2025 24th IEEE Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITherm) | 979-8-3315-2428-9/25/$31.00 ©2025 IEEE | DOI: 10.1109/ITherm55376.2025.11235656

Silicon TTV for Advanced Thermal Investigations of High Powered Lidless Package AI Silicon

Jonathan Stever[1], Cheng Yang[2], Yin Hang[1], Pascale El Kallassi[1], Chloe Xu[1], Chen Wang[3], Yanbo Tang[3], Shuainan Lin[3]
Dongkai Shangguan[4]

[1]Meta Platform, Menlo Park, California, USA
[2]STATS ChipPAC Pte. Ltd, Singapore
[3]JCET Group Co., Ltd., Shanghai, China
[4]Thermal Engineering Associates, Inc., Santa Clara, California, USA
dongkai@thermengr.com

***Abstract*—To meet the ever-increasing demands for compute and power density in the AI/ML chips, we believe that more advanced thermal test vehicles (TTVs) are needed for future silicon thermal characterization. The transition to large lidless packages has introduced many new critical areas including die warpage and die-to-die thermal cross-talk. These factors significantly impact our understanding of thermal performance and limitations. To address this challenge, we have developed a silicon-based TTV that features a multi-die lidless BGA flip chip design. This design includes a large high-powered main die at the center, surrounded by smaller lower-power dies, effectively replicating a modern AI/ML chip. This enables us to perform thermal characterization at the package level, providing quantitative results on die-to-die thermal interactions, thermal resistances across different layers of the heat transfer stack-up, and temperature gradients within dies of a non-uniform heat flux.**

Our design incorporates 25 independently controlled heating zones for the center die, totaling 3 kW of total heating power for all the dies, along with an array of temperature sensors for each chip with majority located in the center large die and rest located in the other dies. This provides precise control over the heating profile of the chip and accurate real time temperature gradient information.

We have tested the thermal resistance with emerging technologies in the Thermal Interface Material (TIM) space (including phase change materials and liquid metal), evaluating their thermal performance enhancements and impact on power limits due to thermal constraints.

In this paper, we will walk through the TTV design, thermal characterization and thermal interface material evaluation, as well as some lessons learned.

Keywords—thermal test vehicle, lidless package, thermal interface material, thermal resistance, thermal characterization

I. Introduction

With growing computing power in artificial intelligence (AI) and machine learning applications, the power density of IC chips also becomes significantly higher, leading to challenges in appropriate cooling solutions to meet product functionality and reliability. The characterization of the thermal behavior, which is necessary for developing effective thermal management solutions, becomes critical before chip production.

Parameswaran et al. [1] developed an innovative approach to thermal testing, which facilitates high-resolution thermal mapping and precise temperature measurement (±2°C). The work of Parameswaran et al. [2] features a highly flexible and programmable package-level thermal hardware evaluation system enabling the generation of on-die heat maps as a function of various package types, thermal interface materials, and cooling solutions. In their study, Refai-Ahmed et al. [3] investigated the thermal performance of both lidless and lidded flip chip packages, emphasizing the critical role of thermal interface materials (TIMs) in preventing TIM pump-out and ensuring effective thermal management. Their findings indicate that carefully selecting TIMs and optimizing design parameters can profoundly influence the thermal efficiency of advanced electronic packages [3]. In the same study, it was found that lidless package is superior to the lidded package regarding thermal performance.

Historically, a large chip package includes an integrated heat spreader with controlled TIM1 BLT (bond line thickness) to spread the heat out to a larger area for better uniformity and lower heat flux. Including such a heat spreader and TIM1 adds thermal resistance, which needs to be compensated by lower TIM2 thermal resistance due to larger area and better warpage control. (For lidless packages, thicker BLT might be needed to achieve higher coverage of TIM between the die surface and the heatsink/cold plate surface, which thereby causes an increase in thermal resistance.)

Recently, lidless packages have become more popular where TIM1.5 connects the silicon die to the cold plate directly. With 2.5D or MCM (multi-chip module) packages, multiple dies in the same package can heat up and impact each other (thermal cross-talk), and therefore, immediately bringing the heat out without intermediate TIM1 and heat spreader can usually benefit the overall thermal management. However, effects such as warpage may become a bottleneck for lidless package cooling solutions because of poor TIM1.5 performance (due to the contact resistance and bond line thickness).

Steve et Al. [4] plotted the thermal resistance contribution of ASIC and HBM in a typical 2.5D package, as shown in Figure 1 below, which indicates that the structure and thermal cross-talk would have impact on such complicated structures.

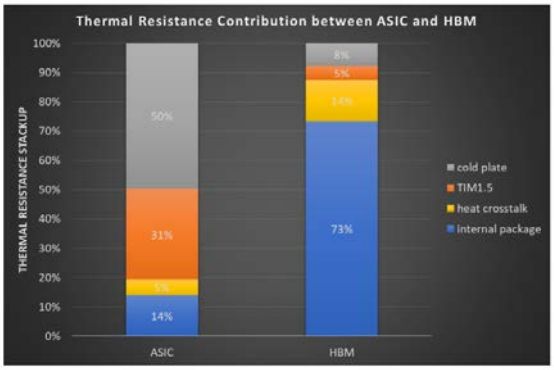

Figure 1. Plot of Thermal Resistance Contribution [4]

Thermal Test Chips (TTC) and Thermal Test Vehicles (TTV) are great tools for thermal characterization, and such tools can provide the early-stage thermal design directions while the actual functional chip is still under development. To accommodate a wide variety of applications, the Unit Cell concept is used which is the basic building block for the TTC. Each Unit Cell consists of a heating resistor and an embedded diode temperature sensor. By using multiple Unit Cells in a square or rectangular array configuration, it is possible to approximate almost any size application chip. [5]

Flip chip ball grid array (fcBGA) packages are widely used with liquid cooled cold plate. The TIM is situated between the TTC die surface and the liquid cooling plate. Precise pressure from the cooling plate is essential for optimal contact with the TIM. Direct TIM materials, including PCM (phase change material) and liquid metal, are used in this study.

To demonstrate the TTV design as an effective tool for thermal characterization, we performed θ_{JA} testing representing the thermal resistance between the die junction and the liquid coolant.

II. Thermal Test Vehicle and Package Design

A. *Thermal Test Chip (TTC) Unit Cell*

The TTCs must be able to closely approximate the power input and power density distribution of the application chip, and simultaneously accurately sense the temperature distribution (using integrated sensors) over the entire die (with resolution down to 1mm x 1mm), in real-time. [5]

We used the thermal test chip which is based on a 1mmx1mm silicon Unit Cell and is designed to provide maximum flexibility for thermal characterization of semiconductor packages. An integrated center diode temperature sensor in each Unit Cell enables temperature measurements to be made in multiple locations in any array configuration. [6]

Each Unit Cell is equipped with a resistor of nominal value 10.5 ohms and a temperature sensing diode. Both the heating resistor and the sensing diode are individually accessible, allowing both heating and temperature measurement as shown in Figure 2. The resistor covers >69% of the die area and can be accessed with wire bonding or bump pads. [7]

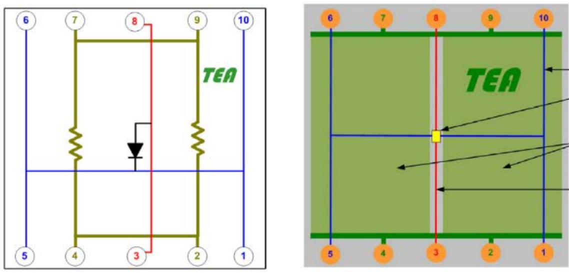

Figure 2. Unit Cell electrical representation and layout

B. *Thermal Test Chip (TTC) Design*

Thermal test chips have been extensively utilized in the qualification of package designs. In our study, we designed a bare die TTC. The TTC comprises an array of 5 x 5 heating zones, with each zone featuring an array of 5 x 6 Unit Cells as shown in Figure 3. The TTC has Cu pillars with lead-free solder cap, with 0.2mm pitch.

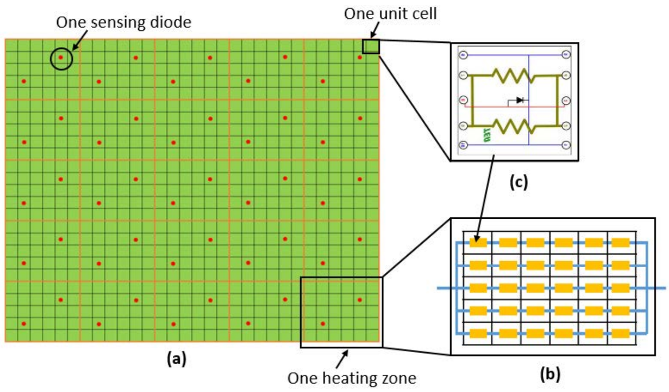

Figure 3. One TTC (a) contains 5 x 5 heating zones (b), and one heating zone contains 5 x 6 Unit Cells (c)

In order to configure the power zones for the desired power distribution for the chip, in each heating zone, six Unit Cells are initially arranged in series, after which five identical configurations are connected in parallel to create a 5x6 matrix as shown in Figure 3(b). Each of the total 25 heating zones shown in Figure 3(a) can be accessed and controlled directly and individually. This design affords the flexibility to connect the power zones in series or in parallel, depending on the desired connection topology.

Up to 750 diode sensors can be available in each TTC. In this design, only two sensing diodes are routed out, making them accessible in each heating zone. The routed-out sensor locations are indicated by the red dots shown in Figure 3 (a). With a total of 25 heating zones, there are 50 sensing diodes evenly distributed across the TTC.

C. *Thermal Test Vehicle (TTV) Design*

A flip chip thermal test vehicle was developed to evaluate the thermal performance of the package. For this TTV, we included eight additional TTC dies encircling the central main TTC on the TTV to replicate the heterogeneous package configuration for various functional dies, as is typical of GPUs in the market. In this paper, we concentrate on the measurement of the central main TTC marked D1 in Figure 4. Further investigation and discussion regarding the surrounding dies will be undertaken in future studies.

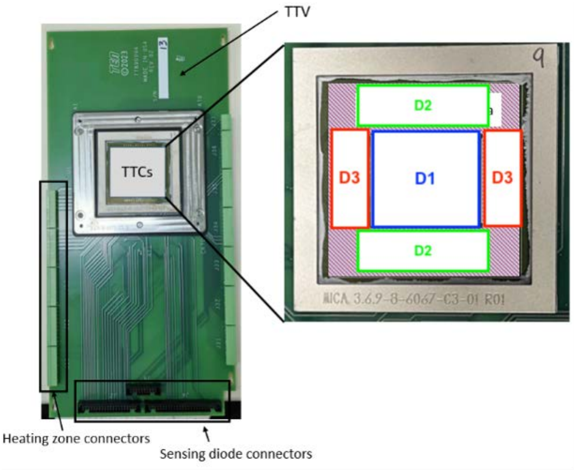

Figure 4. TTV Diagram represents general locations not actual size and locations

For TTV assembly, the TTC is attached to an organic substrate using flip chip and is then underfilled. It is then assembled onto the PCB (printed circuit board) using ball grid array (BGA). Each heating zone and sensing diode is routed out through the TTV substrate to the individual connectors on the PCB.

The cross-sectional view of the FCBGA package together with TIM and the cooling plate is shown in Figure 5. From the top to the bottom is the cooling plate, the TIM layer, the silicon TTC, a layer of underfill around solder bumps, an organic substrate, an array of solder balls, and finally the TTV test board (PCB). The cold plate is designed to cool the D1 as the primary heat source, and the coolant coming from the system loop first impinges on the center of D1 then flows across the entirety of the chip and on to D2 and D3 regions. The cold plate base area is about 50x50mm which covers all the die surfaces (D1-D3) from the TTV. The aim of this design is to more precisely replicate the actual application of a liquid cooled AI chip. The pressure applied by the cooling plate can be adjusted using spring-loaded screws designed for fixation.

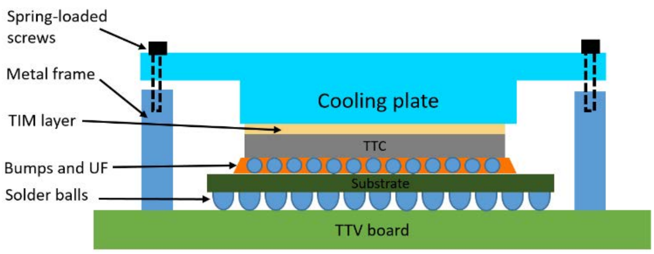

Figure 5. The cross-section view of the TTV system

D. Liquid Cooling

The direct-to-chip liquid cooling system hinges on circulating a fluid through a cooling plate heat exchanger located directly on the chip. The liquid cooling plate is affixed and makes contact with the top surface of the TTC using a TIM to transfer heat from the TTC to the circulating coolant. The cooling plate's surface is composed of copper, and the heat dissipated from the chip is absorbed into the coolant loop. The coolant loop utilizes a LongWin 9670B Liquid Cooling Tester for precision temperature and flowrate control. The heated fluid is then circulated through a large liquid tank which is temperature controlled. The cooling plate is secured to a rectangular metal frame by spring-loaded screws at the corners of the frame, as depicted in Figure 5.

The coolant employed in our experiment consists of a blend of 75% DI (deionized) water and 25% propylene glycol to simulate a common commercially used coolant on the market. The temperature of the coolant is regulated by a heater and a refrigeration unit within the tank. A temperature sensor positioned at the inlet of the cooling plate continuously monitors the coolant's temperature to ensure precise control. A valve and a flow rate meter are also incorporated into the coolant loop to regulate and oversee the coolant's flow rate.

E. Thermal Interface Materials (TIM)

An efficient TIM is essential for optimal heat dissipation performance throughout the TTC to the cooling plate. For our TTV design, a layer of TIM is attached between the TTC top surface and the liquid cooling plate.

Figure 6. Thermal interface material between interfacing surfaces

Given the surface roughness of both the silicon die and the cooling plate as shown in Figure 6, there may be microscopic gaps that could trap air. Direct contact between the cooling plate and the TTC will not achieve optimal thermal dissipation; thus, employing a TIM is better for heat dissipation performance. An appropriate pressure can enhance the contact between the TIM and the TTC and facilitate superior heat dissipation.

A phase change material and a liquid metal are used as TIM in this study.

III. Experiment Setup

A. Temperature Calibration

To convert ADC values from the diode into temperature readings, calibration is necessary. An ESPEC BTZ-175 oven was used to calibrate the ADC data at six distinct temperature points. Data were used to generate a calibration matrix, which can then be used to convert ADC values to temperature values for each individual thermal diode.

Based on the calibration curve shown in Figure 7, we can determine the measurement resolution. A temperature change of one degree Celsius corresponds to 63 ADC units, which means that our resolution is better than +/-0.1 degrees Celsius.

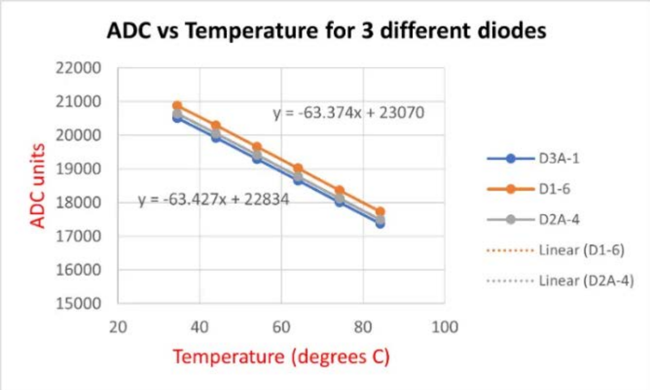

Figure 7. Calibration Curve

B. *Measurement Configuration*

The measurement configuration is shown in Figure 8. The test board is connected to the TTV by applying a sensing current and gauging the voltage of the sensing diode with the calibrated ADCs. A DC power supply is connected to the connectors of the heating zones of the TTV by imparting a substantial heating power to heat up the dies. A multiplexer situated on the test board is also capable of switching across all the sensors on the TTCs within a few milliseconds. Owing to this configuration, we can effortlessly obtain the temperature readings of all the sensing diodes. The coolant is maintained at a constant temperature of 40 °C and flow rate of 1.5 liters per minute (LPM) throughout, from before the heating phase until its conclusion.

The θ_{JA} can be assessed through the thermal steady-state condition based on the TTC design. It is determined using the junction temperature (T_j) during the thermal equilibrium, the ambient temperature (T_{amb}), and the applied heating power (P_H), with the formula below. T_{amb} is the coolant inlet temperature, which is set to 40 °C in our experiment. We can measure the voltage on each sensing diode during the thermal equilibrium, thereby ascertaining the temperature distribution map across the entire TTC.

$$\theta\text{JA} = \frac{T_j - T_{LC25\_inlet}}{P_H}$$

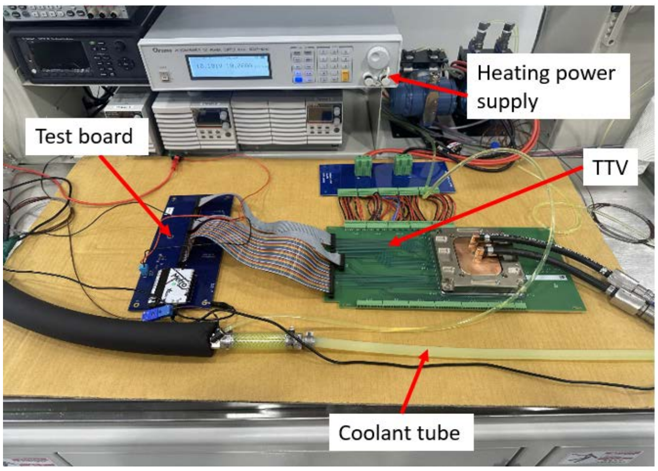

Figure 8. Measurement Configuration

IV. Results and Discusiion

A. *Warapge Validation*

We know that the TIM performance is sensitive to bond line thickness (BLT), which is related to warpage of the die top surface for lidless packages [9]. In this work, warpage of the package is measured by Akrometrix TherMoiré Measurement System. TherMoiré measures vertical displacement by exploiting geometric interference between a reference grating and the shadow it casts on a sample as shown in Figure 9. This method requires a Ronchi-ruled grating, a light source positioned at 45°, and a camera oriented perpendicularly to the sample. Additionally, phase stepping enhances resolution by translating the sample vertically relative to the grating, which results in the automatic ordering of the interference fringes.

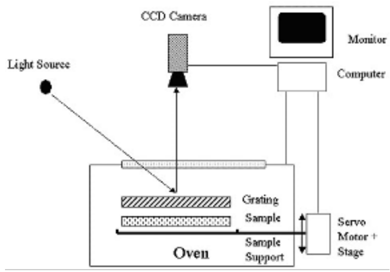

Figure 9. Akrometrix TherMoiré Measurement System

Results in Figure 10 show warpage of the center die (with TTV mounted on PCB) at temperatures from the room temperature to 90 ° C. It is noted that the die center is lower ("smile face") and the warpage increases with temperature. This is different from the "normal" case (with a free-standing FCBGA) which typically shows a "cry face" at the room temperature and the warpage decreases with temperature (flat at the reflow temperature). It may be due to the impact of the package stiffener and the thick PCB, the cold plate ring and the back plate. Further study is needed to fully understand the reason behind this inversion.

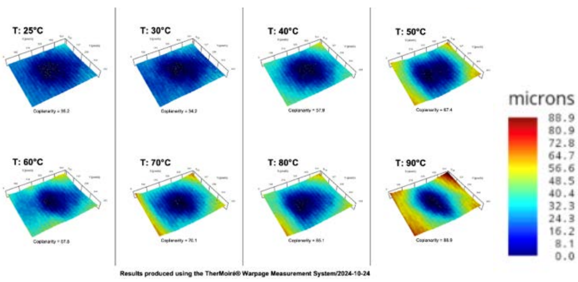

Figure 10. Warpage of the Center Die

B. *Temperature Map*

As expected, the die warpage has a strong impact on the thermal performance of TIMs that do not compress well but has very little effect on more pliable TIM materials such as the liquid metal and the phase change material. Hence, results with the liquid metal and the phase change material TIMs are presented below for discussions.

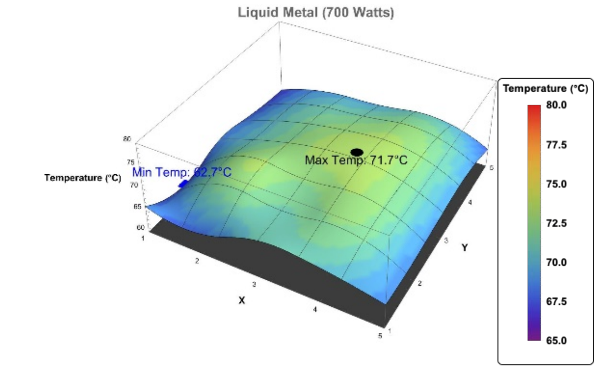

Figure 11. Temperature Map of D1 with 700W for the Center Die and Liquid Metal TIM, with inlet coolant at 40 °C

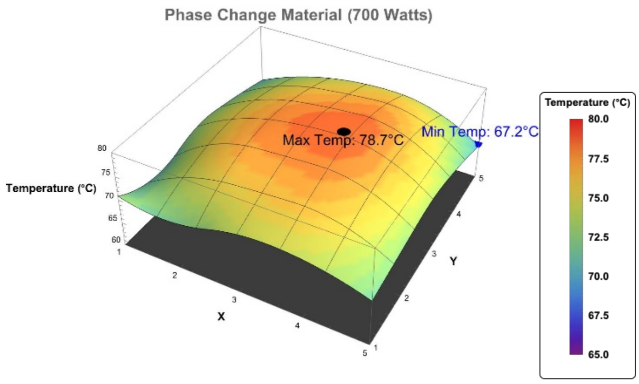

Figure 12. Temperature Map of D1 with 700W for the Center Die and Phase Change Material TIM, with inlet coolant at 40 °C

Figures 11 and 12 show the temperature maps of D1 with 700W, for liquid metal and phase change material TIM, respectively. The temperature at the die edge and corner are lower where the warpage measurement shows a higher surface leading to thinner TIM. As the actual warpage is also impacted by the actual temperature distribution (whereas the warpage measurement is in an oven with a uniform temperature) and by the constraining force by the cold plate, the highest temperature location might not be always at a consistent location at the peak point of the convex shape in the warpage map (Figure 10). The general trend, where the temperature map appears as an inverted version of the mechanical warpage map (cry vs smile face), is as expected. The hottest point is at the center of the die, which corresponds to the lowest region shown in Figure 10. This aligns with the highest TIM BLT and R_TIM. In contrast, the die edges, which exhibit the lowest BLT, correspond to the coolest temperatures.

Figures 10 and 11 indicate that the liquid metal TIM has a better TIM performance than the phase change material, which is mainly due to the higher thermal conductivity of the material itself.

C. *Thermal Cross-Talk*

As multiple dies are mounted on the substrate and heating of different dies can lead to thermal cross-talk, experiments with different heating powers for D1/D2/D3 were conducted.

As the thermal cross-talk (i.e., effect of heating power of nearby dies on the target silicon die) is usually linear (true when the representative maximum temperature of the silicon die is at a fixed location since thermal conduction dominates), DOEs (design of experiments) with different powers of D1/D2 (while D3 power is fixed to reduce complexity, as practically this chip power has only very small variations during operation) were conducted to capture the thermal cross-talk effect.

When the liquid metal TIM is used, Figure 13 shows the thermal map on D1 when D2 and D3 are powered up.

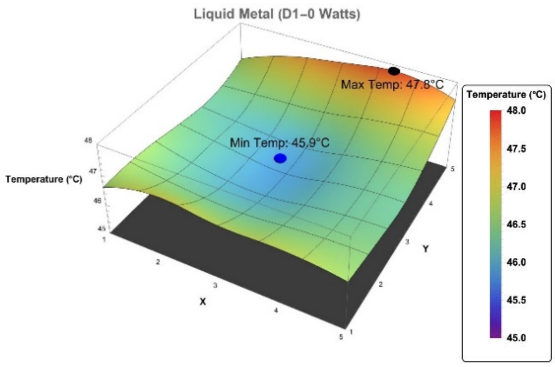

Figure 13. Temperature Map when D1 is not powered, and D2 and D3 are powered, with inlet coolant at 40 °C

As shown in Table 1, D3 is powered with a constant power, and with different power levels of D1 and D2, we conducted an experiment of 14 cases and derived the following equations to quantify the thermal cross-talk as shown in Table 2.

Table 1. Power value of the 14 cases

Case	P_D1 (W)	P_Per_D2 (W)	P_Per_D3 (W)
1	65%	75%	100%
2	81%	75%	100%
3	97%	75%	100%
4	65%	88%	100%
5	81%	88%	100%
6	97%	88%	100%
7	65%	100%	100%
8	81%	100%	100%
9	97%	100%	100%
10	0%	113%	100%
12	94%	85%	100%
13	113%	0%	100%
14	81%	98%	100%

Table 2. Calculated Coefficients of the Thermal Cross-talk

Calculated Coefficients			
Liquid Metal		PCM	
A1	0.037	A1	0.051
B1	0.034	B1	0.069
C1	6.529	C1	2.935
A2	0.011	A2	0.015
B2	0.226	B2	0.322
C2	3.351	C2	1.125

$$T_{D1} = A1 \cdot P_{D1} + B1 \cdot P_{D2} + C1 + T_{inlet}$$
$$T_{D2} = A2 \cdot P_{D1} + B2 \cdot P_{D2} + C2 + T_{inlet}$$

To determine the coefficients of the characteristic equation, we employed Microsoft Excel's Generalized Reduced Gradient (GRG) Nonlinear solver, chosen for its effectiveness in handling nonlinear optimization problems. Our objective was to minimize the Sum of Squared Errors between the model predictions and the empirical data , and we achieved a value below 10. As T_{D1} and T_{D2} have a maximum junction temperature limit, we can use the above correlated equation to derive the power limit that the specific cooling solution can support.

Figure 14 shows the maximum D1/D2 power allowed with both the liquid metal TIM and the phase change material TIM, calculated from the equations above. As the maximum power is limited by the maximum junction temperature (different for D1 and D2), the percentage itself does not mean higher or lower thermal cross-talk; it is only used to indicate that the thermal cross-talk will have an impact on the temperature rise and hence limit the maximum power level of dies in the MCM package. This effect is evident from the slopes of the plots. For example, the maximum power supportable by D2 (HBM) decreases by nearly 20% due to thermal cross-talk as D1's power increases. In contrast, increasing D2's power impacts D1 by approximately 8% with PCM or 4% with liquid metal. These results indicate that thermal cross-talk predominantly flows from D1 to D2. Figure 12 also confirms that the liquid metal TIM has a better thermal performance (with a greater effect on D1, maybe because of thicker BLT from the warpage data, leading to relatively larger TIM thermal resistance, on D1 as compared with D2).

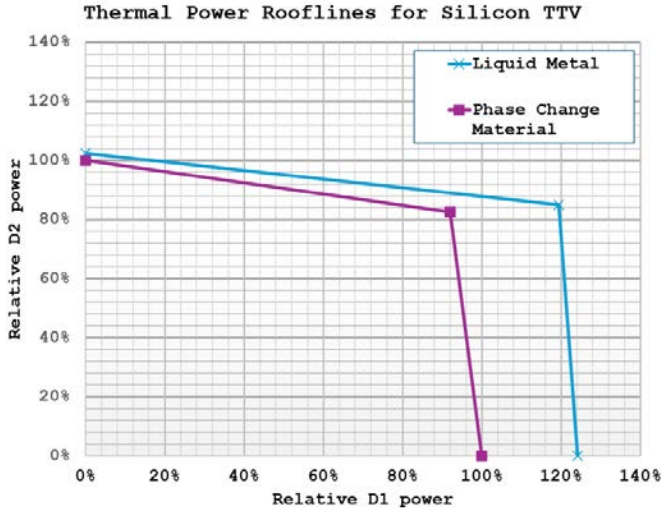

Figure 14. Thermal limited power dissipation of D1 and D2

The performance difference shown in Figure 14 can likely be attributed to the thermo-mechanical characteristic differences between the materials. Notably, thermal conductivity plays a significant role: liquid metal TIMs, such as the one used here, are predominantly composed of indium–gallium alloys that can achieve effective thermal conductivities exceeding 30 W/(m·K), whereas silicone-based PCMs typically offer thermal conductivities below 10 W/(m·K). Additionally, due to the chip's pronounced die warpage, we expect the bond line thickness to reach approximately 100 μm in the worst-case areas (i.e., the center of the chip), thereby magnifying the TIM layer's impact on overall thermal resistance.

D. Summary

A MCM thermal test vehicle was developed to thermally characterize a lidless package with a liquid cooled cold plate designed for high power AI applications.

The measurement results show that die warpage can lead to significant temperature variations within the die under uniform power heating as it plays a significant role in TIM (between bare die surface and cold plate) performance.

Significant thermal cross-talk has been observed among dies at a close distance. Such cross-talk impacts the maximum power limit of each die under different operation conditions.

With significantly larger die sizes in 2.5D and the emergences of wafer level (or even panel level) computing devices, how to control the warpage and MCM die height variations would be critical to minimize thermal resistance caused by TIM. One possible long-term solution might be using direct-to-chip liquid immersion cooling which eliminates TIM in the cooling path. However, as liquid direct contact of the silicon would bring some concerns of contamination and leakage, the short-term solution with a liquid cooled cold plate may require some innovations to overcome the large die warpage and MCM die height variations to get more uniform TIM performance.

ACKNOWLEDGMENT

The authors wish to acknowledge Mr. Bernie Siegal of Thermal Engineering Associates, Inc. for his contribution to the project, andAkrometrix for their contribution to the TTC warpage characterization.

REFERENCES

[1] S. Parameswaran, S. Balakrishnan and B. Ang, "Versatile chip-level integrated test vehicle for dynamic thermal evaluation," IEEE International Conference on Microelectronic Test Structures (ICMTS), Austin, TX, USA, pp. 122-127, 2018

[2] Parameswaran, S, Refai-Ahmed, G, Ramalingam, S, & Ang, B. "Next Gen Test-Vehicle to Simulate Thermal Load for IoT FPGA Applications." Proceedings of the ASME 2018 International Technical Conference and Exhibition on Packaging and Integration of Electronic and Photonic Microsystems. San Francisco, California, USA. August 27–30, 2018

[3] G. Refai-Ahmed et al., "Lidless and lidded Flip Chip Packages for Advanced Applications," IEEE 22nd Electronics Packaging Technology Conference (EPTC), Singapore, Singapore, pp. 104-111, Debember 2020

[4] 30°C COOLANT - A DURABLE ROADMAP FOR THE FUTURE https://ocp-all.groups.io/g/ocp-acf/attachment/379/0/30C%20Durable%20Coolant%20Whitepaper%20Rev0.2.pdf

[5] Thermal Engineering Associates, Inc. Thermal Test Chip, https://www.thermengr.net/html/ttc.html

[6] Digital and Physical: Thermal Twins for Developing Thermal Management Solutions, https://www.3dincites.com/2023/04/digital-and-physical-thermal-twins-for-developing-thermal-management-solutions/

[7] B. Siegal and J. Galloway, "Thermal Test Chip Design and Performance Considerations," 2008 Twenty-fourth Annual IEEE Semiconductor Thermal Measurement and Management Symposium, San Jose, CA, USA, 16-20 March, 2008

[8] Thermal Engineering Associates, Inc. Tech Briefs, https://www.thermengr.net/html/tech_briefs.html

[9] A. Jain, D. P. Parekh, M. Bergendahl, and K. Sikka, "Analytical Solution for a Chip Temperature Distribution in a Multilayer Chip - Package with Thermal Warpage," 2021 20th IEEE Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (iTherm), San Diego, CA, USA, Jun. 1 - 4, 2021, pp. 114-122, 2021, doi: 10.1109/ITherm51669.2021.9503285.