/O driver's capacitance, the capacitance of the device on the other end of the PC board trace, and the length of the trace itself. For most point-to-point connections, an estimate of 10pf can be used with reasonable confidence.

"D" is a data rate factor of the memory architecture being used. For example, D = 0.5 for single data rate devices such as Burst SRAMs and No Bus Turnaround SRAMs; D = 1.0 for double data rate devices, such as SigmaQuad and SigmaDDR SRAMs.

I/O power	= D * F * CL * V^2 * N
	= 1.0 * 333*10^6 * 10*10^-12 * 1.8^2 * 36
	= <u>0.39W</u>

So the total $\mathbf{P_D}$ can be estimated as 1.98W + 0.39W = **<u>2.37W</u>**.

Next, let's determine a die junction temperature. From the datasheet, we see this device has a T_J max of **85^0C** for commercial temperature applications.

Next, let's evaluate the device from a junction-to-case perspective:

T_J = T_C + (<u>2.31</u>^{0}C/W x 2.37W) = T_C + 5.47^{0}C,

Therefore, the maximum package (case) temperature needs to be

T_C = T_J – 5.47^{0}C = **<u>79.5</u>**^{0}C.

Next, let's look at device performance when compared to the PC board. The board temperature at the memory device's location (T_B) is a function of the adjacent components, the capacity of the PC board traces leading away from the memory, and many other factors that are available only via thermal modeling or direct measurement. The goal, therefore, is to find a max PC board temperature that will support the memory's T_J.

T_J = T_B+ (9.29^{0}C/W x 2.37W) = T_B+ 22.02^{0}C

Therefore, the maximum board temperature at the memory needs to be:

T_B = T_J – 22.02^{0}C = **<u>62.98</u>**^{0}C

The die junction to board thermal resistance (θJ_B) is the most reliable predictor of a memory device's thermal performance in a system.

Measurement Techniques

The most common measurement tool is the thermocouple. If a thermocouple is used, some guidelines include using small gauge thermocouple wires to reduce any heat sinking effects of the wire; monitoring as many points as practical, and placing probes as close as practical to the die.

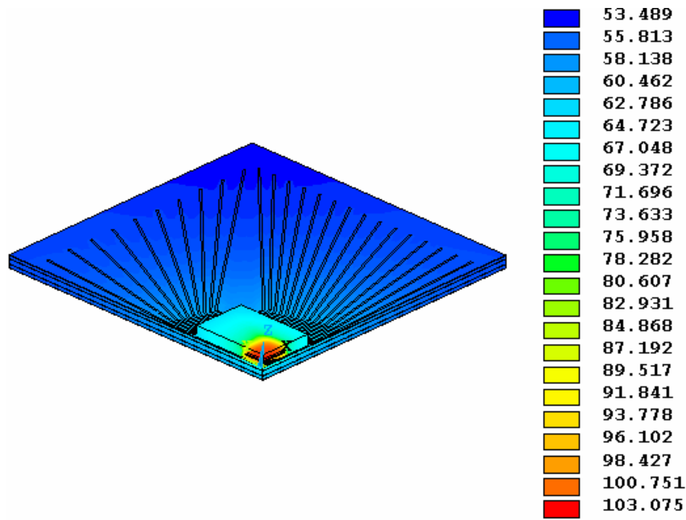
Infrared scanning is a useful technique because it provides an analysis across an entire scanned area. Discovering unexpected hot spots with an infrared scan are very helpful during prototyping when an assembly problem may otherwise go unnoticed. One disadvantage is that the camera must have access to the PCB under test, which could significantly alter its behavior. Handheld, contactless thermometers are inexpensive and easy to use tools for taking spot measurements.

Thermal Simulation Software Options

The calculations presented in the previous section will provide an improved estimate of thermal performance, but only an estimate. More accurate results will require the use of thermal simulation techniques.

There are several commercially available software packages, each with its advantages and disadvantages, with the most important being the numerical method used for solving the simultaneous equations. The most common method is called Computational Fluid Dynamics (CFD). It is most often used because it predicts fluid flow, which is necessary for modeling convection, in addition to calculating conduction and radiation factors.

The two most common thermal analysis packages are provided by ANSYS (CFX, Fluent, Iceboard and Icepak) and Mentor Graphics (Flowtherm). The result in graphical form will often resemble the following diagram:



source: www.aseglobal.com

Thermal resistances can be characterized using a JEDEC standard methodology (JESD51-7 for surface mount packages, or JESD51-9 for BGAs) to provide consistent results. An example is shown below.

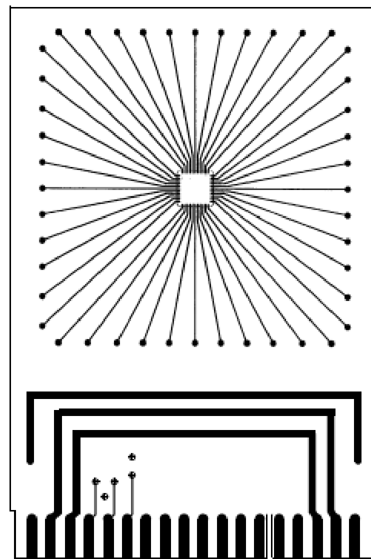


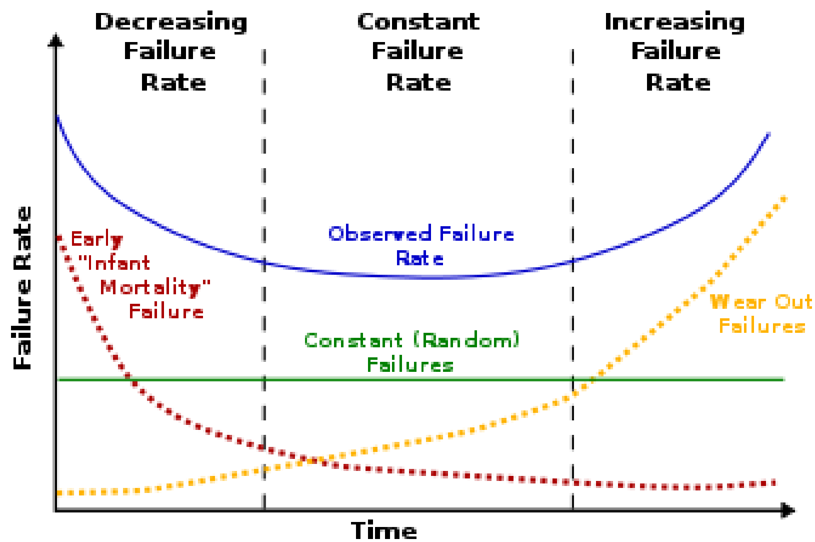
FIGURE 1. TOP VIEW - TYPICAL TEST BOARD

source: www.jedec.org

Temperature and Reliability

Semiconductor device lifetimes are often described using a graphical representation called a “bathtub curve” (see the below diagram). The bathtub curve consists of 1) an infant

mortality period with a decreasing failure rate; 2) a normal lifespan with a low, relatively constant failure rate, and 3) a wear-out period that exhibits an increasing failure rate.



source: en.wikipedia.org/wiki/Bathtub_curve

The bathtub curve **does not** depict the failure rate of a single item, but describes the relative failure rate of an **entire population** of products over time.

Failures during infant mortality are caused by defects and human error: material defects, design issues, assembly issues, etc. When a new product is being prepared for introduction, an Early Failure Rate study (EFR) is conducted to establish the burn-in conditions necessary to drive weak devices to failure during the manufacturing process, before they are shipped to customers. The exact conditions used for the EFR study are a function of the technology used to manufacture the device in question, but typically high temperatures and high voltages are applied to the device, each of which tend to accelerate failures in time. Typically 1200 devices representing three different manufacturing lots are used in the study. The devices are typically stressed at 125^{0}C for 48 hours at an elevated voltage appropriate for stressing the transistors used to manufacture the device. The results of the test are then used as the basis for the design of the Burn-In stress conditions that will be applied to the device prior to final testing that will be used in the course of normal manufacturing of the device.

The Burn-In and test regimen employed as part of the device manufacturing process is designed to weed out virtually all infant mortality failures. Or to put it another way, burn-in

and test is designed to drive the device into the Constant Failure Rate portion of the reliability curve before it is shipped to a customer.

Once Infant Mortality failures have been driven out of a population of devices, they should demonstrate a virtually constant random failure rate until the devices reach the Wear-Out phase of their life cycle. Useful Life is typically regarded as 10 years of normal usage. So before a device is put into normal production, long term reliability testing is conducted to verify that a population of devices that have already been exposed to Burn-In stress and Final Test will demonstrate a nominal long-term failure rate of 50 FITs (1 FIT is 1 failure per 1 billion device-hours of operation)…until the devices wear out. (Yes, semiconductors do wear out…very slowly.)

High Temperature Operating Life Test (HTOL) is designed to accelerate failure mechanisms that are activated by temperature stress and voltage stress. It is used to predict *long-term failure rates* because failure acceleration due to temperature and voltage are well understood. A typical HTOL test looks like the following:

Test Name	Conditions	Sample Size	Operation
HTOL	125^0C for 1000 hours; V_{DD} = Absolute Maximum Voltage	3 wafer lots; 105 devices / lot	Test devices before and after stress (Room, Cold, Hot Temp)

Once the data from the test is in, it is converted into a long term failure rate forecast using the Arrhenius equation and chi-square distribution function. If the forecast shows the devices in the study would have had a long term failure rate of 50 FITs or less under normal use conditions, and has passed a battery of other tests not relevant to this discussion, the device is released ("Qualified") for normal production. But since this paper is about thermal issues, it is important to note what happens to those 315 parts that went into the long term test. They go into an archive. They are NOT shipped to a customer. Why? Because they are now much too close to being worn out.

Note that the accelerated test conditions used for the HTOL test are essentially the conditions shown in the Absolute Maximum section of the device datasheet. The HTOL test is 1000 hours long…about 6 weeks. The message is pretty simple. A silicon device manufacturer does not expect a device exposed to Absolute Maximum conditions for 6 weeks to last too much longer. If the device is operated within Recommended Operating

Conditions, 10 years of useful life can be expected before Wear-Out failure mechanisms begin to surface…and during that 10 years of a population of the devices can be expected to demonstrate a nominal failure rate of 50 or fewer FITs.

Thermal characterization of a board allows verification that the semiconductor devices on the board are operating within their manufacturer's Recommended Operating Conditions; a necessary step toward verifying that the board will meet long term reliability expectations in the field.