Thermal Management of Die Stacking Architecture That Includes Memory and Logic Processor

Bhavani P. Dewan-Sandur, Abhijit Kaisare and Dereje Agonafer
The University of Texas at Arlington,
Box 19018, TX 76019
E-mail: dewan@mech.uta.edu

Damena Agonafer and Cristina Amon
Institute for Complex Engineered Systems,
Carnegie Mellon University

Senol Pekin and Terry Dishongh
Intel Corporation
2111 NE 25th M/S JF3-377
Hillsboro, Oregon 97124

Abstract

The convergence of computing and communications dictates building up rather than out. As consumers demand more functions in their hand-held devices, the need for more memory in a limited space is increasing, and integrating various functions into the same package is becoming more crucial. Over the past few years, die stacking has emerged as a powerful tool for satisfying these challenging Integrated Circuit (IC) packaging requirements.

Previously, present authors reported on the thermal challenges of various die stacking architectures that included memory (volatile and non-volatile) only. In this paper, the focus is on stacking memory and the logic processor on the same substrate. In present technologies, logic processor and memory packages are located side-by-side on the board or they are packaged separately and then stacked on top of each other (Package-on-package [PoP]). Mixing memory and logic processor in the same stack has advantage and challenges, but requires the integration ability of economies-of-scale. Geometries needed were generated by using Pro/Engineer® Wildfire™ 2.0 as a Computer-Aided-Design (CAD) tool and were transferred to ANSYS® Workbench™ 10.0, where meshed analysis was conducted. Package architectures evaluated were rotated stack, staggered stack utilizing redistributed pads, and stacking with spacers, while all other parameters were held constant. The values of these parameters were determined to give a junction temperature of 100°C, which is an unacceptable value due to wafer level electromigration. A discussion is presented in what parameters need to be adjusted in order to meet the required thermal design specification. In that light, a list of solutions consisting of increasing the heat transfer co-efficient on top of the package, the use of underfill, improved thermal conductivity of the PCB, and the use of a copper heat spreader were evaluated. Results were evaluated in the light of market segment requirements.

Keywords: Stack package, Thermal resistance, Logic processor, Memory.

Nomenclature and Subscripts

h: Heat transfer coefficient ($\text{W}/\text{m}^2\text{°C}$).
k: Thermal conductivity ($\text{W}/\text{m}^2\text{°C}$).
P: Power dissipation (W).
Q: Heat flux (W/m^2).
T: Temperature, ($^{\circ}\text{C}$).
R: Thermal resistance, ($^{\circ}\text{C}/\text{W}$)

a: Ambient
j: Junction

I Introduction

Present day market demands shorter product life cycles with increased product complexity. The union of broadband communications and digital technology has increased product opportunities [1].

Integrated Circuits (IC's) have encompassed almost all aspects of modern technology. Wireless communication and computing products constitute over 75% of the world's semiconductor consumption. Consumer electronics market is demanding reduction in size, weight, cost and power. This cannot be met without technical innovation. One of the enabling technologies is stacked packaging [2]. Integrating chips vertically in a single package multiplies the amount of silicon that can be crammed in a given package footprint and more silicon functions per cubic centimeter (cm) of application space, conserving system-board real estate.

Stacked packaging can be carried out using two or more dice within a single package, or by stacking and connecting completed packages (Expansion in Z-direction). A stacked package has significant advantages over conventional packaging with regard to size (40 – 50 % reduction), volume (5-6 times reduction compared to Multi-Chip-Modules [MCM]), weight (2-13 times reduction compared to MCM), silicon efficiency (100 % as compared to 20 - 90% in case of MCM), delay, noise, power consumption, and greater speed [3]. At the same time, stack package enables shorter routing of interconnects from chip to chip, which speeds signaling between them. Another benefit is the simplification of surface-mount system-board assembly because fewer components must be placed on the board. Figure 1 shows the die stacking trend [4].

Figure 1: Die Stacking Trend [4]

Some of the major challenges involved in realizing stacked packaging are modeling and simulation of the thermal, mechanical, and electrical performance of the complete chip, package & heat dissipation structures as a single system [2]. Previous work has been done on investigating the modeling and thermal simulation of various stacked packages. These included package-on-a-package (PoP), package-in-a-package [PiP], multi-chip-modules (MCM), and multi-chip-package (MCP). These involved mainly memory dice or memory and logic dice packaged separately and stacked on top of one another [5, 6, 7, 8, 9, 10, 11, and 12]. This paper concentrates on the thermal challenges of stacking memory and the logic processor on the same substrate. Three different stacking architectures were evaluated, viz. [a] Rotated stack, [b] Staggered stack utilizing redistribution pads, and [c] Stacking with spacers. Based on the maximum junction temperature, thermal management strategies that involve increasing heat transfer co-efficient on top of the package, the use of underfill, increasing the PCB core thermal conductivity, and use of copper heat spreader are presented.

II Modeling Procedure

A molded Ball Grid Array [BGA] stacked package is considered for the analysis. The substrate is 13x15 mm and 0.21 mm thick. A fully populated solder ball matrix with a ball count of 399 and a pitch of 0.6 mm is used. The stand off height after reflow is 0.2 mm. The mold cap is 0.90 mm thick and has the same dimensions as the package substrate. Both memory and logic processor modules were stacked on the same substrate.

Three different package architectures were modeled, viz. [a] Rotated stack, [b] Staggered stack utilizing redistribution pads, and [c] Stacking with spacers. The memory module was made up of four dice. Three non-volatile dice (Die 1, Die 2, and Die 3) measuring 4x6 mm, with a thickness of 0.1 mm, form the stack. The bottom volatile memory is made of a die measuring 9x13 mm, with a thickness of 0.1 mm. The logic die measures 12x12 mm, with a thickness of 0.1 mm, and is placed below the memory dice. In the rotated stack, Die 2 was placed at 90° to Die 1 and Die 3. In the staggered stack, Die 2 was offset by 1 mm in the X and Y axes of dice 1 and 3. For the spacer stack, three dummy silicon dice measuring 3x5 mm, with a

thickness of 0.03 mm were placed in between dice 1, 2, and 3. The overall package thickness was not to exceed 1.5 mm. The dimensions are not specific to a particular package. They are based on values found in present market i.e. a typical molded BGA stack package.

The Printed Circuit Board (PCB) for the 5-Die molded BGA stack package was modeled according to the Joint Electron Device Engineering Council (JEDEC) standard EIA/JESD51-5. It is made up of two signal and two power layers (2S2P). The four layered PCB includes 0.070mm thick upper and lower trace layers, 0.035mm thick internal power planes, and 0.45mm thick core layers. The 13x15mm BGA package was positioned at the center of the 76x76x1.57 mm 2S2P PCB. The detailed view of one of the configurations is as shown in Figure 2 [12, 13].

Figure 2: Detailed view of package and PCB (Mold cap hidden)

Table 1 shows the package dimensions, and Table 2 shows the material properties of the components.

Table 1: Package Dimensions

Component	Dimension (mm)
Memory - Die Stack	4 x 6 x 0.1
Memory - Bottom Die	9 x 13 x 0.1
Logic Die	12 x 12 x 0.1
Die Attach Thickness	0.0318
Mold Cap	13 x 15 x 0.9
Substrate	13 x 15 x 0.21
Substrate Mask Thickness	0.0175
Solder Ball Diameter	0.35
Stand-off Height (After reflow)	0.2
Solder Ball Pitch	0.6
PCB	76 x 76 x 1.57

The geometries were modeled as “Parts” using Pro/E® WildFire™ 2.0 as a CAD tool. Then, an “Assembly” was created for each of the stacking architectures. This ensured accurate material property assigning and material continuity of each component. Figure 3 shows the CAD model of the package and PCB. Figures 4, 5, and 6 show the CAD models of the individual architectures in the package.

Table 2: Material Properties

Material	Thermal Conductivity [W/m°C]
Die	120
Die Attach [Non-Conductive]	0.3
Substrate Mask	250
Substrate Core	0.3
Molding Compound	0.88
Solder	50.6
PCB Core [FR4]	0.18
Copper	393

Figure 3: CAD model – Package and PCB

Figure 4: Rotated Stack – PCB and mold cap hidden

Figure 5: Staggered Stack – PCB and mold cap hidden

Figure 6: Spacer Stack – PCB and mold cap hidden

The thermal resistance model of the assembly is shown in Figure 7. There are two paths for heat dissipation to the ambient. One is from the die stack, through the mold cap to the ambient. The second is from the die stack to the substrate, through the solder balls to the PCB, and through the PCB layers to the ambient. There are additional paths that are not that significant; for example, through the wirebonds.

About 80% of the heat dissipation is through the latter path, as shown in Figure 8. This is because of high thermal resistance of the interface between the die - mold cap, and the mold cap - ambient.

Figure 7: Thermal resistance model [14]

The efficiency of heat dissipation is largely dependent on the thermal resistance of the interface of the package - PCB (includes thermal resistance of the interface for PCB - ambient). This can be represented as:

$$T_j = P * (R_{pkg} + R_{pcb}) + T_a$$

$$R_{pkg} = \Sigma (R_{dice}, R_{datch}, R_{ovrmld}, R_{subs})$$

Figure 8: Heat dissipation path [15]

Given the maximum junction temperature, $[T_j]_{\max}$, the maximum power that can be dissipated through the package and the PCB to the ambient air is determined by,

$$P_{\max} = ([T_j]_{\max} - T_a) / (R_{\text{pkg}} + R_{\text{pcb}})$$

This implies that in order to reduce the junction temperature or to dissipate more power from the package, the thermal resistances of the interface between the package - PCB and the PCB - ambient have to be minimized. Thermal resistance of the package is determined by the package size, and materials [15].

Table 3 shows the constant power levels assigned to each die.

Table 3: Power levels of each die

Die	Power (W)
Die1 - Die Stack - Top	0.3
Die2 - Die Stack - Mid	0.3
Die3 - Die Stack - Bottom	0.3
Die4 - Bottom (Volatile)	0.4
Die5 - Logic Die	1

For each of the stacking architecture, steady state thermal simulation was carried out as a baseline study to determine the maximum junction temperature using Ansys® WorkBench™ 10.0 as a Finite Element Analysis (FEA) Tool. The CAD models were imported into Ansys® WorkBench™ 10.0. Material properties were already assigned during CAD modeling. Figure 9 shows the FEA model of the package and PCB. Initial analysis resulted in nodes greater than 300,000. Mapped face meshing with an edge size of one division was introduced to reduce the number of nodes to less than 30,000. Figures 10, 11, and 12 show the FEA models for each of the individual architectures.

Baseline study further considers an application of an effective heat transfer co-efficient of $5 \text{ W/m}^2\text{-}^\circ\text{C}$ on top and bottom surfaces of the PCB, and on top of the mold cap. All the other properties held constant. The ambient temperature T_a was assumed to be 25°C . Even though the ambient temperature was 25°C , it resulted in a high junction temperature. It should be noted that in real world applications, the ambient temperature can be greater than or equal to 50°C . Based on the maximum junction temperature, the following thermal management strategies were investigated:

1. Increasing the effective heat transfer coefficient on top of the mold cap to $20 \text{ W/m}^2\text{-}^\circ\text{C}$, keeping the effective heat transfer coefficient on top and bottom surfaces of the PCB constant ($5 \text{ W/m}^2\text{-}^\circ\text{C}$).
2. Increasing the thermal conductivity (k) of the PCB core to $0.8 \text{ W/m-}^\circ\text{C}$ from $0.18 \text{ W/m-}^\circ\text{C}$.
3. Introducing underfill ($k = 0.8 \text{ W/m-}^\circ\text{C}$) for the solder balls.
4. Introducing a copper spreader with the mold compound.

Additionally, the die attach on the logic die was changed to a conductive die attach, with value of $k = 3 \text{ W/m-}^\circ\text{C}$ [16]. This did not have a significant impact on the temperature distribution, and thus, not included in this paper.

Figure 9: FEA model – Package and PCB.

Figure 10: FEA model – Rotated Stack
No. of Nodes: 23908

Figure 11: FEA model – Staggered Stack
No. of Nodes: 23908

Figure 12: FEA model – Spacer Stack
No. of Nodes: 25756

III Results and Discussion

Baseline

For the baseline simulation, an effective heat transfer coefficient of $5\text{ W/m}^2\text{-}^\circ\text{C}$ was applied on top of the mold cap, and the top and bottom surfaces of the PCB. This resulted in the maximum junction temperature of around 112°C . The baseline simulation temperature profiles of the three architectures (PCB and mold cap are hidden) are shown in Figures 13, 14, and 15. The temperature contours are from 80°C to 115°C .

Figure 13: Temp. profile of the Rotated Stack

Figure 14: Temp. profile of Staggered Stack

Figure 15: Temp. profile of Spacer Stack

Strategy 1

Increasing the heat transfer coefficient ($h = 20\text{ W/m}^2\text{-}^\circ\text{C}$), representative of forced convection condition, on top of the mold, resulted in a decrease in the maximum temperature by nearly 7% in all of the three architectures. Of these, the spacer architecture showed more improvement because the maximum temperature occurs near the top of the mold cap. The higher heat transfer rate on top of the mold cap provides the necessary cooling advantage. It should be noted that in particular systems, like cellular phones, implementation of forced convection will not be viable from reliability and economic considerations.

Strategy 2

Increasing the thermal conductivity of the PCB core thermal conductivity to $0.8\text{ W/m-}^\circ\text{C}$, from $0.18\text{ W/m-}^\circ\text{C}$, results in a reduction of around 3% of the maximum temperature in each of the architectures. This is because majority of the heat generated by the package is dissipated via the PCB to the ambient.

Strategy 3

An underfill ($k = 0.8\text{ W/m-}^\circ\text{C}$) was introduced for the solder balls. Using an underfill may be required for improving the solder ball joint reliability. This was evaluated to access its effect on the overall thermal profile of the package. There was no significant decrease in temperature.

Strategy 4

Introducing a copper spreader on top of the mold compound resulted in a decrease of 1.2% for the spacer architecture, while it did not improve the performance of the other two architectures. This is because the top die surface is closest to the mold cap – ambient interface in the spacer configuration. The higher thermal conductivity of the copper spreader provides the necessary cooling advantage. Since these packages are used in a cost competitive market, adding a copper spreader would not be viable economically either.

Table 4 shows the maximum junction temperature of the different simulations. Figure 16 shows the temperature comparison between the three architectures and the five cases evaluated and Figure 17 shows the temperature reduction percentage for different architectures with different strategies.

Table 4: Max. Junction Temperatures [°C] of different simulations

	Rotated	Staggered	Spacer
Baseline	112.45	112.81	112.2
Strategy 1	104.63	104.99	104.38
Strategy 2	109.17	109.54	108.92
Strategy 3	112.37	112.72	112.09
Strategy 4	112	112.42	110.86

Figure 16: Stack Architecture vs Maximum Temperature
Of the four strategies evaluated, Strategy 1 provides the best performance.

Figure 17: Stack Architecture vs Percentage Temperature Reduction

IV Summary and Conclusions

A parametric study and thermal management strategy was conducted on a five die stacked package, which included both memory and logic processor on the same substrate. Three different configurations, viz. rotated, staggered, and spacer, were evaluated. Temperature profiles of the overall package were evaluated with a focus on maximum junction temperature. Four different thermal management strategies were assessed from the design and boundary condition point of view.

Of the three architectures evaluated, the spacer architecture was the easiest to improve. However, none of the thermal management strategies evaluated offered significant improvement in performance. This implies that the root cause for the elevated temperature profiles is the high power buildup

in the package die stack i.e. memory and logic dice. The three stack architectures evaluated assume that all the dice are functioning simultaneously – the worst case scenario. However, in practice, the various dice do not function simultaneously. The degree of functionality of each die, especially the memory dice, is determined by the bus architecture. The thermal management strategies evaluated in this paper have demonstrated the significance of system design. Efforts must be concentrated on the system design processes (placement of components on the PCB, bus architecture, etc.) to improve the overall performance of the package.

References

1. International Electronics Manufacturing Initiative (iNEMI) – 2004 Overview of Technology Roadmap - <http://www.inemi.org>
2. International Technology Roadmap for Semiconductors (ITRS) – 2005: Executive Summary, and Assembly and Packaging - <http://www.itrs.net>
3. Said F., et al, “A review of 3 D packaging Technology”, IEEE transactions on CPMT, Vol. 21, Issue 1 (1998), pp. 2-14
4. Intel Technology Journal (ITJ) – Vol. 9, Issue 4, <http://intel.com/technology/itj/2005/volume09issue04>
5. Beyne, E., “3D interconnection and packaging: impending reality or still a dream?”, IEEE Digest of Technical Papers on ISSCC, Vol. 1. (2004), pp. 138 - 139
6. Carson, F., et al, “The development of a novel stacked package in package”, IEEE/CPMT/SEMI EMTS, (2004), pp. 91 - 96
7. Kappor, R., et al, “Package design optimization and materials selection for stack die BGA package”, IEEE/CPMT/SEMI EMTS, (2004), pp. 113 - 118
8. Solberg, V., et al, “Performance evaluations of stacked CSP memory modules”, IEEE/CPMT/SEMI EMTS, (2004), pp. 301 - 308
9. Guixiang Tan, et al, “Modeling of stacked packaging thermal performance in module/system environment” IEEE SEMITHERM, (2005), pp. 270 - 274
10. Haehyung Lee, et al, “Thermal characterization of high performance MCP with silicon spacer having low thermal impedance”, IEEE SEMITHERM, (2005), pp. 322 - 326
11. Garcia, E.A., “Two-resistor compact modeling for multiple die and multi-chip packages”, IEEE SEMITHERM, (2005), pp. 327 - 334
12. Akhter, R., et al, “Thermal Comparison of Die Stacking Architectures for Flash Memory Applications” Proceedings of IMECE, (2005) HT2005 - 72853
13. Joint Electron Device Engineering Council (JEDEC) [EIA / JESD 51, 51(1) to(6)] - <http://www.jedec.org>
14. Rencz, M., “Thermal issues in stacked die packages” IEEE SEMITHERM, (2005), pp. 307 - 312
15. “PCB Design and SMT Assembly/Rework Guidelines for MCM-L Packages” – Application Notes - <http://www.skyworksinc.com>
16. ZipCURE™ Die Attach Adhesive Datasheet <http://www.polysciences.com>
17. Pro/E® WildFire™ 2.0 Reference Manual
18. Ansys® WorkBench™ 10.0 Reference Manual