

Within-Tier Cooling and Thermal Isolation Technologies for Heterogeneous 3D ICs

Yue Zhang, Hanju Oh, and Muhannad S. Bakir

School of Electrical & Computer Engineering, Georgia Institute of Technology, Atlanta, GA

Abstract—Thermal management in 3D ICs not only requires cooling, but may also require thermal isolation in scenarios in which high-power chips (e.g. logic chips) are stacked along with low-power and temperature-sensitive tiers (e.g. memory or silicon nanophotonic chips). A hybrid thermal solution combining within-tier microfluidic cooling for the high-power tier and within-tier thermal isolation for the low-power tier is proposed for the first time. In this paper, we report 1) within-tier microfluidic cooling in a processor-on-processor stack 2) TSVs with 23:1 aspect ratio integrated in the microfluidic heat sink, and 3) the integration of air/vacuum cavity in the low-power tier to ‘protect’ it from the temperature variation and nonuniformity of the high-power chip. Thermal modeling shows that the low-power tier temperature only increases by 4 °C when the power density of the processor tier increases from 50 W/cm² to 100 W/cm², compared to 22 °C temperature increase without thermal isolation.

Index Terms— 3D ICs; thermal management; high aspect ratio through-silicon via (TSV); air/vacuum cavity; thermal isolation.

I. INTRODUCTION

Three-dimensional ICs offer opportunities for improving performance and reducing power dissipation by enabling shorter on- and off-chip interconnects and heterogeneous integration [1]. Due to the increased power densities and the thermal resistance of “inner” tiers, heat removal in 3D ICs is challenging with top-attached air-cooled heat sink (ACHS) [2]. The problem is exacerbated when multiple high-power tiers are stacked. Embedded microfluidic cooling is considered a promising solution to solve the described thermal issue due to its high cooling capability, microscale form factor, and CMOS microfabrication compatibility [3][4]. Of course, development of TSVs that are compatible with the microfluidic cooling solution is critical and will be described in this paper.

However, in some applications where high-power chips (e.g. logic chips) are stacked along with low-power and temperature-sensitive components (memory or silicon nanophotonic chips, for example), thermal management will not only require effective cooling, but may also require effective thermal isolation to ‘protect’ the temperature-sensitive components from the time-varying power dissipation of other chips in the stack. By placing such tiers next to each other, the thermal coupling between them will be significant, leading to possibly undesirable junction temperature variation

in the temperature-sensitive tier as a result of the high-power chips. In the case of silicon nanophotonics, stacking such chips in a stack adjacent to logic and memory has been explored [5][6]. However, the temperature sensitivity of the optical elements presents a great challenge for integration; for example, a microring modulator with 5 μm diameter is reported to have a wavelength drift of 0.11 nm/°C in [7]. A temperature change of 13.5 °C will result in a complete passband mismatch between transmitter-receiver pairs in 64-channel wavelength-division multiplexing (WDM) [6]. In applications involving a memory stack, it has been shown that stacking logic on memory can cause a 30-to-40 °C temperature increase in the SRAM tier [8]. The increased temperature not only causes the leakage power to increase by approximately 2 times, but also causes the average cache access time to increase by 50 ps (28% performance degradation) [8]. As such, there is a need for wafer-level batch fabricated within-tier thermal isolation technologies in order to minimize thermal coupling between the high-power logic chip and the low-power and temperature-sensitive chips in the stack. To this end, we propose within-tier microfluidic cooling (for heat removal, when needed) and within-tier thermal isolation by forming micro-scale air gaps/vacuum cavities. An example illustration of a logic-silicon nanophotonic stack using within-tier cooling and isolation is shown in Fig. 1. The figure illustrates a high-power processor tier with an embedded microfluidic heat sink (MFHS) stacked above a low-power silicon nanophotonic chip with air/vacuum cavities formed on its back-side (side facing the logic chip) to provide thermal isolation. The power of the silicon nanophotonic chip is dissipated through the silicon interposer. Note that the within-tier thermal isolation concept using air/vacuum cavities is independent of the within-tier microfluidic cooling; it can be used with an air-cooled heat sink when microfluidic cooling is not necessary or is not an option. Forming the air/vacuum thermal isolation cavities is simple as it only requires back-side silicon etching of the silicon nanophotonic chip. The device layer that contains the nanophotonic components, such as microring resonators, photodetectors, waveguides, and couplers, faces the silicon interposer. Local thermal tuning is commonly used to maintain the microring resonators at a constant temperature [10]. Local microcavities beneath the temperature-sensitive components have been explored in [9] and [10] to reduce the thermal coupling from the surrounding area. In [10], a local undercut microcavity is created beneath the resonator. The undercut is shown to reduce the tuning power by an order of magnitude. However, to our knowledge, there has been little work on thermally decoupling the silicon

nanophotonic or memory chips from the adjacent tiers in a 3D stack. Adopting the chip-scale air/vacuum cavity in the low-power chip brings about smaller temperature variation to deal with. Moreover, the local thermal isolation method based on undercut microcavity can be built in addition to ensure a constant local temperature [10]. Thus, by combining our chip-scale air/vacuum cavity with the local undercut microcavity, along with the 3D stack architecture, new opportunities for improved heterogeneous system integration and miniaturization become possible.

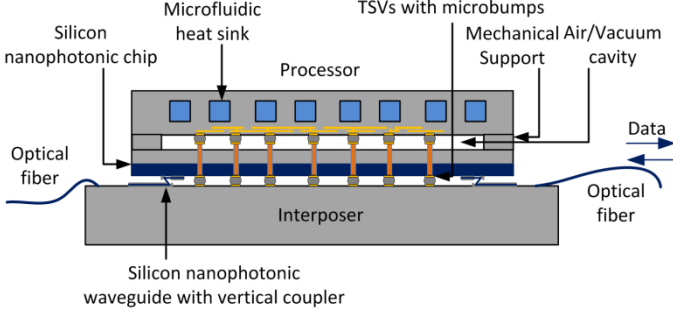


Fig. 1. A cross-sectional view of a 3D stack of processor and silicon nanophotonic chip with hybrid thermal management solution: within-tier microfluidic cooling in processor and air/vacuum cavity to thermally isolate the silicon nanophotonic chip.

II. WITHIN-TIER MICROFLUIDIC THERMAL TESTBED AND EXPERIMENTAL RESULTS

This section of the paper reports experimental results relating to within-tier microfluidic cooling.

A. Within-tier microfluidic thermal testbed

A 2-tier microfluidic cooled thermal testbed was developed to benchmark against an air-cooled 2-tier stack. The details of the fabrication process are reported in [11]. The MFHS utilizes a micropin-fin array with the following dimensions: micropin-fin diameter of $150\ \mu\text{m}$, a pitch of $225\ \mu\text{m}$ and a height of $200\ \mu\text{m}$, as labeled in Fig. 2. Each tier has 4 independent Pt-based thin-film segmented heaters/thermometers in order to emulate the heating of a simplified multicore microprocessor (Fig. 3). The dimensions of each heater are $0.22\ \text{cm} \times 1\ \text{cm}$ with a spacing of $0.03\ \text{cm}$ between them. The total heating area in each tier is $1\ \text{cm} \times 1\ \text{cm}$. The heaters are independently controlled. Moreover, each tier has its own set of inlet and outlet ports allowing independent flow rate control within each tier. The two tiers are bonded using a thermal interface material with a thermal resistance of $0.28\ \text{K/W}$.

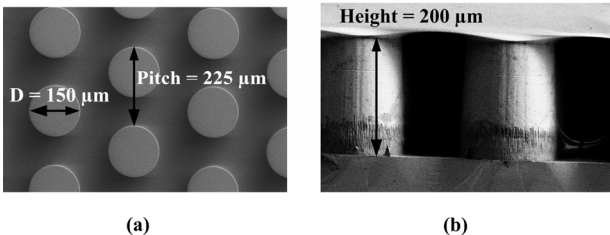


Fig. 2. (a) Top and (b) cross-section views of the silicon micropin fin heat sink in each tier.

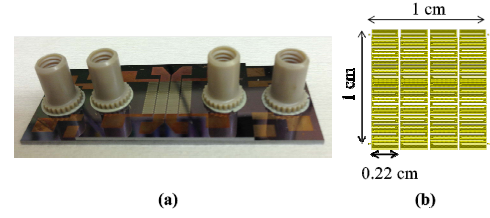


Fig. 3. (a) Micrograph of a 2-tier chip thermal testbed with segmented heaters. (b) Illustration of the segmented heaters on each tier.

Two gear pumps are connected to the two tiers in the testbed to supply DI water independently. The temperature of the inlet DI water is $20 \pm 1\ ^\circ\text{C}$. The temperature of each segmented heater is sensed by the on-chip Pt heater/thermometer and recorded at a rate of $1\ \text{Hz}$ using an Agilent data logger.

In order to attain an initial insight into the benefits of embedded microfluidic cooling, a 3D air-cooled heat sink (ACHS) testbed is constructed similarly without the embedded microfluidic heat sink. This will be used to benchmark the thermal results to the MFHS cooled chip.

B. Single tier with uniform power dissipation

To capture the lateral temperature increase as the coolant flows from the inlet to the outlet ports, uniform power dissipation is applied in a single tier. Fig. 4 illustrates the temperature of each heater on the chip as the total chip power density ramps from $25\ \text{W/cm}^2$ to $100\ \text{W/cm}^2$. The DI water flow rate is $80\ \text{mL/min}$ in all of the measurements unless specified otherwise. In the high power density case ($100\ \text{W/cm}^2$), the junction temperature of heater 4 (i.e., the heater closest to the outlet) increases by $33\ ^\circ\text{C}$ while that of location 1 increases only by $17\ ^\circ\text{C}$. This result is expected since the coolant temperature increases as it flows from inlet to outlet, and thus, the chip junction temperature also increases [3]. The chip design was simulated under a power density of $100\ \text{W/cm}^2$ using ANSYS fluent. The results are also included in Fig. 4 for reference. The difference between the experimental

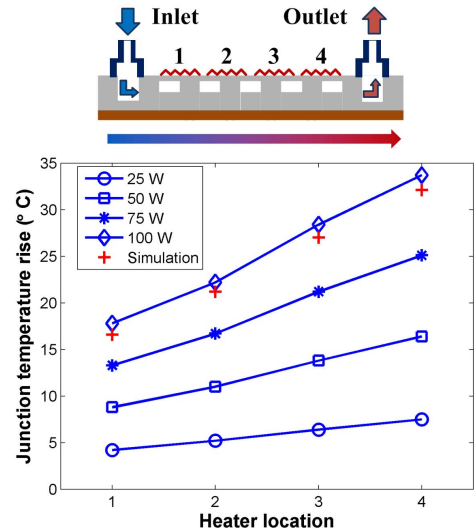


Fig. 4. Junction temperature rise at different heater locations on the chip for different power dissipations. ANSYS simulation for $100\ \text{W}$ case is also plotted for reference.

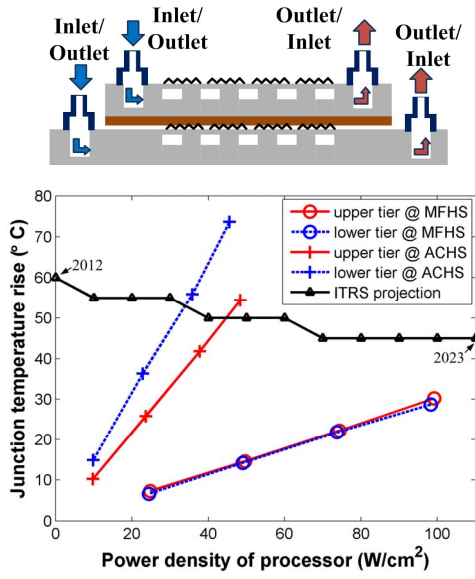


Fig. 5. Junction temperature rise in a processor-on-processor stack under MFHS and ACHS.

results and the simulations is less than 1.6 °C. The lateral thermal gradient across the chip becomes exacerbated for higher power densities. One way to mitigate the thermal gradient is to increase the flow rate. However, the pressure drop and the pumping power will increase. An alternate approach was reported in [12].

C. Processor-on-processor stack

In Fig. 5, the results from microfluidic- and air-cooled 2-tier chip stacks that dissipate up to 100 W/cm² per tier are shown. A MFHS is integrated into each tier; the flow rate in each tier is 100 mL/min. Two sets of measurements were performed for the MFHS cooled stack, and the average junction temperature rise above the inlet coolant temperature in each chip is plotted in Fig. 5. The difference in the two measurements did not exceed 1.1 °C. As seen from the plots, when the power density in each tier is 100 W/cm², the junction temperature rise in either tier is 30 °C, resulting in an absolute junction temperature of 50 °C. Similar experiments were conducted with ACHS. The results are also plotted in Fig. 5 for comparison. The testbed under ACHS has a temperature rise of more than 54 °C at 50 W/cm². The maximum junction temperature rise trend according to ITRS is also plotted as a reference. At 50 W/cm², the junction temperature of both tiers under air-cooling exceeds the ITRS projection. The processor-on-processor stack cooled using MFHS can dissipate more than 100 W/cm² in each tier without reaching the projected maximum junction temperature. Please note that the thermal results obtained by ACHS testbed may have been better if a better thermal interface material was used.

III. TSVs INTEGRATED IN MICROPIN-FINS

Due to the insertion of the 200 µm tall micropin-fin heat sink, the total wafer thickness must increase for 3D ICs. A

high aspect ratio TSV technology is developed to be compatible with the embedded MFHS. This section reports the fabrication process and results of the high aspect ratio TSVs within micropin-fins. This work extends our previous paper [13], where TSVs with an aspect ratio of 18:1 are integrated in micropin-fins. A summary of the fabrication process flow is shown in Fig. 6. In Step 2, high aspect ratio TSVs (~23:1) are fabricated using a standard Bosch process that alternates between SF₆ (plasma etch step) and inert C₄F₈ (passivation step). Thermal oxide is then grown to electrically isolate the TSVs from the substrate in Step 3. Titanium and copper are deposited onto the back side of the wafer as a seed layer in Step 4. Next, vias are pinched off through an electroplating process forming a conductive layer at the bottom of the TSV hole. Using this newly formed layer, bottom-up pulsed electroplating is performed to fill the vias with copper in Step 5 with Enthone DVF plating solution. Following electroplating, the sample is polished and excess silicon is removed using chemical mechanical polishing (CMP). A second Bosch process is used to fabricate the micropin-fins in Step 6. The fabricated TSVs are 13 µm in diameter and 300 µm deep (23:1). An SEM image of an angled view of TSVs integrated in micropin-fins is shown in Fig. 7 (a). Each micropin-fin contains a 4 × 4 array of copper filled TSVs. A cross-sectional view of the fabricated TSVs within the micropin-fin is shown in Fig. 7 (b). The TSVs are fully plated with no voids. In summary, a die spanning 1 cm x 1 cm can have 1,936 micropin-fins with a diameter of 150 µm at a pitch of 225 µm. Each micropin-fin has 16 electrical TSVs, providing a total of 30,976 electrical I/Os that can be used to connect adjacent tiers of the 3D stack.

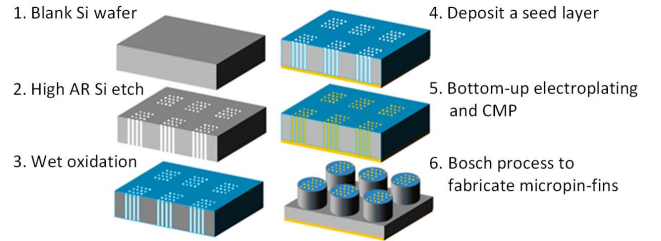


Fig. 6. Process flow of integrating TSVs in micropin-fin heat sink.

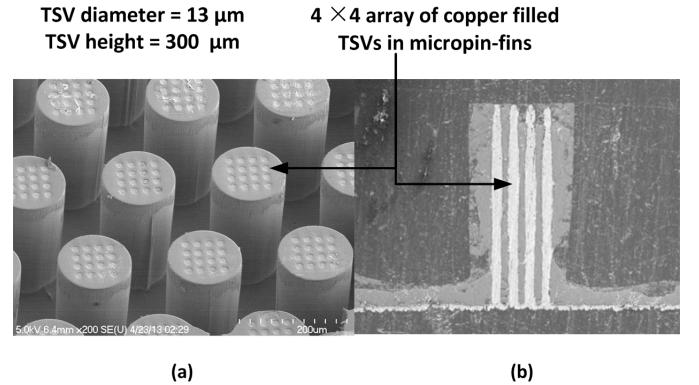


Fig. 7. SEM images of a angled view (a) and a cross-sectional (b) view of the high aspect ratio TSVs integrated in micropin-fins (TSV dimensions: 13 µm diameter, 24 µm pitch and 300 µm tall).

IV. WITHIN-TIER AIR/VACUUM CAVITY ISOLATION

This section of the paper reports results relating to within-tier air/vacuum cavity isolation. As noted in the introduction, stacking of high-power chips along with low-power and temperature-sensitive chips presents challenges. The time-dependent temperature variation in the processor tier, which is workload dependent, is directly coupled to nearby stacked chips. In current 3D IC technology, it is common to use an adhesive with high thermal conductivity between stacked chips to ensure the thermal resistance between each tier as small as possible. This method helps remove the heat from within the stack to the top most of the stack where an air-cooled heat sink is placed. However, the thermally conductive adhesive will also enhance the thermal cross-talk between the processor tier and other chips in the stack that are temperature-sensitive (memory and silicon nanophotonics, for example) leading to temperature variation. We propose using a hybrid thermal solution that combines cooling and within-tier thermal isolation to enable the stacking of logic and low-power temperature-sensitive chips. Fig. 8 illustrates an example of the hybrid thermal solution. The 3D stack contains a processor and a low-power chip (i.e. bottom chip). In this case, an air-cooled heat sink is used for the high-power processor. An air/vacuum cavity is integrated within the bottom chip to thermally decouple the two tiers, reducing the temperature variation in the bottom chip. Air, with a thermal conductivity of 0.026 W/m-K, is good for thermal isolation. The way to create the air cavity is to etch a thin layer (a few micrometers) of silicon from the back of the silicon nanophotonic chip. Vacuum is reported to have a thermal conductivity as low as 0.004 W/m-K below 10 Pa [14]. Creating and maintaining vacuum at such low pressure was reported in [15] using Si-Si direct bonding.

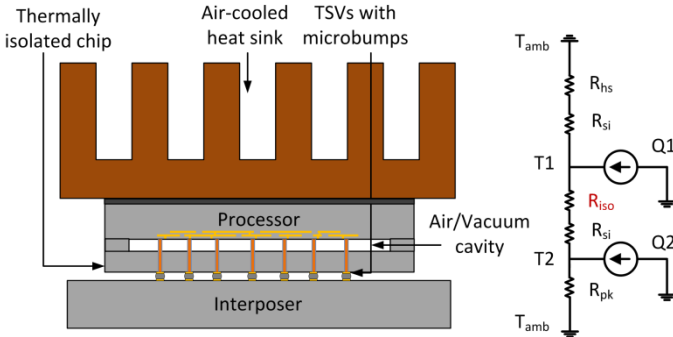


Fig. 8. Air/Vacuum cavity integrated between the logic and thermally isolated tier (left) and the corresponding thermal resistance network (right).

The modeling effort begins by neglecting the impact of TSVs within the air/vacuum cavity. Using the 1D thermal resistance network model in Fig. 8, the junction temperature of the bottom tier as a function of the power dissipated in the logic chip is plotted in Fig. 9 (air-cooling) and Fig. 10 (microfluidic cooling). The values assumed in this model are shown in Table I. The results are also confirmed with ANSYS simulations. When the power density of the processor tier increases from 50 W/cm² to 100 W/cm², the processor temperature increases from 46 °C to 68 °C. This trend is

Table I
VALUES OF THE VARIABLES IN THE THERMAL RESISTANCE NETWORK (FIG. 8)

Variables	Value
R_{hs} (K/W)	0.5 (ACHS)/ 0.25 (MFC)
R_{pk} (K/W)	2.5
Thickness of each tier (μ m)	50
Thickness of air gap (μ m)	5
$Q2$ (W)	5

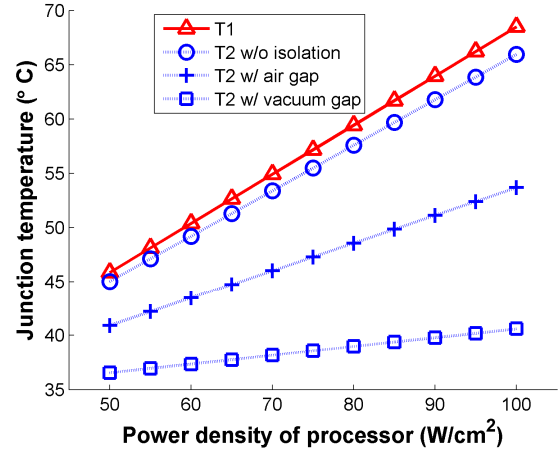


Fig. 9. Junction temperature of processor tier (T1) and the low-power tier (T2) w/ and w/o thermal isolations under air cooling (i.e. Fig. 8).

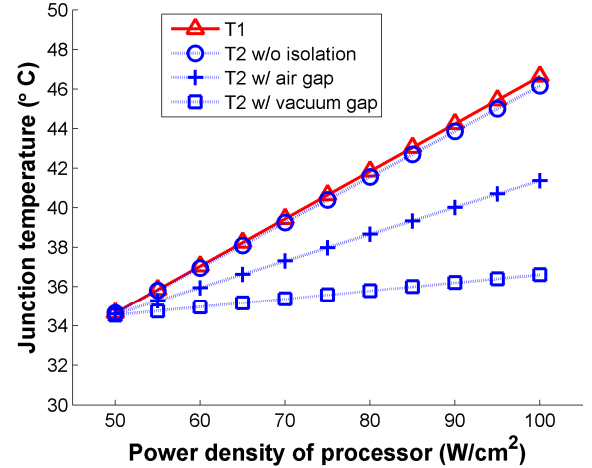


Fig. 10. Junction temperature of processor tier (T1) and the low-power tier (T2) w/ and w/o thermal isolations under microfluidic cooling (i.e. Fig. 1).

similar with and without thermal isolation. Without any thermal isolation, the temperature of the bottom tier follows the same trend. With a 5 μ m thick air cavity, the temperature of the bottom tier increases with a smaller slope from 41 °C to 54 °C. If vacuum is created between the two tiers, the temperature of the bottom tier will only increase by 4 °C. According to [6], a 13.5 °C temperature difference will cause a 1.48 nm wavelength drift on a microring resonator with a radius of 5 μ m, leading to communication error in a 64-

channel WDM. In Fig. 10, with microfluidic cooling, the junction temperature increase in the bottom tier is smaller than all the corresponding cases with air cooling. With an air cavity, the temperature of the bottom chip varies by 6 °C, which is half of the case in air cooling. A smaller temperature change can be achieved by increasing the thickness of the air/vacuum cavity.

ANSYS simulations were performed to analyze the impact of routing TSVs through the air/vacuum cavity. The TSVs are partially embedded in the bottom chip and partially exposed in the air/vacuum cavity (as shown in Fig. 11). Due to the heat conduction through the TSVs, the thermal coupling between the two tiers is stronger. The results are plotted in Fig. 12. TSVs are assumed to have a diameter of 2 μm and a silicon dioxide liner of 1 μm on a 100 μm $\times$ 100 μm pitch. The temperature change of the bottom tier is more obvious with TSVs. The bottom tier suffers from a temperature increase of 16.5 °C when TSVs are used and 13 °C without the TSVs. The results are expected since TSVs with copper cores have good thermal conductivity, causing undesired thermal coupling between the two tiers. One potential option to eliminate this effect is to use contactless communications such as capacitive/inductive coupling for signaling [16] and TSV for power/ground only. Of course smaller diameter TSVs and perhaps lower thermal conductivity TSVs would be helpful to reduce the thermal coupling. Smaller diameter TSVs would also have smaller capacitance.

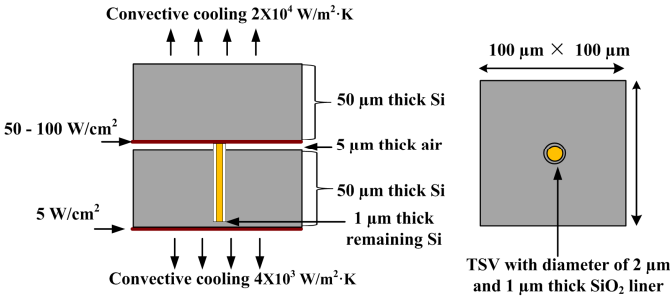


Fig. 11. Cross-sectional (left) and top (right) view of the structures simulated in ANSYS to represent the case where TSVs are routed through the air cavity.

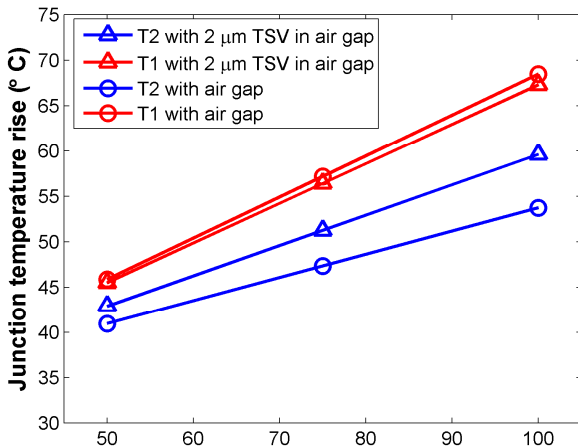


Fig. 12. ANSYS simulation results of junction temperature of the two tiers when TSVs are routed through the air gap.

V. CONCLUSION

This paper explores a hybrid thermal solution for heterogeneous 3D ICs. The hybrid solution features within-tier cooling for high-power processor tiers and within-tier thermal isolation for the low-power and temperature-sensitive tiers, e.g. processor and memory/silicon nanophotonic chip. Within-tier microfluidic cooling technology is firstly presented for the cooling of the high-power unit. The MFHS is shown to maintain the temperature of a 2-tier processor-on-processor stack below 50 °C at 200 W/cm². TSVs with 23:1 aspect ratio are integrated with the micropin-fin heat sink, resulting in a density of 30,976 TSVs/cm². The use of air/vacuum cavities to thermally isolate tiers in a die is proposed for the first time. Thermal modeling demonstrates that, without the impact of TSVs, the use of an air or vacuum cavity will cause a 13 °C or 4 °C temperature increase, respectively, in the thermally isolated tier when the power density of the processor tier increases from 50 W/cm² to 100 W/cm². If no isolation is used between the tiers, the temperature increase in the bottom tier will be 22 °C. If the impact of TSVs is considered, the thermally isolated tier will suffer from a temperature increase of 16.5 °C according to ANSYS simulation.

ACKNOWLEDGMENT

This work has been carried out in part under support from Defense Advanced Research Projects Agency (DARPA) with Grant # N66001-12-1-4240.

REFERENCES

- [1] P. G. Emma and E. Kursun, "Is 3D chip technology the next growth engine for performance improvement?," *IBM J. of Research and Development*, vol. 52, pp. 541-552, 2008.
- [2] S. M. Sri-Jayantha, G. McVicker, K. Bernstein, and J. U. Knickerbocker, "Thermomechanical modeling of 3D electronic packages," *IBM J. of Research and Development*, vol. 52, pp. 623-634, 2008.
- [3] D. B. Tuckerman and R. F. W. Pease, "High-performance heat sinking for VLSI," *IEEE Electron Device Lett.*, vol. 2, pp. 126-129, 1981.
- [4] T. Brunswiler, S. Paredes, U. Drechsler, B. Michel, W. Cesar, G. Toral, Y. Temiz, and Y. Leblebici, "Validation of the porous-medium approach to model interlayer-cooled 3D-chip stacks," in *Proc. of 3D System Integration*, 2009, pp. 1-10.
- [5] D. Brunina, D. Liu, and K. Bergman, "An Energy-Efficient Optically Connected Memory Module for Hybrid Packet- and Circuit-Switched Optical Networks," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 19, 2013.
- [6] L. Zheng et al., "Reliability Modeling and Management of Nanophotonic On-Chip Networks," *IEEE Trans. on VLSI System*, vol. 20, pp. 98-110, 2012.
- [7] S. Manipatruni et al., "Wide temperature range operation of micrometerscale silicon electro-optic modulators," *Opt. Lett.*, vol. 33, no. 19, pp. 2185-2187, 2008.
- [8] S. Chatterjee, C. Mink, R. Rao, and S. Mukhopadhyay, "Impact of die-to-die thermal coupling on the electrical characteristics of 3D stacked SRAM cache," in *Proc. of IEEE Semi-Therm*, 2012, pp. 14-19.
- [9] C. Holzwarth et al., "Localized substrate removal technique enabling strong-confinement microphotonic in bulk si cmos processes," in *Proc. of Lasers and Electro-Optics and Conf. on Quantum Electron. Laser Sci.*, 2008, pp. 1-2.
- [10] P. Dong et al., "Thermally tunable silicon racetrack resonators with ultralow tuning power," *Opt. Express*, vol. 18, 2010.
- [11] Y. Zhang, A. Dembla, and M. S. Bakir, "A Silicon Micropin-fin Heat Sink with Integrated TSVs for 3D ICs: Trade-off Analysis and Experimental Testing," *IEEE Trans. on Components, Packaging and Manufacturing Technology*, to be published.

- [12] Y. Zhang, L. Zheng, and M. S. Bakir, "Tier-Independent Microfluidic Cooling for Heterogeneous 3D ICs with Nonuniform Power Dissipation," in *Proc. of IEEE International Interconnect Technology Conference*, 2013.
- [13] A. Dembla, Y. Zhang, and M. S. Bakir, "High aspect ratio TSVs in micropin-fin heat sinks for 3D ICs," in *Proc. of IEEE Nanotechnology*, 2012, pp. 1-6.
- [14] D. Randjelovic et al., "Vacuum MEMS Sensor Based on Thermopiles – Simple Model and Experimental Results," in *Proc. of International Conference on Microelectronics*, 2008.
- [15] I. Toshihiro et al., "Room Temperature Vacuum Sealing Using Surface Activated Bonding Method," in *Proc. of International Conference on Solid State Sensors, Actuators and Microsystems*, 2003, pp. 1828-1831.
- [16] A. Fazzi et al., "3-D Capacitive Interconnections for Wafer-Level and Die-Level Assembly," *IEEE J. of Solid-State Circuits*, vol. 42, pp. 2270-2282, 2007.