2012 12th IEEE International Conference on Nanotechnology (IEEE-NANO)
The International Conference Centre Birmingham
20-23 August 20112, Birmingham, United Kingdom

High Aspect Ratio TSVs in Micropin-Fin Heat Sinks for 3D ICs

Ashish Dembla, Yue Zhang and Muhannad S. Bakir

***Abstract*— Future high performance 3D systems require a systematic co-design of their electrical interconnect network and their heat removal mechanism. This paper presents fine pitch (35 µm) and high aspect ratio (20:1) TSVs integrated in a benchmarked micropin-fin heat sink capable of removing power density of 100 W/cm^2/tier at a junction temperature below 50 ºC.**

I. Introduction

In the last decade, 3D IC technology has emerged as a means of extending Moore's Law. Stacking dice vertically reduces the physical distance between dice that leads to reduction in interconnect wire lengths [1]. Three-dimensional stacking reduces the interconnect lengths that enables low energy and low latency communication. For example, by stacking multicore processors with DRAM, the entire off-chip interconnect network can be reduced from several millimeters to few hundred microns or less. High performance systems require integration of state-of-the-art technologies into a single 3D system. State-of-the art analog mixed signal systems, digital logic, memory and MEMS can be fabricated using dedicated processes on independent wafers. These dice can then be integrated into a single multifunctional system using 3D IC technology. Next generation computing at exa-scale or peta-scale requires design of high performance and energy efficient computing systems that utilize 3D integration effectively.

Requirements of such high performance 3D systems presented in the ITRS assembly and packaging roadmap are plotted in Figure 1 [2]. The projection data shows that a drop in power dissipation for chip stacks is required to address the limited heat removal capability of conventional air-cooled heat sinks and to operate at lower junction temperatures. Also, the number of dice in a stack is expected to increase leading to a proportional increase in power density of high performance 3D stacks. Such 3D stacks may have power densities in excess of 100-200W/cm^2. Conventional air-cooling is limited to cooling the stack from one particular direction, generally from the top of the package. One scalable approach to enhanced cooling in 3D stacks is through embedding a microfluidic heat sink technology within the die stacks [3]. The embedded microfluidic heat sink provides targeted cooling to internal tiers making the approach scalable. In comparison to conventional air-cooling in 3D systems, liquid cooled 3D systems show better thermal performance [4,5].

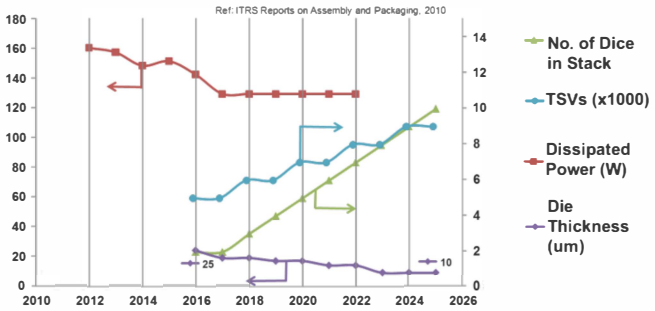

Figure 1: ITRS data for high performance 2D/3D stacked ICs [2]

Integration of high performance 3D systems faces several challenges. These can broadly be classified into manufacturing challenges and design challenges. The challenges from a manufacturing perspective deal with issues such as reliability, warpage, and electromigration, while the design challenges are efficient power delivery (i.e. reduced power supply noise), high bandwidth inter-tier signaling, low energy inter-tier communication, and, perhaps the most important, efficient heat removal. There have been several approaches to address these issues independently as shown in [6],[7]&[8] respectively. The design challenges show stronger interdependence for the high performance segment of 3D ICs, requiring multidimensional analysis and co-design.

The work presented here aims to address the described design challenges in a unified manner while laying emphasis on the simultaneous improvement in thermal and electrical interconnect performance of the high performance 3D stacks.

II. Heat Removal in 3D ICs: Challenges and Opportunities

The number of stacked high performance dice determines the aggregate power density of a 3D stacked IC. Tuckerman and Pease demonstrated heat removal capability up to 790W/cm^2 using embedded microchannel heat sinks in a single chip [9]. Pressure drop and thermal resistance (R_{tot}) are metrics used to quantitatively evaluate heat sink performance. The thermal resistance comprises of three

Manuscript received July 4, 2012. This work has been carried out in part under Semiconductor Research Corporation contract 2011-KJ-2189 and Interconnect Focus Center contract 2009-IT-2050.

Ashish Dembla is with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332 USA Phone: 678-702-2065; e-mail: adembla@ gatech.edu

Yue Zhang is with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332 USA; e-mail: yzhang324@ gatech.edu

Muhannad S. Bakir is an Associate Professor with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332; e-mail: muhannad.bakir@mirc.gatech.edu

parts: R_{cond}, arising due to the conductance from the circuit through the wafer and heat sink interface; R_{conv} accounts for the convection from the heat sink to the liquid; R_{heat} is due to the increase of the fluidic temperature as it flows through the heat sink [10]. The geometric design of an interlayer microfluidic heat sink can significantly affect the convective heat transfer. In theory, the convective component of thermal resistance is inversely proportional to the interfacial contact area between the coolant and the silicon heat sink. Therefore, to increase the heat removal capability of single phase microfluidic systems, the contact surface area between the heat sink and the flowing coolant needs to increase. This increase in surface area can be accomplished through insertion of high surface area cylindrical pillar like obstructions (micropin-fins) in the direction of liquid flow. The obstructions (micropin-fins as shown in figure 2) can significantly increase the interfacial contact area in microfluidic heat sinks while keeping the height of the heat sink unaltered. Thus, micropin-fin heat sinks are able to provide large background heat removal in excess of 100 W/cm^2/tier. The applicability of large background heat removal is not limited to high-performance systems only, but such large background heat removal technologies can be used to provide large localized cooling through other technologies that are capable of spreading heat from hot spots [12,13]. Initial co-design of the heat sink has been shown by the co-authors [4].

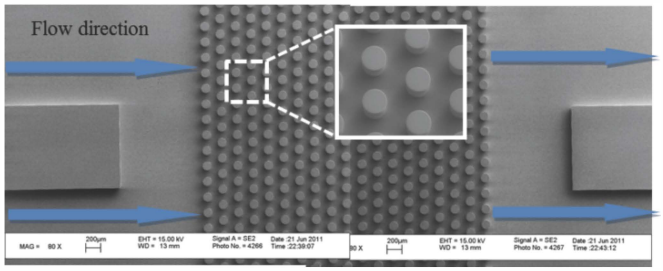

Figure 2: Staggered micropin-fin heat sink structure (top view image) showing liquid flow direction

Simulations performed using analytical models presented in [10-11] show that there is a good design spot from both an electrical and thermal perspective wherein, the diameter, height and pitch of micropin-fins is 150 µm, 200 µm and 225 µm, respectively. Table 1 shows the measured total ther-mal resistance for air cooled heat sink and micropin-fin heat sink at 100 W/cm^2. Benchmarking of air-cooled and micropin-fin heat sinks is accomplished by measuring the junction temperatures at different power densities. Figure 3 shows the benchmarked data compared to ITRS projected data for junction temperature requirements up to the end of the roadmap. While advances in air-cooled heat sink can be made, microfluidic heat sinks provide superior performance even at high power densities. The measured data shows that junction temperatures as low as ~50 ºC can be achieved at 100 W/cm^2 using single phase staggered micropin-fin heat sink.

Table 1: Comparison of air cooled heat sink (ACHS) and micro-fluidic heat sink (MFHS)

Property	ACHS	MFHS (70 mL/min)
$R_{thermal}$ (measured) K cm^2/W	0.518	0.269
$R_{thermal}$ (model) K cm^2/W	-	0.274
Power Density (W/cm^2)	100.2	103.4

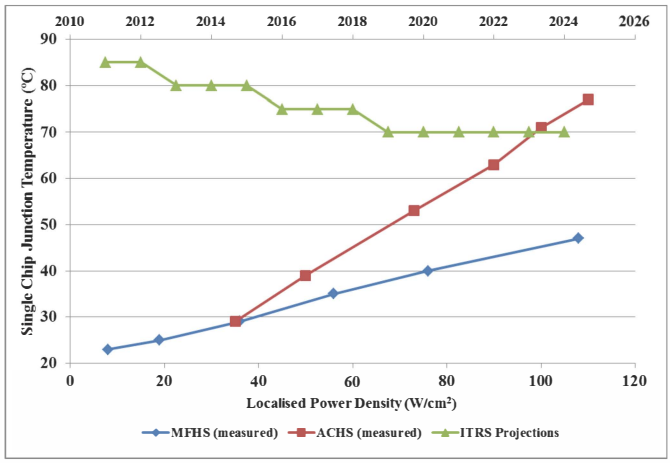

Figure 3: Benchmarking air-cooled heat sinks (ACHS) and microfluidic heat sinks (MFHS) at flow rate of 70mL/min. Top axis represents ITRS projections for single chip junction temperature

Microfluidic heat sinks provide multiple low temperature junctions (coolant) that are embedded in the 3D stack. Schematic of the thermal resistance network for a two-die stack is shown in Figure 6. The presence of embedded low temperature junctions allows increased heat transfer across low thermal resistance paths to the flowing coolant. The thermal test-bed shown in Figure 4 was fabricated to study the impact of microfluidic interlayer cooling in three different scenarios of a two-die stack: (a) processor on processor (b) memory on processor and (c) processor on memory. The thermal test-bed comprises of gear pumps to deliver coolant to each micropin-fin heat sink in the stack, flow meter to measure the flow rate of the coolant, differential pressure gauge to measure the pressure drop across each tier, and thermocouples to measure the inlet and outlet temperatures. An Agilent power analyzer is used to deliver power density up to 100W/cm^2/tier to an on-chip platinum RTD, and an Agilent data logger is used to measure the resistance of the RTD and monitor the delivered power at a sampling rate of 1 Hz. To emulate a high-performance processor power dissipation, >50 W/cm^2 is delivered to the platinum RTD. On the other hand, to emulate memory chip power dissipation, ~5 W/cm^2 is delivered to the platinum RTD. Furthermore, interlayer cooling was investigated for identical flow rates in different tiers and power adaptive flow rates in two tiers. A summary of the results obtained is shown in Table 2.

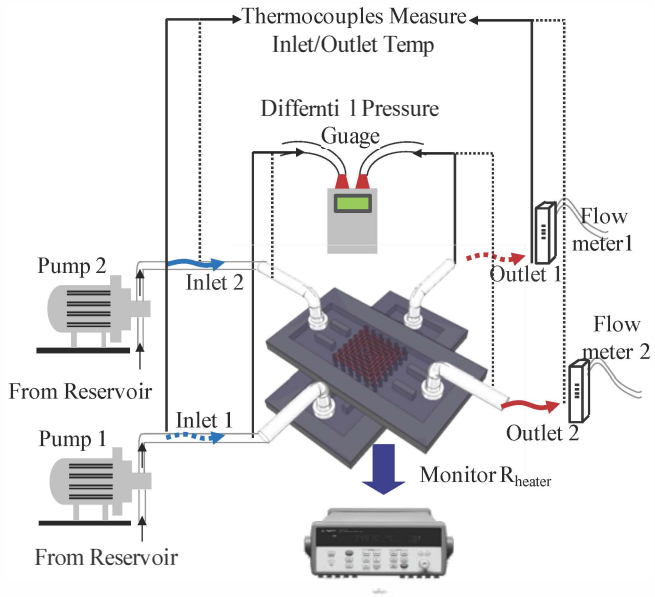

Figure 4: Schematic diagram of test-bed used to measure thermal performance of a microfluidic cooled 3D stack in three configurations (a) processor on processor (b) processor on memory (c) memory on processor

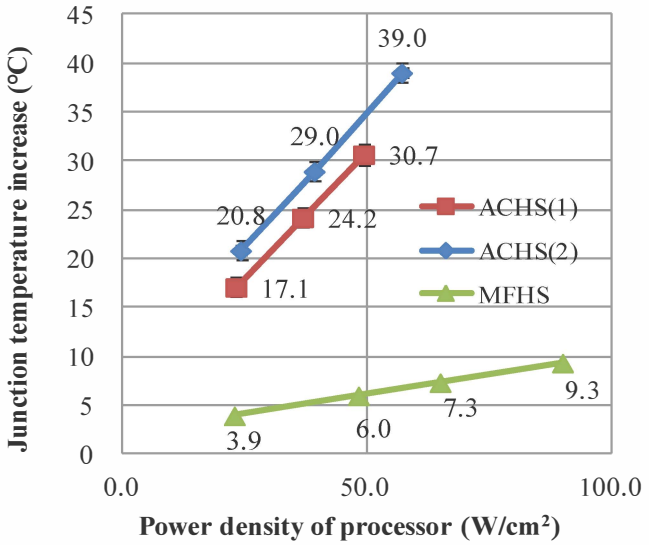

Figure 5: Increase of memory chip junction temperature as a function of different 3D stack configration. ACHS (1) is the architecture where memory is closer to the fan. ACHS (2) is the case when the processor is closer to the fan. For the case of the MFHS, only the processor is liquid cooled.

Figure 5 shows the increase in junction temperature from ambient in the memory-processor test cases. The air-cooled cases include memory on processor and processor on memory; microfluidic cooling is used to cool the processor only. In the case of air-cooling, proximity of the processor to the air-cooled heat sink reduces the junction temperature by 4-9 °C, depending on the power dissipated in the stacked dice. In the same scenario the embedded micropin-fin heat sink outperforms the air-cooled heat sink by reducing the junction temperature by 21 °C compared to the air-cooled heat sink at peak power density of 100 W/cm^2, which is approximately twice the peak power density in the air-cooled heat sink. Reduction in junction temperatures helps reduce the leakage power and improve device reliability by mitigating thermal degradation mechanisms.

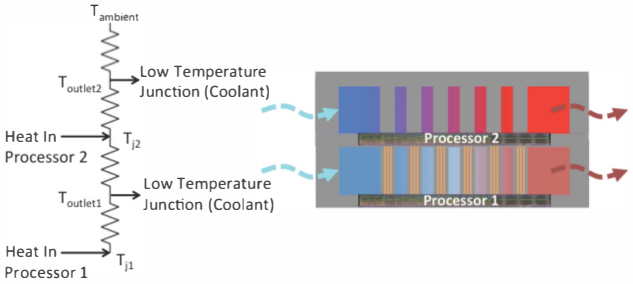

Figure 6: Thermal resistance network of a two-die 3D stack showing embedded low temperature junctions that provide low resistance heat flow paths for enhanced heat removal.

Table 2: Summary of some representative data points for microfluidic cooling (white) and air-cooling (blue)

	Power density (W/cm^2)		Flow rate (ml/min)		Junction T (°C)	
	B	T	B	T	B	T
Two Processors	54.8	56.0	60	60	36.4	34.4
	104.5	106.1	60	60	47.6	47.0
	45.6	48.4	-	-	95.9	76.7
Memory + Processor	5	48.3	0	60	25.3	31.2
	5	90.0	0	60	28.6	41.4
	5	57.1	-	-	61.2	59.0
	49.3	5	-	-	79.3	50.7
Identical flow rate	102	55	45	45	50.5	38.5
Different flow rate	105	55	70	40	45.5	38.5

*B=Bottom Chip, T=Top Chip

III. Electrical TSVs in Microfluidic Heat Sinks

A schematic representation of the high-performance 3D stack envisioned in this work is shown in Figure 7. The system consists of multiple 3D stacks on a CTE matched silicon substrate (silicon interposer). In the schematic, an array of 3D stacks each containing two processors and multiple memory stacks is shown.

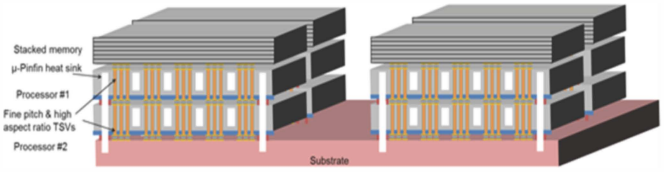

Figure 7: Schematic of micro-pinfin heat sink with integrated fine pitch and high aspect ratio TSVs.

The stacks can communicate on the interposer through high density planar interconnect networks fabricated by conventional BEOL processing. More detailed work that describes 2D interconnect design can be found in [14-15]. Additionally technologies that may be required for hermetic

sealing of fluidic interconnect networks may be found in [16]. The 3D system depicted here has high bandwidth and low energy-per-bit interconnect network between adjacent tiers of a 3D stack along with an embedded microfluidic heat sink capable of removing power densities up to 100 W/cm^2/tier. In this paper, critical design aspects of TSVs that are required to connect 3D stack with interlayer cooling are described.

Electrical interconnect networks need to be integrated within the described micropin-fin based heat sink. The integration of these technologies challenges the electrical interconnect design in two ways. Firstly, the length of the vertical interconnects (TSVs) is proportional to the height of the heat sink. Figure 8, shows that an increase in micropin-fin height leads to reduction in thermal resistance and a linearly proportional increase in TSV capacitance. The increase in TSV capacitance follows a linear trend as long as TSV aspect ratio can keep increasing i.e. TSV is taller but has the same diameter. For taller microfluidic heat sinks, the aspect ratio may not keep up and the increase in capacitance becomes quadratic instead of linear. The second challenge in integrating TSVs in tall microfluidic heat sinks arises as a result of increasing TSV diameter due to the aspect ratio limited fabrication leading to a quadratic decrease in TSV density.

Compact physical models for TSV resistance and capacitance were adapted from Katti et al. [17]. The analytical correlation based models for thermal resistance were adapted from Peyes et al. [10]. Using the analytical models, a trade-off between thermal resistance, TSV capacitance and TSV density was assessed as shown in Figure 9. TSVs are placed periodically in 1% of silicon area in a 200 μm thick micropin-fin heat sink having an area of 1 cm^2. The typical oxide thickness is 500 nm. TSV height is the sum of the micropin-fin height and the processor die thickness (the die thickness is depicted as the grey zone in Figure 4, i.e., the base of the die).

Given that microfluidic cooling enables the stacking of high power chips, delivering power up to 100 W per tier is as important and challenging as the cooling. TSV resource allocation to signal connectivity and power delivery should be done in a way to minimize silicon area consumption. TSVs with aspect ratios (~10:1) are used in general for 3D connectivity in low power 3D stacks [14]. Using the same TSVs to connect dice with interlayer cooling, at 10:1 aspect ratio, approximately 3,200 TSVs can be integrated between two adjacent tiers of the 3D stack (requiring 1mm^2 of die area). TSVs are required to simultaneously address signaling and power-delivery requirements of the 3D stack. The limited availability of vertical interconnects constrains the number of high-performance dice that can be interconnected within a 3D stack. TSV density can be increased by dedicating more area to vertical interconnect routing. Scaling aspect ratio provides a way to increase the TSV density without sacrificing precious silicon real estate (although resistance of the TSVs dedicated to power delivery must be considered). Additionally, increasing the aspect ratio at constant TSV height helps reduce the capacitance from 1.6 pF at 5:1 AR to 0.4 pF at 20:1 AR (Figure 9).

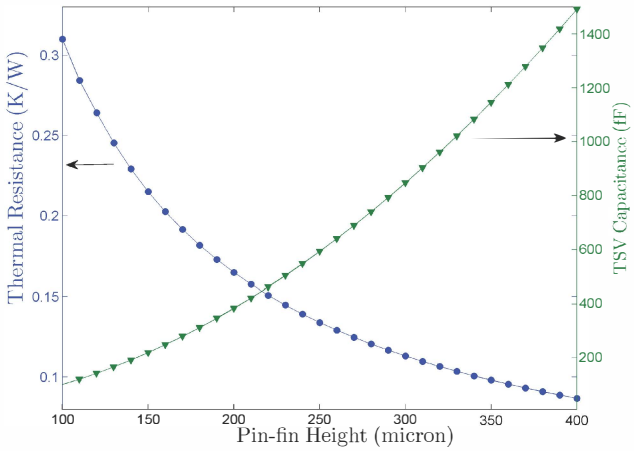

Figure 8: Impact of pin-fin height (at fixed AR=20) on electrical and thermal performance of a micropin-fin heat sink

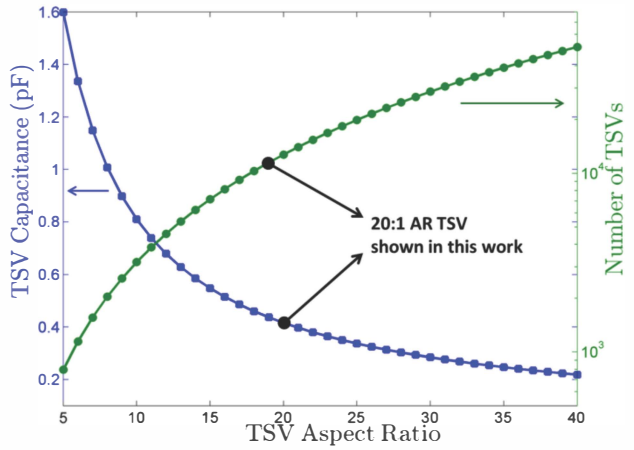

Figure 9: Impact of TSV aspect ratio on capacitance and TSV density in 200 µm tall micropin-fins with 1µm SiO_2 sidewall liner. Number of TSVs is calculated in 1% of silicon area of 1cm^2 die.

By increasing the TSV aspect ratio, smaller diameter TSVs can be fabricated that have relatively smaller capacitance. Therefore, the height reduction from staggered micropin-fin heat sinks [4] combined with high aspect ratio TSVs can provide a lower-energy interconnect solution in high power and high-performance systems. High aspect ratio TSVs in micropin-fins can provide higher TSV densities and a corresponding reduction in their capacitance; further reduction in capacitance can be achieved through other means e.g. low-k dielectric liners.

The fabrication of such high aspect ratio (> 15:1) structures is challenging and requires novel TSV etch processes and thorough TSV 3D metrology. The details of fabrication are provided in the next section.

IV. Fabrication & Measurement

The fabrication of TSVs within micropin-fins starts with a 300µm thick double side polished silicon (DSP) wafer. High aspect ratio TSVs (~20:1) are fabricated using a standard

Bosch process that alternates between SF_6 (plasma etch step) and inert C_4F_8 (deposition step). To achieve the high aspect ratio etching the gas flow rate, cycle times and RF power is ramped over the process duration. Cross-section image of a 23:1 aspect ratio via etched to a depth of 284 μm is shown in Figure 11 (b). Thermal oxide is then grown to isolate the TSVs from the substrate. Bottom-up pulsed electroplating with Enthone DVF plating solution is used to fill the vias with copper. Cross-section image of electroplated TSV is shown in Figure 11 (a).

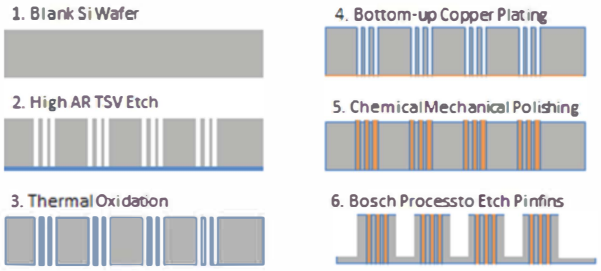

Figure 10: Fabrication process flow for TSVs in micropin-fins.

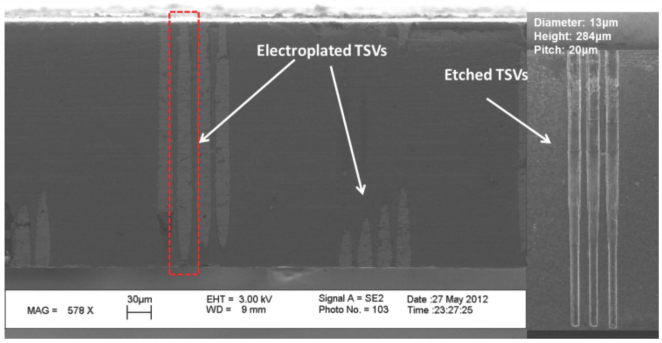

Figure 11:SEM Images of fully plated TSVs, manual polishing creates conical shapes of electroplated cylinders(left) & SEM image (right) of high aspect ratio TSVs etched using Bosch process (10 μm diameter, 35 μm pitch and 178 μm tall).

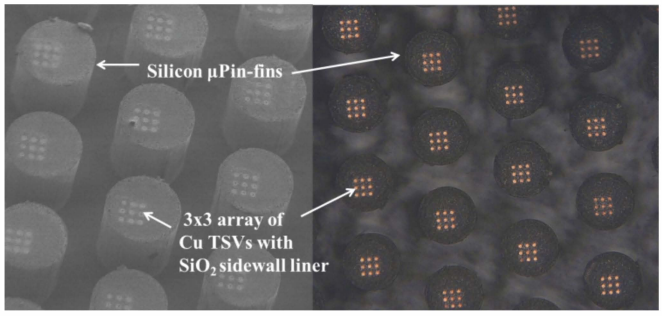

Figure 12:SEM (left) & optical images (right) of high aspect ratio TSVs integrated in micropin-fins (10 μm diameter, 35 μm pitch and 178 μm tall).

Following electroplating, the sample is polished and excess silicon is removed using the Bosch process to make 200μm tall micropin-fins. SEM and optical images of a micropin-fin with integrated TSVs is shown in Figure 12. The fabricated TSVs are ~10 μm in diameter and 200 μm deep (20:1). An overview of the process flow for fabricating TSVs in micropin-fin heat sinks is shown is Figure 11. The TSVs are fully plated with no voids. To test the presence of voids, the sample was dipped in 45% KOH to remove any surrounding silicon; free standing high aspect ratio copper plated TSVs were observed. Figure 13 shows a more aggressively fabricated structure that consists of a 4x4 array of TSVs within the micropin-fins.

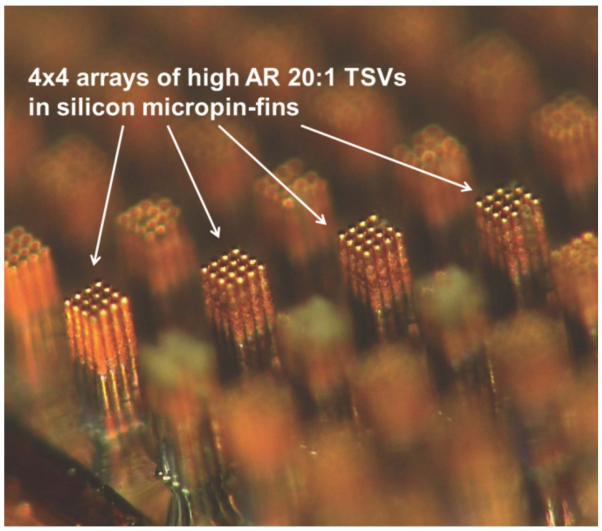

Figure 13: Free standing copper plated TSVs within the silicon micropin-fin heat sink (micropin-fins removed using etching).

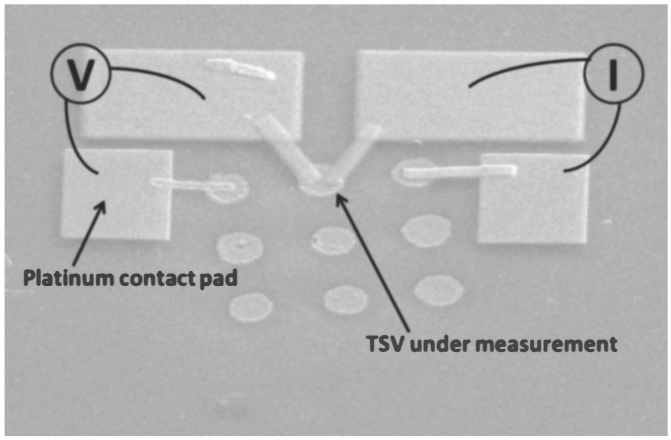

Figure 14: Resistance measurement of TSVs using four point tech-nique. The platinum pads were deposited selectively using Focused Ion Beam (FIB) deposition.

This shows the absence of voids in the electroplated TSVs. The resistance of these TSVs was measured using four point technique; platinum pads were deposited selectively using Focused Ion Beam (FIB) deposition (Figure 14). The theoretical value of TSV resistance for a 178 μm tall TSV is 38 mΩ, this is within the error bar of the measured value of 36.5 ± 1.5 mΩ. In summary, a die spanning 10mm x 10mm can have 1,936 micropin-fins with diameter of 150μm at a pitch of 225μm. Each micropin-fin has 9 electrical TSVs, providing a total of 17,424 electrical I/Os can be used to connect adjacent layers of the 3D stack while using only 1.36% of the die area.

V. Conclusion

This work demonstrates key enabling technologies for energy efficient high-performance 3D systems. Fine pitch TSVs were integrated within micropin-fins and their

resistance was measured to be with reasonable error limits. These TSVs were integrated in a silicon micropin-fin heat sink that has been experimentally shown to remove 100 W/cm^2 with junction temperatures below 50 ºC. Electrical performance metrics (e.g. bandwidth density and energy per bit) can be traded for thermal performance metrics (e.g. thermal resistance) by altering the micropin-fin design. This shows that the electrical and thermal design of next generation high-performance 3D systems is inseparable and needs to be considered in a holistic manner to sustain performance gains in next-generation silicon technologies.

ACKNOWLEDGMENT

We would like to specially thank Dr. Thomas Brunschwiler (IBM Zurich) and Dr. Ramakanth Alapati (GLOBA-LFOUNDIRIES) for their invigorating discussions.

REFERENCES

[1] J. W. Joyner, P. Zarkesh-Ha, J. A. Davis, and J. D. Meindl, "A Three-Dimensional Stochastic Wire-Length Distribution for Variable Separationof Strata," in Proc. IEEE Int. Interconnect Technology Conference, June 2000, pp. 126–128.

[2] ITRS Reports, "Assembly and Packaging", 2010.

[3] F. Schindler-Saefkow. O. Wittler, D. May, B. Michel Thermal Management in a 3D-PCB-Package with Water Cooling, *in Proc. Electronic Systems Integration Technology Conference*, pp.107-110, 2006.

[4] Yue Zhang, C. R. King, J. Zaveri, Yoon Jo Kim, V. Sahu, Y. Joshi, and M. S. Bakir, "Coupled electrical and thermal 3D IC centric microfluidic heat sink design and technology," in Electronic Components and Technology Conference (ECTC), 2011 IEEE 61st, 2011, pp. 2037–2044.

[5] T. Brunschwiler, B. Michel, H. Rothuizen, U. Kloter, B. Wunderle, H. Oppermann, and H. Reichl, "Forced convective interlayer cooling in vertically integrated packages," in Thermal and Thermomechanical Phenomena in Electronic Systems, 2008. ITHERM 2008. 11th Intersociety Conference on, 2008, pp. 1114 –1125. H. Yu, J. Ho, and L. He, "Allocating power ground vias in 3D ICs for simultaneous power and thermal integrity," ACM Trans. Des. Autom. Electron. Syst., vol. 14, no. 3, pp. 41:1–41:31, Jun. 2009.

[6] G. Huang, M. S. Bakir, A. Naeemi, and J. D. Meindl, "Power Delivery for 3-D Chip Stacks: Physical Modeling and Design Implication," Components, Packaging and Manufacturing Technology, IEEE Transactions on, vol. PP, no. 99, p. 1, 2012.

[7] D. H. Woo, N. H. Seong, D. L. Lewis, and H.-H. S. Lee, "An optimized 3D-stacked memory architecture by exploiting excessive, high-density TSV bandwidth," in High Performance Computer Architecture (HPCA), 2010 IEEE 16th International Symposium on, 2010, pp. 1 –12.

[8] L. Zhang, H. Y. Li, S. Gao, and C. S. Tan, "Achieving Stable Through-Silicon Via (TSV) Capacitance with Oxide Fixed Charge," IEEE Electron Device Letters, vol. 32, no. 5, pp. 668–670, May 2011.

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

[10] Y. Peles, A. Koşar, C. Mishra, C.-J. Kuo, and B. Schneider, "Forced convective heat transfer across a pin fin micro heat sink," International Journal of Heat and Mass Transfer, vol. 48, no. 17, pp. 3615–3627, Aug. 2005.

[11] G. Katti, M. Stucchi, K. De Meyer, and W. Dehaene, "Electrical Modeling and Characterization of Through Silicon via for Three-Dimensional ICs," IEEE Transactions on Electron Devices, vol. 57, no. 1, pp. 256–262, Jan. 2010

[12] C. Green et al., *ASME J. Electronic Packaging*, 2009.

[13] V. Sahu et al., *Nano. Micro. Thermophys. Eng.*, 2009.

[14] B. Banijamali, S. Ramalingam, K. Nagarajan, and R. Chaware, "Advanced reliability study of TSV interposers and interconnects for the 28nm technology FPGA," in Electronic Components and Technology Conference (ECTC), 2011 IEEE 61st, 2011, pp. 285 –290.

[15] 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 8x 10-Gb/s Source-Synchronous I/O System Based on High-Density Silicon Carrier Interconnects," Solid-State Circuits, IEEE Journal of, vol. 47, no. 4, pp. 884 –896, Apr. 2012.

[16] C. R. King, J. Zaveri, M. S. Bakir, and J. D. Meindl, "Electrical and fluidic C4 interconnections for inter-layer liquid cooling of 3D ICs," in Electronic Components and Technology Conference (ECTC), 2010 Proceedings 60th, 2010, pp. 1674–1681.

[17] G. Katti, M. Stucchi, K. De Meyer, and W. Dehaene, "Electrical Modeling and Characterization of Through Silicon via for Three-Dimensional ICs," IEEE Transactions on Electron Devices, vol. 57, no. 1, pp. 256–262, Jan. 2010.

978-1-4673-2200-3/12/$31.00 ©2012 IEEE