Silicon Interposer with Embedded Microfluidic Cooling for High-Performance Computing Systems

Li Zheng[1], Yang Zhang, Xuchen Zhang and Muhannad S. Bakir[2]
School of Electrical and Computer Engineering
Georgia Institute of Technology
791 Atlantic Drive NW
Atlanta, Georgia 30332

[1]lizheng@gatech.edu
[2] muhannad.bakir@mirc.gatech.edu

Abstract

A silicon interposer platform utilizing microfluidic cooling is proposed to address the off-chip signaling and cooling challenges facing future high-performance computing systems. A test vehicle with microfluidic I/Os and a micropin-fin heat sink was used to evaluate microfluidic cooling performance. De-ionized water (~20 °C) was used as the coolant. At a flow rate of 50 mL/min, the measured temperature was 55.9 °C for a power density of 97.0 W/cm^2. Compared to air cooling, microfluidic cooling significantly improves cooling performance and thermal isolation. Moreover, 3-D integration of two silicon dice with microfluidic I/Os on a silicon interposer is demonstrated.

I. Introduction

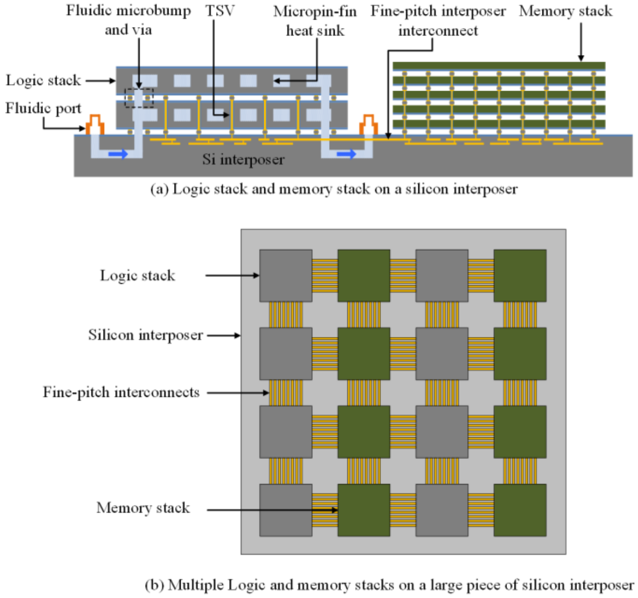
Off-chip signaling and cooling are two major challenges facing future high-performance computing systems. Recently, silicon interposer technology has been widely explored due to its potential for high-bandwidth-density signaling [1], [2]. Data rates of 10 Gbps per channel using silicon interposer interconnects 2 µm to 6 µm wide and up to 6 cm long have been demonstrated [3]. Moreover, silicon interposer technology enables heterogeneous integration of logic, memory, MEMS, and optoelectronics and in many instances leads to a smaller form factor system and reduction in the thermal stresses due to the coefficient of thermal expansion match to silicon dice [1], [2], [4].

Microfluidic cooling, originally proposed by Tuckerman et al. in 1981 [5], has been demonstrated as a promising solution to large power density electronic systems. More recently, Zhang et al. demonstrated cooling of a two-die stack with 100 W/cm^2 per tier at a maximum junction temperature of 47 ºC utilizing a staggered micropin-fin heat sink [6]. Moreover, microfluidic I/Os consisting of solder-based fluidic microbumps and fluidic vias, which enable coolant delivery from a silicon interposer to an on-die microfluidic heat sink, have been demonstrated [7].

In this paper, we propose a silicon interposer platform utilizing microfluidic cooling for high-performance 3-D computing systems, as shown in Fig. 1. A logic stack, which has an embedded microfluidic heat sink in each tier, and a memory stack are assembled side-by-side on a silicon interposer. High-bandwidth signaling between the two stacks is achieved using fine-pitch interconnects on the silicon interposer. With a large silicon interposer, multiple logic and memory stacks can be integrated to form a high-performance and high-capacity system, as shown in Fig. 1(b).

Microfluidic cooling is critical to the proposed platform since stacking of high-performance logic dice would lead to a large power density. As shown in Fig. 1(a), a coolant is pumped into the fluidic channels in the silicon interposer and distributed to the microfluidic heat sink in each tier through the fluidic microbumps and vias. Thus, each tier in the stack can be equally cooled. Fig. 2 (a) shows SEM images of the microfluidic I/Os (fluidic via and microbumps), micropin-fin heat sink, and fine-pitch electrical microbumps. Fig. 2 (b) shows an X-ray image of two silicon dice bonded on a silicon interposer. In addition to this envisioned system, we also envision the direct integration of an embedded microfluidic heat sink within the silicon interposer (discussed in Section II).

This paper is organized as follows: Section II discusses the thermal benefits of the proposed platform; thermal measurements using the test vehicle with microfluidic I/Os and micropin-fin heat sink are reported. Air- and microfluidic cooling configurations are compared using thermal simulations. Section III presents the 3-D stacking of two silicon dice with microfluidic I/Os on a silicon interposer. The conclusion is presented in Section IV.



(a) Logic stack and memory stack on a silicon interposer

(b) Multiple Logic and memory stacks on a large piece of silicon interposer

Fig. 1 A high-performance computing system based on a silicon interposer platform utilizing microfluidic cooling: (a) side view of a logic stack and a memory stack heterogeneously integrated on a silicon interposer, (b) an array of logic and memory stacks on a 'large' silicon interposer with fine-pitch interconnects for high-bandwidth communication

978-1-4799-8609-5/15/$31.00 ©2015 IEEE

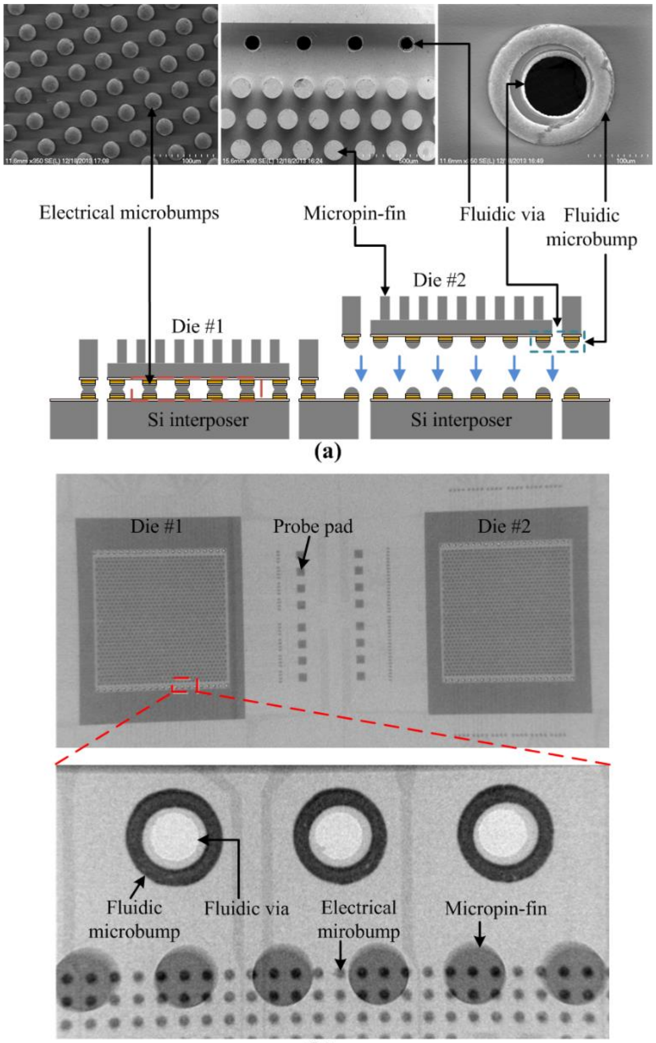


Fig. 2 SEM images of the fabricated silicon die and X-ray images of the bonded silicon dice and interposer: (a) SEM images of the electrical microbumps, fluidic vias and microbumps, and micropin-fins; two silicon dice are sequentially bonded to a silicon interposer; (b) X-ray image of two bonded silicon dice on a silicon interposer; close-up of bonded fluidic and electrical microbumps.

II. Thermal Benefits of the Proposed Platform

To evaluate the cooling performance of the proposed platform, a test vehicle was fabricated and assembled for thermal measurements. Moreover, microfluidic cooled and air cooled silicon interposer based systems were compared via thermal modeling.

A. Thermal Measurements

The experimental setup for the thermal measurements is illustrated in Fig. 3. A silicon die with the micropin-fin heat sink and electrical and fluidic microbumps is flip-chip bonded onto a silicon interposer. The micropin-fin heat sink is capped using a silicon substrate with a platinum heater, which simultaneously serves as a heat source and resistance temperature detector (RTD). Inlet/outlet ports are attached to the back side of the interposer.

De-ionized (DI) water at room temperature (~20 °C) is used as the coolant. During the experiment, the adjustable digital gear pump drew the DI water from a reservoir. The DI water flowed through a mass flow meter and a polyester-based filter to remove particles (>20 μm) that would possibly clog the fluidic vias and micropin-fin heat sink. A differential pressure gauge was used to measure the pressure at the input port. After flowing across the micropin-fin heat sink, the DI water exits the chip into another reservior. The temperature of the DI water was measured at both the inlet and output ports.

Once coolant flow commenced, the thin film Pt heater was powered by an Agilent N6705B power analyzer to mimic the power dissipation of a functional die. The electrical resistance of the heater/RTD was recorded with an Agilent 34970A data logger.

Flow rate is an important factor that affects the cooling performance of the micropin-fin heat sink. Different flow rates, from 10 mL/min to 50 mL/min, were applied during the experiment. A power density of up to 100 W/cm^2 was applied to the heater.

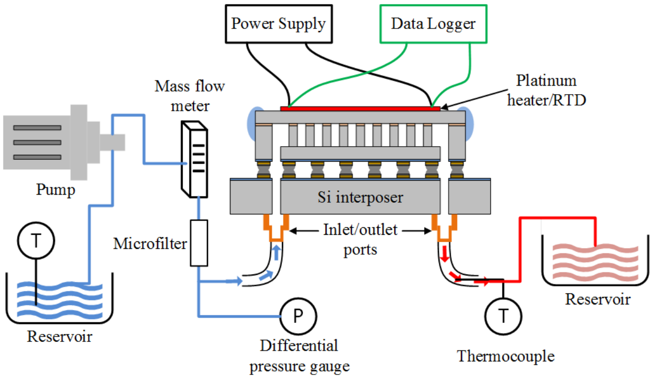


Fig. 3 Microfluidic cooling experiment setup

The results of the experiments are shown in Fig. 4. As expected, the temperature of the heater increases linearly with increased power density, and the temperature decreases as the flow rate increases for a given power density. The measured junction temperature is 55.9 °C at a power density of 97.0 W/cm^2 with a flow rate of 50 mL/min.

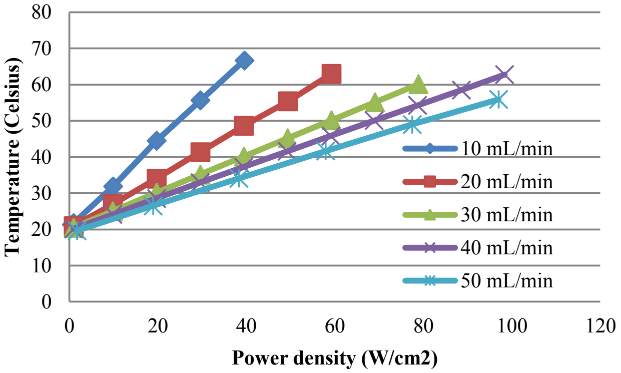


Fig. 4 Heater/RTD temperature vs. power density for different flow rates (DI water at room temperature ~20 °C)

The resistance of the electrical microbumps were also measured using the 4-point measurement method. Table 1 lists the measured resistance of three different assembled samples to validate the bonding quality.

Table 1: Resistance of the electrical microbumps (mΩ)

Microbump	sample #1	sample #2	sample #3
#1	11.6	12.8	13.8
#2	10.6	11.3	11.8
#3	12.9	13.2	12.6
Average	11.7	12.4	12.7

B. *Thermal Simulation*

In this section, thermal modeling based on the finite volumn method [8] is used to compare different air and microfluidic cooled silicon interposer based systems. The thermal models, to which a convective boundary is applied, have been validated using ANSYS with an error of less than 3% [8].

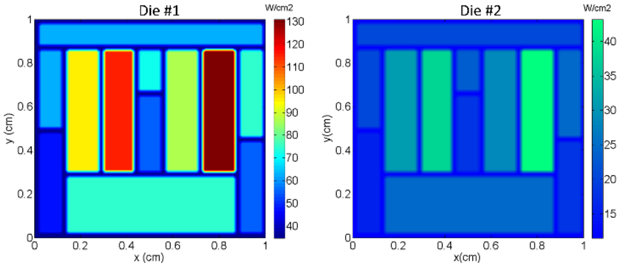


Fig. 5 Power maps of the two dice (die #1 74.63 W and die #2 24.88 W)

For the silicon interposer based system, we assume two silicon dice (1 cm x 1 cm each) are assembled side-by-side on a silicon interposer with a 1 mm gap between the dice. The power maps of the two logic dice are based on the *Intel* i7 microprocessor [9], as shown in Fig. 5. We further assume that the left die (Die #1) is operating at a maximum power of 74.63 W, and the right die (Die #2) is operating at one third of the maximum power. The size of the silicon interposer is 2 cm x 3 cm.

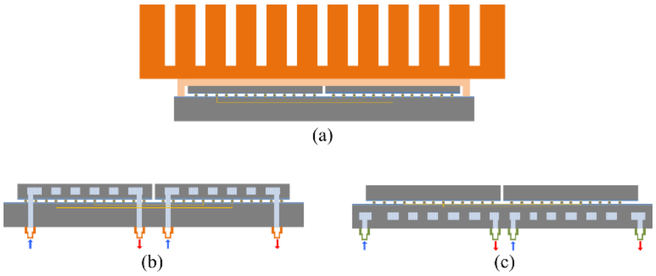


Fig. 6 Different cooling scenarios for a silicon interposer based system assuming two dice: (a) air cooing; (b) microfluidic cooling in silicon dice; (c) microfluidic cooling in silicon interposer.

For the above silicon interposer based system, we consider three different cooling scenarios, as shown in Fig. 6. Fig 6(a) shows an air cooling solution with a heat spreader and a bulky air-cooled heat sink placed on top of the two dice; Fig. 6(b) is the microfluidic cooling scenario in which a microfluidic heat sink is embedded in each of the two dice; in Fig. 6(c), a microfluidic heat sink is embedded in the silicon interposer. This configuration would reduce the system complexity by avoiding microfluidic heat sinks in active dice and fluidic interconnections between the dice and interposer.

The simulated temperature maps of the silicon dice and interposer for the three scenarios (corresponding to scenario (a), (b), and (c) in Fig. 6) are shown in Fig. 7. The maximum temperature of the dice and the average temperature (over the region between the center of the two dice) of the interposer are labeled on top of each temperature map. One obvious effect we see from the temperature maps is that with air cooling, there is a strong temperature coupling between the high-power die and low-power die. As expected, the microfluidic cooling scenarios have much lower temperatures than the air cooling.

Fig. 8 shows the temperatures (maximum temperature of the dice and average temperature of the interposer) of the three scenarios for comparison. Among the three cooling scenarios, as expected, scenario (a) has the highest temperatures. Comparing scenario (b) and (c), the temperature of die #1 is much higher in scenario (c) due to the lack of direct contact to the microfluidic heat sink, while the temperature of die #2 is close in the two scenarios due to its low-power and the thermal isolation with microfluidic cooling. The average temperature of the interposer in the two scenarios is very close to each other.

From the simulations, we can see that, compared to air cooling, microfluidic cooling significantly improves cooling performance and reduces thermal coupling.

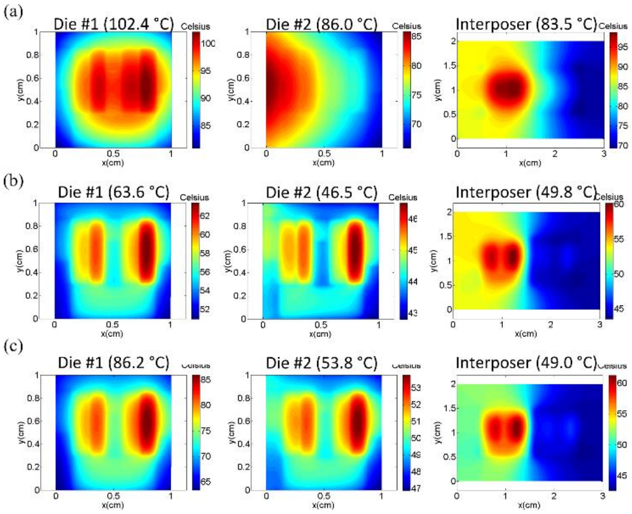


Fig. 7 Simulation temperature maps for the silicon dice and interposer for the three scenarios in Fig. 6: (a) air cooing; (b) microfluidic cooling in silicon dice; (c) microfluidic cooling in silicon interposer.

Based on the thermal simulations, we also found that the number of the electrical microbumps, which determines the thermal conductivity between the silicon dice and interposer, impacts the temperature distribution. Increasing the number of electrical microbumps causes the temperature to become more evenly distributed between the silicon dice and interposer due to the improved thermal conductivity. In the previous simulations, we assume 3,600 electrical microbumps on each die for all three scenarios.

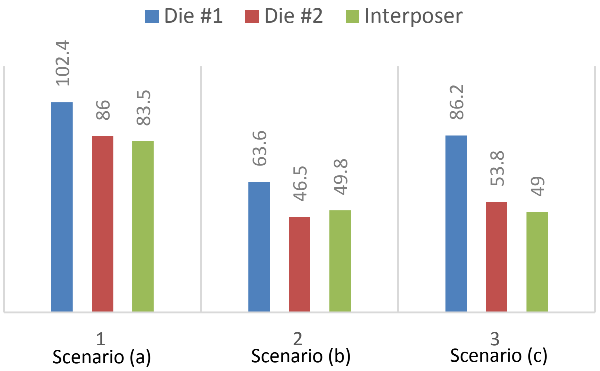


Fig. 8 Comparison of the temperatures (Celsius) of the silicon dice and interposer in the three scenarios: (a) air cooling; (b) microfluidic cooling in silicon dice; (c) microfluidic cooling in silicon interposer.

III. 3-D Stack with Microfluidic Cooling

3-D integration is critical to future high-performance computing systems. In this section, we present the 3-D integrated microfluidic cooled silicon interposer based platform. In this effort, two silicon dice with fine-pitch electrical and fluidic I/Os are stacked on a silicon interposer, as shown in Fig. 9.

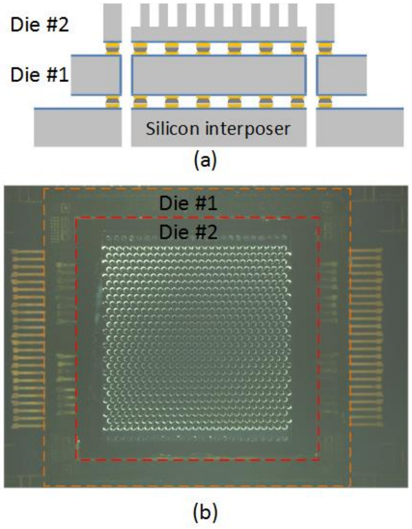


Fig. 9 (a) schematic of the 3-D stack with electrical and fluidic I/Os; (b) optical image of the assembled 3-D stack

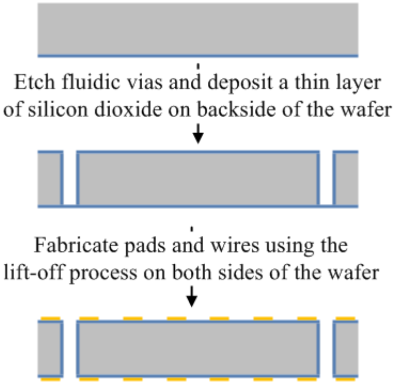

Etch fluidic vias and deposit a thin layer
of silicon dioxide on backside of the wafer
Fabricate pads and wires using the
lift-off process on both sides of the wafer

Fig. 10 Fabrication process of the middle die (die #1)

The fabrication process of Die #1 is illustrated in Fig. 10. The process begins with a 300 μm thick 4-inch wafer with a thin layer of SiO_2 deposited on the back side. Fluidic vias are etched through from the front side using the BOSCH process followed by an SiO_2 deposition step. Next, a thin layer of Ti/Cu/Au is evaporated on both sides of the wafer to form the wires and pads using the lift-off process. The last step is to remove the SiO_2 membrane hanging over the fluidic vias. The fabrication process of Die #2 and the interposer are identical to the process for silicon die and interposer presented in [7], [10]. Table 2 lists the parameters of the Die #1, Die #2, and the silicon interposer.

Table 2: Die and interposer parameters

Parameter	Value
Die #1 size	~ 0.8 cm x 0.8cm
Die #2 size	~ 1 cm x 1cm
Interposer size	~ 1.5 cm x 1.5 cm
Number of fluidic microbumps	48 (24 each row)
Number of electrical microbumps	~7,600

Following fabrication, the two silicon dice are stacked on the silicon interposer in two bonding steps, as shown in Fig. 11. Die #1 is flip-chip bonded on the interposer in the first step. Next, Die #2 is flip-chip bonded on top of Die #1. The temperature and force applied during the bonding process are listed in Table 3.

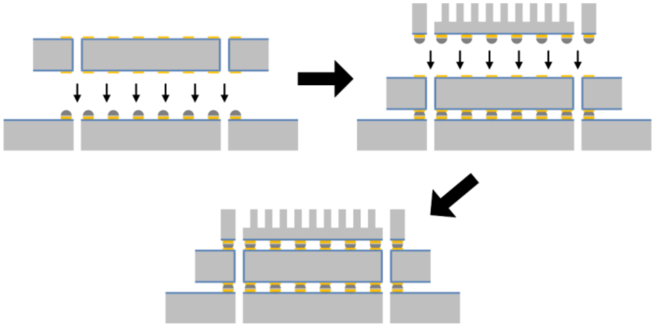


Fig. 11 Flip-chip bonding of two dice on a silicon interposer sequentially

Table 3: Flip-chip bonding and chip parameters

Parameter	Step 1 (Die #1)	Step 2 (Die #2)
Temperature ramp rate	2 ºC/s	2 ºC/s
Peak temperature	230 ºC	230 ºC
Peak temperature time	15 s	15 s
Bonding force	~3.5 N	~4 N

The assembled 3-D stack was inspected using an X-ray imager. Fig. 12 (a) and (b) show the top and angled view of the 3-D stack from which the two silicon dice and interposer can be clearly seen. Fig. 12(c) shows the close-up of a column of fluidic I/Os. The small ‘dots’ and large circles to the left of the fluidic I/Os are the electrical microbumps and micropin-fins, respectively. Fig. 12(d) is the angled close-up in which we observe the fluidic I/Os on both silicon dice.

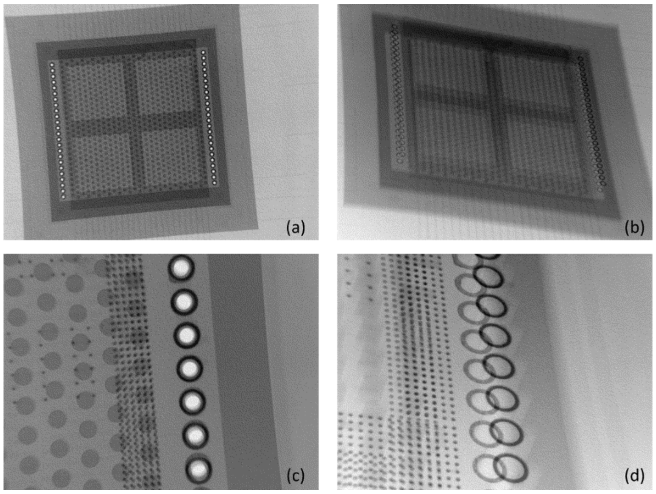


Fig. 12 X-ray images of the 3-D stack with electrical and fluidic I/Os: (a) top view of the stack; (b) angled view of the stack; (c) top view of the fluidic I/Os; (d) angled view of the fluidic I/Os on both tiers.

IV. Conclusions

A platform utilizing silicon interposer and microfluidic cooling technologies is proposed for high-performance computing systems. The cooling performance of the platform is evaluated experimentally. Compared to air cooling, microfluidic cooling significantly reduces temperature and thermal coupling. This paper also reports 3-D stacking of two silicon dice with electrical and fluidic I/Os on a silicon interposer.

Acknowledgments

The authors would like to thank DARPA ICECool Fundamentals and Applications for support of this work under contract numbers W31P4Q-12-1-0014 and HR0011-14-1-0002.

References

[1] J. Knickerbocker, P. Andry, B. Dang, R. Horton, C. Patel, R. Polastre, K. Sakuma, E. Sprogis, C. Tsang, B. Webb, and S. Wright, "3D silicon integration," in *Proc. 58th IEEE Electron. Compon. Technol. Conf.*, Lake Buena Vista, FL, USA, May 2008, pp. 538–543.

[2] R. Ho, P. Amberg, E. Chang, P. Koka, J. Lexau, Guoliang Li, F.Y. Liu, H. Schwetman, I. Shubin, H.D. Thacker, Xuezhe Zheng, J.E. Cunningham, A. V. Krishnamoorthy "Silicon Photonic Interconnects for Large-Scale Computer Systems," *Micro IEEE*, vol.33, no.1, pp.68-78 Jan.-Feb. 2013.

[3] T. O. Dickson, Y. Liu, S. V Rylov, B. Dang, C. K. Tsang, P. S. Andry, J. F. Bulzacchelli, H. A. Ainspan, X. Gu, L. Turlapati, M. P. Beakes, B. D. Parker, J. U. Knickerbocker, and D. J. Friedman, "An 8x10-Gb/s Source-Synchronous I/O System Based on High-Density Silicon Carrier Interconnects," *IEEE J. Solid-State Circuits*, vol. 47, no. 4, pp. 884–896, 2012.

[4] P. Thadesar and M. Bakir, "Novel photo-defined polymer-enhanced through-silicon vias for silicon interposers," *IEEE Trans. Components, Packag. Manuf. Technol.*, vol.3, no.7, pp.1130-1137, July 2013.

[5] D. B. Tuckerman and R. F. W. Pease, "High-performance heat sinking for VLSI," *IEEE Electron Device Lett.*, vol. 2, no. 5, pp. 126–129, May 1981.

[6] Y. Zhang, A. Dembla, Y. Joshi, and M. Bakir, "3D stacked microfluidic cooling for high-performance 3D ICs," in *Proc. 62nd Electron. Components Technol. Conf.*, 2012.

[7] L. Zheng, Y. Zhang, G. Huang, M.S. Bakir, "Novel Electrical and Fluidic Microbumps for Silicon Interposer and 3-D ICs," *IEEE Trans. Components, Packag. Manuf. Technol.*, vol.4, no.5, pp.777-785, May 2014.

[8] Y. Zhang, Y. Zhang, M. S. Bakir, "Thermal design and constraints for heterogeneous integrated chip stacks and isolation technology using air gap and thermal bridge," *IEEE Trans. Compo. Packag. Manuf. Technol.*, vol.4, no.12, pp.1914-1924, December 2014.

[9] Intel, "New 2010 Intel® CoreTMi7 Processor Extreme Edition," http://newsroom.intel.com, [Available Online].

[10] L. Zheng and M. Bakir, "Design, fabrication and assembly of novel electrical and microfluidic I/Os for 3-D chip stack and silicon interposer," in *Proc. 63rd IEEE Electronic Components and Technology Conference (ECTC)*, Las Vegas, NV, USA, May 2013.