Microfluidic Cooling for 3D-IC with 3D Printing Package

Jun-Han Han, Karina Torres-Castro, Robert E. West, Walter Varhue, Nathan Swami, and Mircea Stan
Electrical and Computer Engineering Department
University of Virginia
Charlottesville, VA, USA
jh2vs@virginia.edu, mircea@virginia.edu

***Abstract*—Microchannel cooling method for stacked 3D-IC has process and cost issues. The proposed method, the thermal fluid chamber, exploits the gap between the upper and lower stacks. The method advantages are achieving controllable coolant flow and low-cost process. The microfluidic cooling system is composed of a thermal test chip, thermal test board, 3D printing package, and microfluidic apparatus. The thermal test chip with C4 micro-bumps is attached on the test board. The thermal generator and sensor in the chip are operated by the test board. The microfluidic chamber of 68 µm thick is integrated between the chip and board. Thermal sensing and microfluidic tests demonstrate the validity of the thermal fluidic chamber.**

***Keywords*—microchannel, 3D-IC, cooling, 3D-printing, IC package**

I. INTRODUCTION

With the continued increase of transistor density, three-dimensional integrated circuits (3D-ICs) have become a promising technology for high-performance computing. The main bottleneck of 3D-IC technology is a limited heat removal capacity. In the past, both passive and active methods have been used to address this issue [1]-[3]. The passive methods remove the heat using ventilation, heat sinks, or heat pipes. However, since the heat is being generated throughout the volume of the 3D-IC and removed only along the surface, these methods have limited heat removal capacity. On the other hand, active methods such as microchannel cooling can solve the mismatch between limited heat removal capacity and the increase of the volumetric power density inherent in 3D-ICs [1]-[6].

Microchannel cooling methods require additional fluidic structures to deliver the coolant into the 3D stacks. The main fluidic structures are the microchannels, fluidic via, and fluidic pipe. The fluidic via enables coolant to move into the 3D stacks, and the fluidic pipe connects the upper and lower fluidic vias. The microchannels remove the heat produced by the dies by allowing the fluid delivered by the vias to flow across the dies [5]-[6]. However, these fluidic structures require additional fabrication processes and can conflict with electrical structures. To fabricate microchannels, channels must be patterned and etched, and then a capping layer must be attached to make the channels watertight. For the fabrication process of the fluidic via, another etching process is required because the fluidic via is only several hundred micrometers in size, and thus cannot be processed simultaneously with the Through-Silicon Vias (TSVs), which are only a few micrometers in size. During the fabrication process of the fluidic pipe, it is possible for conflicts to arise between electrical and fluidic structures in 3D stacking. When the upper and lower layers are electrically connected through the TSVs and the micro-bumps, the fluidics are also connected through the fluidic vias and pipes. During the reflow process of micro-bumps, the distance between the layers is determined [7]. Therefore, to prevent an electrical interconnection issue, the fluidic pipes fabrication process cannot affect the reflow process of the microbumps. Additionally, since the fluidic pipes are several hundred micrometers in size, the die area must increase for the fluidic pipes to fit. These additional structures, processes, and the increased area will all increase the fabrication cost of the 3D ICs.

Besides the increased fabrication cost, using microchannels as a cooling method has a limitation caused by a fixed one-directional straight coolant flow [2],[8],[9]. After fabrication, the location and direction of microchannels are fixed. The cooling profile cannot be altered based on the temperature profile. This is especially an issue in some modern technologies, such as Processing-In Memory (PIM) or FPGAs, that can have thermal issues caused by multiple hot spots at changing locations [10],[11].

In this study, we propose a cost-effective microfluidic cooling method for 3D-ICs and validate the method with basic experiments. Instead of fabricating the microchannels, fluidic vias, and pipes on wafers, the proposed method utilizes the cavity that already exists between the dies.

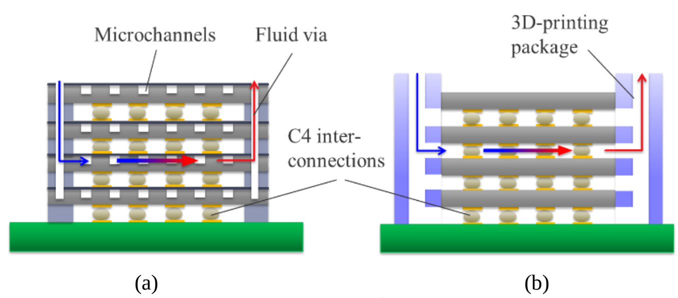
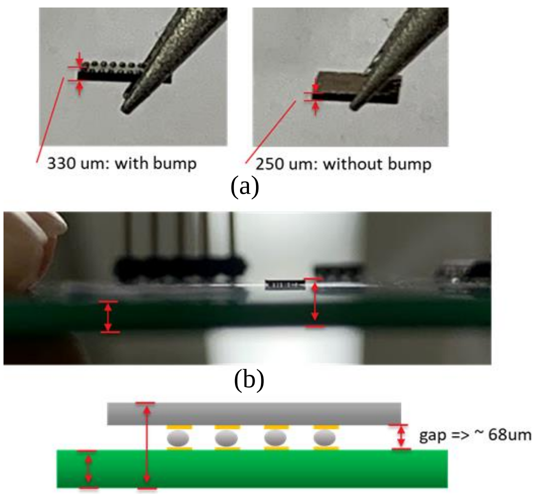


Fig. 1. Cross-section schematic diagrams of (a) Microchannel cooling method and (b) fluidic chamber with 3D-printing package.

II. COOLING METHOD

A device structure of previous microchannel cooling method is described in Figure 1 (a). The conventional approaches require microchannels along with the silicon dice and trenched thermal vias through the silicon dice for circulating fluid. Instead of fabricating additional structures, we proposed a microfluidic cooling method that exploits the gap between the silicon dice. By using the empty space as a microfluidic chamber, the proposed method can achieve multidirectional coolant flow. The device structure in Figure 1 (b) shows the proposed device structure comprised of 3D stacked ICs and a 3D printing package. The method does not require additional micro-structure on the wafers for circulating fluid — alternately, the 3D-printing apparatus constructs inlet, outlet, and channels for the coolant flow of the stacked 3D-IC structure. This approach can decrease the process complexity and the cost by eliminating the additional structures on the silicon wafers.



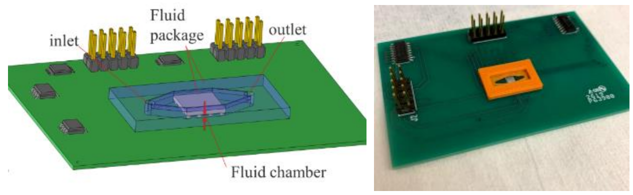
(a)
(b)
(c)

Fig. 2. Images of (a) thermal test chips, (b) microfluidic chamber, and (c) schematic diagram of microfluidic chamber.

III. Device Configuration

We demonstrate the microfluidic chamber method of cooling using a thermal test chip and a printed circuit board. Twenty-four micro bumps with 80 µm height were attached on the silicon chip before assembling the thermo-fluidic chamber. The stacked thermal chips in Figure 2 shows the gap between the chips. The height of the chamber is measured using a digital caliper and a profilometer. The average gap of four samples is 68 µm. Note that due to the reflow process, the height of the microbumps decreased from 80 µm. The thermal test chip has 2540 µm × 2540 µm dimension: two heat generating resistors and four thermal sensing diodes. Two resistors generate the heat flux up to 6W each. Under a constant current driving condition, the forward voltage of diodes varies with the temperature change. The forward voltages of the diodes are read through 12-bit ADC of Texas Instruments AMC 7823. Heat flux generation and temperature sensing operations are controlled by a RaspberryPi system.

To investigate the fluidic-cooling experiment the thermal test chip was directly mounted on the printed circuit board. The gap between the chip and the board constitutes the thermal chamber in this configuration. A Fused Deposition Modeling (FDM) type 3D printer Stratasys F170 is used for printing the apparatus. Figure 3 (a) shows the 3D design result using SolidWorks and Figure 3 (b) shows the image of actual device. The device has an inlet and an outlet for the circulating the coolant. It also embeds a distribution structure for the coolant and acts as a cue for the fluid chamber structure.



(a) (b)

Fig. 3. (a) Microfluidic chamber design and (b) actual device using 3D printing package.

A dielectric coolant of EC120 ElectroCool® was used as the coolant for the fluid chamber. The dielectric coolant has a high resistivity of $1x10^{14}$ Ω/cm and dielectric constants of 2.1. Dielectric coolant is required to ensure proper electric insulation of the microbumps that act as pillars between stacks. The dielectric coolant has a kinematic viscosity of 5.02 cSt which is higher than the 0.658 cSt of deionized water at 40°C

A constant current of 100 µA drives the thermal sensing diode, and a heat flux of 0.5W is applied to the resistor. The forward voltage of the diodes decreases when the resistor generates heat. The forward voltages of the diodes are sensed using the 12-bit ADC. The change of temperature can be represented by the difference in the forward voltage with a correlation constant K. The equation is written as:

$$K = \left| \frac{T_{high} - T_{low}}{V_{low} - V_{high}} \right| \tag{1}$$

The unit of constant K is °C/mV, and the calibrated value in the experimental set-up is 0.2 °C/mV.

IV. Results and Discussion

Figure 4 shows the converted temperature obtained as the product of forward voltage and constant K. After 90 seconds of heat flux generation period, the flux generator is turned off. The black and red curves are the baselines to show the temperature behavior with and without heat flux generation.

10 µl of the dielectric coolant at 10ºC is injected through the fluid chamber to wet the surface and facilitate the fluid flow. A micropipette tip was used as coolant reservoir at the inlet, and a passive pumping mechanism was placed at the outlet. The passive pumping mechanism we used consisted of a paper pump that acts upon the capillary effect and the porosity of the paper to withdraw coolant from the fluid chamber.

When the chilled water flows for 60 sec, the temperature drops 10 °C during 3 seconds and re-heated by the heat flux generator. The blue curve of figure 4 shows the cooling effect of the coolant flow. This change of temperature due to the coolant flow confirms that the pumping effect of the coolant through the microcavity effectively cools down the thermal test chip.

The green curve of figure 4 shows the microfluidic cooling behavior of the fluidic chamber method. During the 90 seconds of heat flux generation period, coolant flow is applied at 60 seconds. The paper pump is used at an outlet, and 1ml pipette tip is attached to the inlet. The 1ml of coolant gradually runs down by gravity until the pressure between inlet and outlet is balanced and then the paper pump is placed at the outlet with a primer coolant drop to enable paper pumping by capillary effect. During 14 seconds of the cooling period, despite the 0.5W heat flux is generating, the temperature gradually decreased by 14 °C and start to re-heated by the heat flux generator after the end of coolant flow. These preliminary results show that our thermo-fluidic chamber cooling method is feasible for a 3D stack of IC.

This study demonstrates that the microfluidic chamber method with 3D printing package is a valid alternative to the micro-channel method. This method can achieve controllable coolant flow with a low-cost fabrication process. Structural and material improvements are required to achieve watertight 3D-printing package for microfluidic cooling of 3D-IC and other paper pump geometries will be explored to achieve longer pumping times.

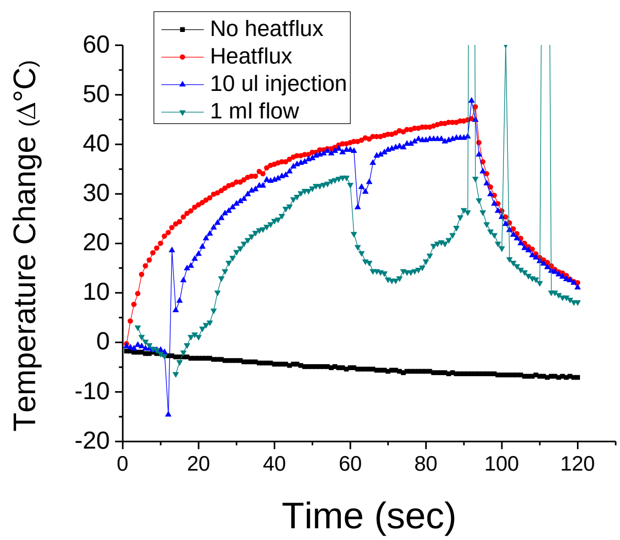


Fig. 4. Converted temperature change from the forward voltage of thermal sensing diode of the thermal test chip

ACKNOWLEDGMENT

This work was supported in part by Semiconductor Research Corporation (SRC) Joint University Microelectronics Program (JUMP) Task 2780.003: Thermal and Power Management. Engineered Fluid Inc. sponsored the dielectric coolant EC120 ElectroCool®.

REFERENCES

[1] P. Ruch, T. Brunschwiler, W. Escher, S. Paredes, and B. Michel, “Toward five-dimensional scaling: How density improves efficiency in future computers,” IBM J. Res. Dev., vol. 55, no. 5, p. 15:1-15:13, 2012.

[2] A. Sridhar, A. Vincenzi, D. Atienza, and T. Brunschwiler, “3D-ICE: A compact thermal model for early-stage design of liquid-cooled ICs,” IEEE Trans. Comput., vol. 63, no. 10, pp. 2576–2589, 2014.

[3] Wang, S. et al. 3D Integrated Circuit Cooling with Microfluidics Micromachines. 2018 Jun 7;9(6).

[4] A. Bar-Cohen, J. J. Maurer, and J. G. Felbinger, “DARPA’s Intra/Interchip Enhanced Cooling (ICECool) Program,” CS MANTECH Conf., pp. 171–174, 2013.

[5] D. Sekar et al., “A 3D-IC technology with integrated microchannel cooling,” 2008 IEEE Int. Interconnect Technol. Conf. IITC, pp. 13–15, 2008.

[6] B. Dang, M. S. Bakir, D. C. Sekar, C. R. King, and J. D. Meindl, “Integrated microfluidic cooling and interconnects for 2D and 3D chips,” IEEE Trans. Adv. Packag., vol. 33, no. 1, pp. 79–87, 2010.

[7] B. Dang et al., “3D chip stacking with C4 technology,” IBM J. Res. Dev., 2008.

[8] G. Chen, J. Kuang, Z. Zeng, H. Zhang, E. F. Y. Young, and B. Yu, “Minimizing Thermal Gradient and Pumping Power in 3D IC Liquid Cooling Network Design,” no. 1, pp. 1–6, 2017.

[9] H. Zhu, F. Hu, H. Zhou, D. Z. Pan, D. Zhou, and X. Zeng, “Interlayer Cooling Network Design for High-Performance 3D ICs Using Channel Patterning and Pruning,” IEEE Trans. Comput. Des. Integr. Circuits Syst., vol. 37, no. 4, pp. 770–781, Apr. 2018.

[10] L. Nai, R. Hadidi, H. Xiao, H. Kim, J. Sim, and H. Kim, “CoolPIM : Thermal-Aware Source Throttling for Efficient PIM Instruction Offloading,” 2018 IEEE Int. Parallel Distrib. Process. Symp., pp. 680–689, 2018.

[11] G. Deshpande, D. K. Bhatia, “Microchannels for thermal management in FPGAs”, Int. Conf. on ReConFigurable Comp. and FPGAs (ReConFig), Dec 4-6, 2017, pp.1-5