Thermal Challenges for Heterogeneous 3D ICs and Opportunities for Air Gap Thermal Isolation

Yang Zhang, Thomas E. Sarvey and Muhannad S. Bakir
Georgia Institute of Technology
791 Atlantic Drive
Atlanta, Georgia, USA 30332
Email: steven.zhang@gatech.edu

***Abstract*—Thermal crosstalk within a heterogeneous 3D IC results in higher temperatures for low-power dice; this is particularly true in memory-logic, photonic-logic, and MEMS-logic stacks. The elevated temperatures may consequently impact the performance of the low-power devices. This paper describes a thermal solution for both heat removal as well as thermal isolation within a 3D chip stack. Based on the evaluated memory-logic 3D architecture and compared to conventional air-cooling, the proposed technologies reduce the maximum temperature of the memory die from 75.6 °C to 36.7 °C and processor die from 75.9 °C to 60.1 °C.**

Keywords—3D IC; DRAM; microbumps; microfluidic heat sink (MFHS); multicore processor; TSVs

I. Introduction

While significant work has addressed the thermal challenges of 3D integration, e.g. the increasing power density and inter-stack thermal resistance [1], relatively fewer efforts have been proposed to mitigate the negative effects of thermal coupling between different dice in a 3D heterogeneous stack, and in particular, to minimize inter-die thermal coupling.

For instance, in a DRAM-processor stack, the DRAM will usually have a relatively higher temperature due to strong thermal coupling [2] even though the DRAM itself dissipates much lower power than the processor. However, a higher DRAM temperature (in extended temperature range) degrades memory performance by 8.6% and results in 16.1% additional power [3]. Likewise, in silicon nanophotonics, the ring resonators are sensitive to temperature [4]. Similarly, there are temperature coupling challenges for 3D stacking of MEMS and their readout circuits [5] [6]. From the above examples, there is a need for novel technologies to reduce the thermal crosstalk within the stack to 'protect' the low-power and temperature sensitive dice.

To resolve this thermal coupling problem, we investigate a 3D stack architecture with an interposer embedded microfluidic heat sink, an air gap between the stacked dice, and a low-resistance thermal path to cool the isolated die [7]. The low-resistance thermal path, which we call the 'thermal bridge,' is a copper plate cooled by an auxiliary heat sink. The 3D stack configuration is shown in Fig. 1.

In this paper, we explore the opportunities of the proposed architecture and develop a thermal model to benchmark the proposed architecture with baseline stacks that utilize an air-cooled heat sink and standalone microfluidic cooling (without air gap and thermal bridge).

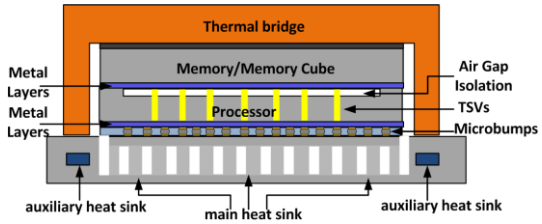

Fig. 1. Proposed architecture with interposer-embedded heat sink, thermal bridge, and air-gap isolation.

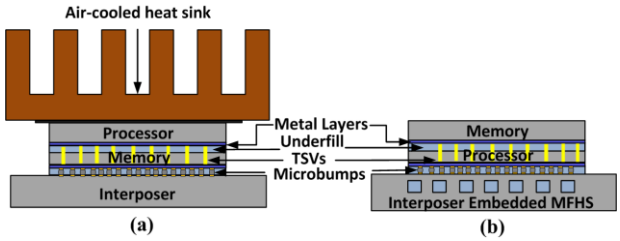

Fig. 2. 3D stack (a) with conventional air cooled heat sink (b) with interposer embedded microfluidic heat sink (MFHS)

II. Benchmark Architecture

A. Proposed and baseline stack architectures

The proposed architecture, shown in Fig. 1, has three key features: 1) A microfluidic heat sink (MFHS) is integrated in the interposer and consists of two separate parts. The main MFHS is under the processor die. It serves as the main thermal path for the stack. The auxiliary MFHS is located at the peripheral of the interposer and is used to cool the thermal bridge (to be discussed later). 2) An air-gap thermal isolation is integrated between the high-power and low-power dice to reduce the thermal crosstalk, and 3) a "thermal bridge" is attached on top of the isolated low-power die to provide an 'external' low-resistance thermal path for the isolated die.

Fig. 2 shows two 3D stacks with different cooling solutions that are used as benchmarks for our proposed approach. The first 3D stack is based on a conventional air-cooled heat sink (with heat spreader). The second 3D stack is cooled using a microfluidic-cooled interposer, as shown in Fig. 2(b). Even if the microfluidic-cooled interposer can lower the stack temperature compared to the air-cooled heat sink, the thermal coupling between the two dice remains as an unsolved challenge.

978-1-4799-8472-5/14/$31.00 ©2014 IEEE

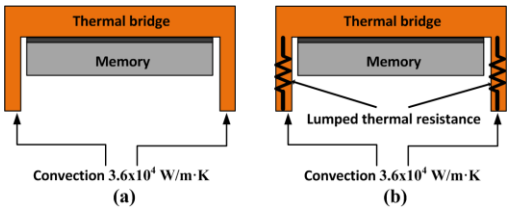

Fig. 3. (a) Physical structure of the thermal bridge; (b) Lumped resistance modeling for bridge fins and TIM

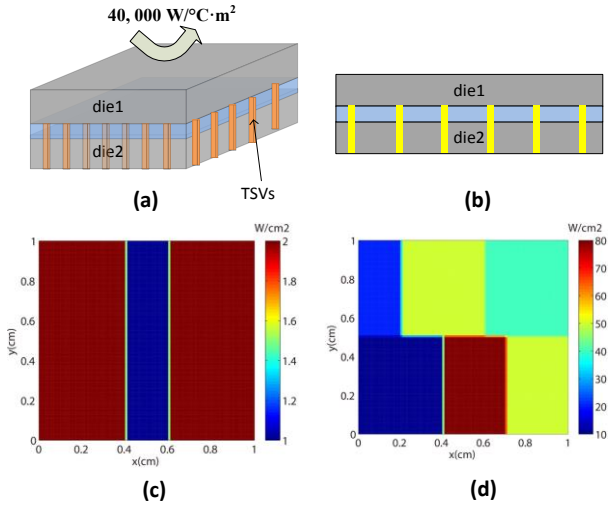

Fig. 4. (a) Configuration used for the validation example. (b) The cross-sectional view of the 3D stack. (c) Power map of die #1. (d) Power map of die #2

B. *Thermal bridge*

Without an effective thermal path for the isolated die, the temperature of the isolated die may be relatively large. In Fig. 1, this need is addressed using the "thermal bridge," which can be formed using a modified copper spreader. Fig. 3(a) shows the physical structure of the thermal bridge. The top surface of the copper thermal bridge is 1.5 cm x 1.5 cm with a thickness of 500 µm (assuming chip size is 1cm x 1 cm). A convective boundary condition of 3.6×10^4 W/m^2·K is applied. To simplify the structure, we model the bridge fins and TIM (attaching the bridge to the interposer) as lumped thermal resistors shown in Fig 3(b); the width of the fin (2 mm) justifies this simplification.

C. *Thermal modeling*

By using non-conformal grids in the chip and interposer [8] and the weighted thermal conductivity calculation in the chip domain [9], we implement a thermal model using the finite difference method. The schemes are described in [10]. We use the backward Euler scheme [10] to implement the transient analysis. To model the thermal interactions between the fluidics and the chip, we added the energy balance equation described in [11] into our finite difference scheme.

Fig. 4 (a) (b) shows an example 3D stack that was used to validate the thermal model with ANSYS. The power map of each of the stacked chips is shown in Fig. 4 (c) (d). All surfaces are adiabatic except for the top surface, which is defined to have a convection heat transfer coefficient of 40,000 W/°C·m^2.

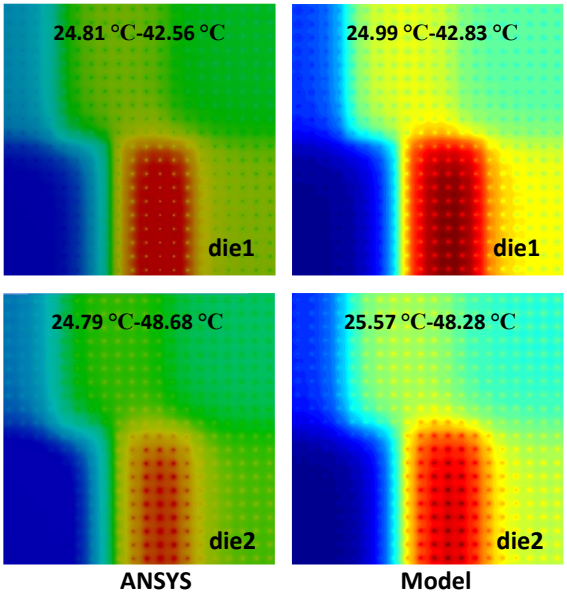

Fig. 5. Thermal map of both stacks using ANSYS and the model.

The chip size is 1 cm × 1 cm. To reduce the meshing and analysis complexity in ANSYS, we only use 400 uniformly distributed TSVs between the two dice in this validation example. The TSV diameter is 50 µm, and we assume there is no liner (again, to simplify ANSYS meshing). The thickness of both dice is 50 µm and the bonding layer is 5 µm. The thermal maps of both dice using ANSYS and the thermal model are shown in Fig. 5 and match to within a maximum error of 7% for this example.

III. COMPARISON OF DIFFERENT 3D STACKS

In this section, we benchmark the three memory-processor stacks described in section II in order to gain insight into the benefits and challenges of our architecture.

A. *Specification*

Table I lists the thickness and material properties of all layers (structures) modeled in all 3D architectures considered. The chip size is assumed to be 1 cm × 1 cm. The interposer is 2 cm × 1.5 cm. The heat spreader is 4 cm x 3.5 cm (scaled according to our chip size) and the total thermal resistance from the heat spreader to ambient is 0.218 K/W [12]. The interposer embedded microfluidic heat sink is assumed to be the same size as the chip; we assume thermal characteristics similar to those reported in [1] at a flow rate of 100 ml/min. The modeled thermal bridge is shown in Fig. 3(b). The ambient temperature is set to 25 ºC for all three scenarios.

B. *Power density maps*

Fig. 6 illustrates the power maps of the memory and processor dice. The memory die layout is based on a 3D DDR3 DRAM design from *Samsung* [13]. The layout of the processor die is based on the *Intel* i7 microprocessor [14]. The total DRAM power is set to 2.82 W [15], and the total processor power is set to 74.49 W based on the *Intel* Core i7 processor [16]. The processor TSV diameter is assumed to be 5 µm with

TABLE I
PARAMETERS

	Conductivity (W/K·m)	Thickness (μm)	Heat capacity (J/Kg·K)	Mass density (Kg/m^3)
TIM	3	20	1000	2900
Memory die	149	100	705	2329
Underfill layer	0.9	5	1000	2100
Air gap	0.024	5	1030	1.23
Processor die	149	100	705	2329
Micro-bump	60	40	227	12000
Interposer	149	200	705	2329
Copper	400	N/A	385	8690
SiO_2	1.38	N/A	705	2648
fluidics	0.6	N/A	4187	1000

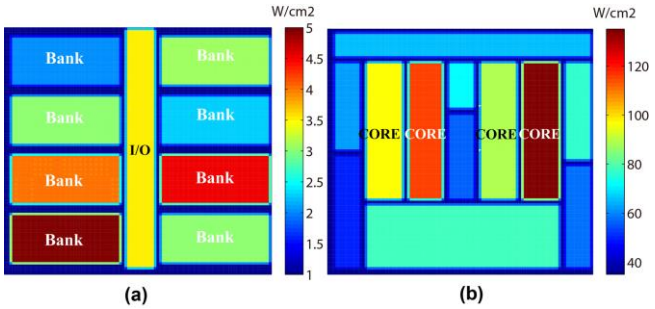

Fig. 6. Power density distribution: (a) Memory die (b) Processor die

a silicon dioxide liner thickness of 0.5 μm. A total of 10,000 TSVs are assumed to be uniformly distributed throughout the chip. There are 1,600 uniformly distributed microbumps with a diameter of 40 μm between the bottom die and the interposer.

C. *Steady state thermal comparison of the three stacks*

The two baseline stacks are shown in Fig. 2. Because the TSVs influence the decoupling results of the air-gap thermal isolation concept, we evaluate the proposed stack with and without TSVs to give a 'worst' and 'best' case analysis.

TABLE II
COMPARISON OF DIFFERENT STACKS

Unit:°C	T_{max} (Memory)	T_{max} (Processor)
Stack with air cooled heat sink	75.56	75.89
Stack with interposer embedded MFHS	61.49	62.12
Proposed stack w/o TSVs	36.69	60.08
Proposed stack with TSVs	46.76	54.46

Table II illustrates the maximum temperature of each die in all 3D stack scenarios. From the results, we find our proposed architecture has the lowest temperature. Moreover, our proposed architecture decouples the heat from the processor and the memory. In the first two scenarios, the memory exhibits a temperature value that is similar to the processor. For the proposed stack with thermal isolation, the memory temperature is only 36.69 °C even though the processor temperature is as high as 60.08 °C (assuming no TSVs).

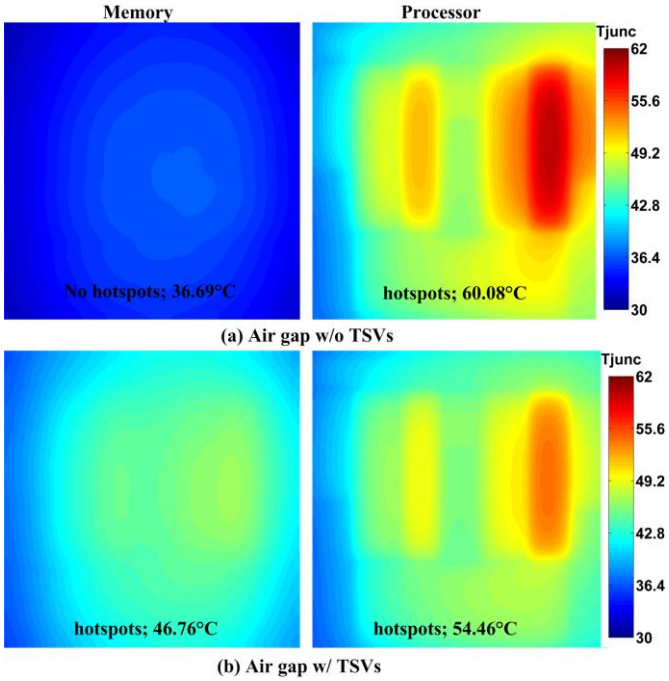

Fig. 7. Thermal maps of proposed stack (a) without TSVs (b) with TSVs

However, when TSVs are inserted in our proposed stack, the thermal coupling increases, as expected, compared to the TSV-free case. The thermal map of each die is shown in Fig. 7. From Fig. 7(b), we find the temperature distribution of the memory die to be similar to that of the processor and the temperature difference of the two dice is only 7.7°C compared to 23.4 °C in Fig. 7(a). Thus, the TSVs clearly impact the thermal isolation. Although the TSV-free case has no practical application in 3D ICs, the clustered-TSV layout design to be discussed in the next section obtains a temperature difference close to this case.

D. *Transient thermal comparison of the three stacks*

Fig 8(a) is a plot of the assumed power dissipation profile (activity) of the processor die from 0 s to 4 s. The initial power is 15 W and is increased to a peak power state (75 W) for 2 seconds and finally brought back to its initial state. Fig. 8(b) shows the processor and memory temperatures as a function of time for all three stack scenarios. Our proposed stack narrows the temperature range of the memory die. In the first two cases, the temperature range of the memory die is 38.21 °C and 28.23 °C, respectively. With air-gap isolation, the temperature variation of the memory die drops to 7.32 °C for the TSV-free case and 16.44 °C when accounting TSVs. This is useful in applications that contain temperature sensitive components.

IV. THE IMPACT OF DESIGN PARAMETERS

In this section, we thermally study our proposed 3D stack architecture as a function of the cooling capability of the thermal bridge, TSV number, TSV diameter, and TSV distribution. If not specified, the parameters and power maps are the same as those used in the previous section.

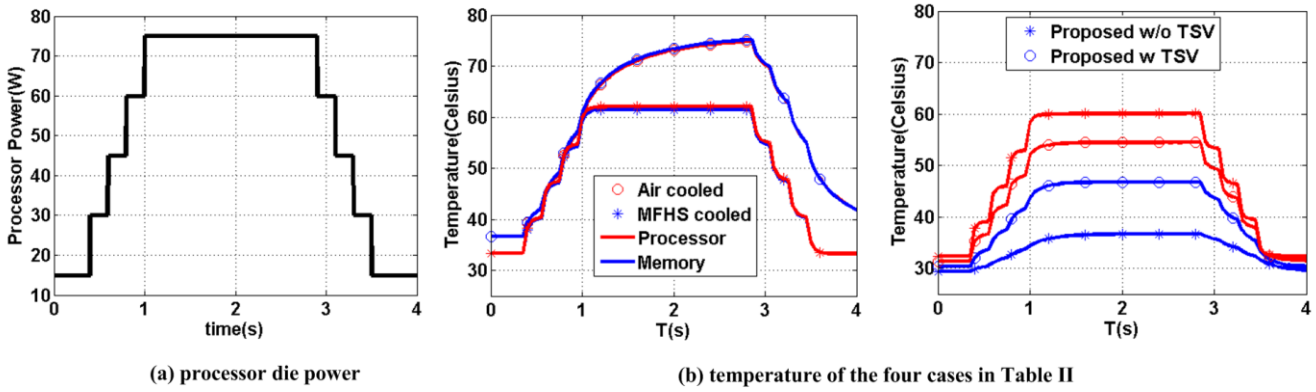

Fig. 8. (a)The assumed power trace of processor die (b) The temperature varying with time in four cases of Table II

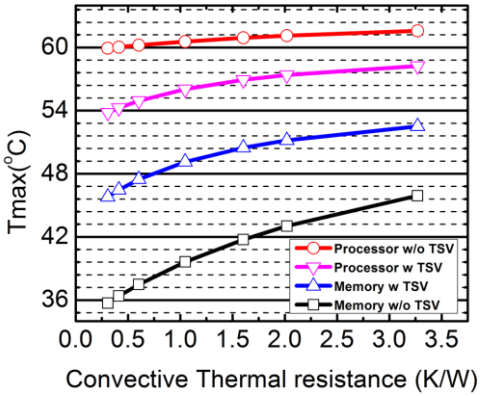

Fig. 9. Thermal impact of the thermal bridge. The x-axis denotes the convective boundary thermal resistance of the bridge fins and the y-axis is the maximum temperature of the dice.

A. Thermal bridge

The thermal bridge influences the maximum temperature of the stack. If its thermal resistance is too high, the isolated die will not be cooled effectively, resulting in a higher temperature.

Therefore, we plot the maximum temperature of each die as a function of the total convective boundary thermal resistance (applied at both fins of the thermal bridge in Fig 3(b)); this is done for the 3D stacks with and without TSVs. The results are shown in Fig. 9. When the boundary thermal resistance of the thermal bridge increases, it has a relatively minimal impact on the temperature of the processor die but has a large temperature impact on the memory die, as expected. This is because the air gap decouples the two dice and the thermal bridge mainly cools the memory die. Fortunately, in our proposed system, we use a separate (isolated) microfluidic heat sink to cool the thermal bridge, and this leads to a boundary thermal resistance of 0.463 K/W. Thus, the memory die can be cooled down effectively.

B. Impact of TSVs

(1) TSV number and diameter

The diameter and number of TSVs impacts the equivalent thermal resistance of the thermal air-gap isolation. In order to evaluate the impact of the TSVs, we simulated the following scenarios: first, we fix the TSV diameter to 5 µm and TSV liner to 0.5 µm and sweep the number of TSVs from 1,600 to 10,000. Next, we fix the total number of TSVs to 10,000 and sweep the TSV diameter from 2 µm to 10 µm (TSV liner thickness unchanged).

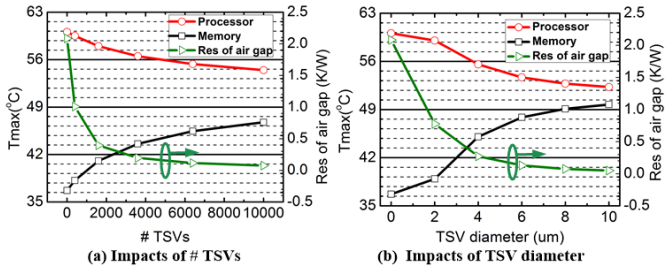

Fig. 10. The impact of the TSVs: (a) number of TSVs, (b) diameter of TSVs.

Fig. 10(a) and 10(b) show the impact of the number and diameter of the TSVs, respectively. As the TSV total volume increases, the air-gap layer becomes more thermally conductive, and the inter-die heat coupling becomes stronger thereby reducing the temperature difference between the two dice. If 2 µm diameter TSVs are used rather than 10 µm, the memory temperature is 38.92 °C compared to 49.73 °C for the 10 µm diameter TSV case. Further scaling of the TSV dimensions will yield additional improvements in the thermal isolation of the air-gap.

(2) TSV distribution

In 3D ICs, the TSVs are good heat conductors due to the high thermal conductivity of copper. The area in which TSVs are located will undergo stronger thermal coupling. Thus, the placement of the TSVs is an important thermal consideration. When the TSVs are clustered (such as in wide-I/O technology), increased thermal coupling is expected to occur locally. In this manner, the thermal coupling from the processor die will be localized in the memory die.

For the memory die, the clustered TSVs act as the I/O pins and are outside of the memory cell circuits (labeled by a dashed rectangle in Fig. 11(a)). Hence, the memory cell circuits will become relatively ‘free’ from the thermal impact of the processor because there are no TSVs in the area of the cell circuits.

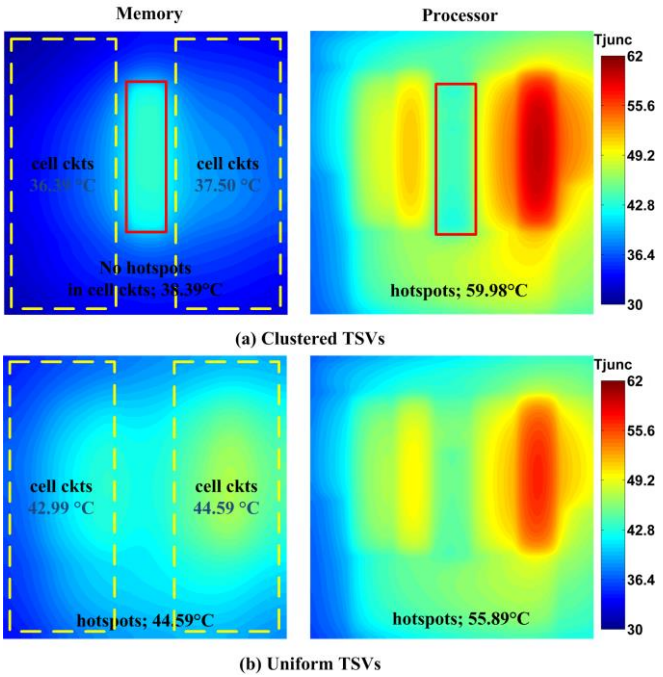

Fig. 11. Thermal maps of clustered TSVs and uniform TSVs. (a) TSVs are clustered in the solid-line box; (b) The same number of TSVs is uniformly distributed

To this end, we cluster the TSVs only in the center. The resulting TSV cluster is assumed to be 1 mm × 5 mm and is assumed to have 49×100 TSVs, which is labeled by the solid rectangle in Fig. 11(a). For a fair comparison, we also consider uniformly distributed TSVs (4,900 total TSVs). The results are shown in Fig. 11(b).

In the clustered TSV case, the maximum temperature of the whole DRAM die drops by 6.20 °C compared to the uniformly distributed TSV case. Moreover, the maximum temperature in the cell circuits area is only 37.50 °C, which is a drop of 7.09 °C and is much closer to the 36.69 °C reported for the (ideal) TSV-free case discussed in Section III.C. By clustering the TSVs far from the memory cells, the most thermally-sensitive portion of the die becomes more isolated from the high power die.

V. CONCLUSION

This paper evaluates air-gap isolation as a potential technology for thermal decoupling in heterogeneous 3D ICs. In the evaluated memory-processor stack with air-gap isolation, the memory temperature is reduced by 38.9 °C compared to conventional bonding with underfill. To maintain the thermal benefits of an air gap, the thermal bridge should provide sufficient cooling to remove the heat from the memory die. Likewise, the TSVs should be carefully designed taking the thermal effects of their number, diameter and layout into account.

ACKNOWLEDGMENT

The authors would like to thank the support from Defense Advanced Research Projects Agency (DARPA) with Grant # N66001-12-1-4240.

REFERENCES

[1] Y. Zhang, et al, "3D stacked on-demand microfluidic cooling for high performance 3D ICs," *in Proc. IEEE Electronic Components and Technology Conference*, 2012, pp. 1644-1650.

[2] H. Oprins, et al, "Numerical and experimental characterization of the thermal behavior of a packaged DRAM-on-logic stack," *in Proc. IEEE 62nd Electronic Components and Technology Conference (ECTC),* pp.1081-1088, 2012

[3] J. Liu, et al, "RAIDR: Retention-aware intelligent DRAM refresh," *in Proc. IEEE/ACM 39th Annual International Symposium on Computer Architecture (ISCA)*, pp.1-12, 2012

[4] Z. Li, et al, "Reliability Modeling and Management of Nanophotonic On-Chip Networks," *IEEE Trans. Very Large Scale Integr.(VLSI) Syst* vol.20, no.1, pp.98-111, 2012

[5] R. Nadipalli, et al, "3D integration of MEMS and CMOS via Cu-Cu bonding with simultaneous formation of electrical, mechanical and hermetic bonds," *in Proc. IEEE International 3D Systems Integration Conference(3DIC),* pp.1-5, 2012

[6] J. Yang, et al, "Thermal Calibration for the Accelerometer Triad Based on the Sequential Multiposition Observation," *IEEE Trans. Instrum. Meas.*, vol.62, no.2, pp.467,482, Feb. 2013

[7] Y. Zhang, et al, "Thermal Design and Constraints for Heterogeneous Integrated Chip Stacks and Isolation Technology Using Air Gap and Thermal Bridge," *IEEE Trans. Compon., Packag. Manuf. Technol.,* 2014

[8] J. Xie, M. Swaminathan, "Electrical-thermal co-simulation of 3D integrated systems with microfluidic cooling and Joule heating effects," *IEEE Trans. Compon., Packag. Manuf. Technol.,* vol. 1, no. 2, pp. 234–246, Feb. 2011.

[9] J. Meng, et al, "Optimizing energy efficiency of 3-D multicore systems with stacked DRAM under power and thermal constraints," *in Proc. ACM/EDAC/IEEE 49th Design Automation Conference (DAC)*, pp. 648-655, 2012

[10] J. Xie, M. Swaminathan, "3D transient thermal solver using non-conformal domain decomposition approach," *in Proc. IEEE/ACM International Conference on Computer-Aided Design (ICCAD),* pp. 333–340, 2012.

[11] Z. Wan, et al, "Electrical/Thermal Co-Design of Multi-Core Architectures and Microfluidic Cooling for 3D Stacked ICs," in Proc. *19th Int. Workshop on Thermal Investigations of ICs and Systems*, 2013, pp. 237-242.

[12] Intel, *Intel Core™ i7Processor Families for the LGA-2011-0 Socket, Thermal Mechanical Specification and Design Guide,* Intel Corporation, November 2013.

[13] U. Kang, et al, "8Gb 3D DDR3 DRAM Using Through-Silicon-Via Technology," *in Proc. IEEE International Solid State Circuits Conference (ISSCC),* pp. 130–132, February 2009.

[14] "3rd Generation Intel® Core™ Processor Family Quad Core Launch Product Information," Intel, [online]. Available: http://download.intel.com/newsroom/kits/core/3rdgen/pdfs/3rd_Generation_Intel_Core_Product_Information.pdf ,APR, 2012

[15] Micron Technology, "DDR3 SDRAM System-Power Calculator," July 27, 2011

[16] S. Li, et al, "Mc-PAT: an integrated power, area, and timing modeling framework for multicore and manycore architectures," *In Proc. 42nd Annual IEEE/ACM International Symposium on Microarchitecture,* pp. 469-480. 2009