# Thermal Characterization of TSV based 3D Stacked ICs

Sahana Swarup Sheldon X. –D. Tan Zao Liu
Department of Electrical Engineering, University of California, Riverside
Email: stan@ee.ucr.edu

**Abstract – This paper studies the thermal impact and characterization of Through Silicon Vias (TSVs) in stacked three dimensional (3D) integrated circuits (ICs) through finite-element based numerical analysis. Realistic 3D stacked ICs are built using a commercial finite-element based modeling and analysis tool, COMSOL. Thermal profiles along with thermal impact of TSVs are studied for two layer and three layer stacked IC structures under practical power inputs. Experimental results show that there is a significant temperature gradient across the stacked dies for both two layer and three layer structures. The cross-layer temperature is seen to grow rapidly from two layer structures to three layer structures with the same power and TSV densities. As a result, stacking of active layers will not be scalable as the maximum temperature can quickly reach the 105 degree Centigrade limit for CMOS technology. Elevated temperature will make thermal-sensitive reliability issues major challenge for 3D stacked ICs. Advanced lower power, cooling, thermal management and new architecture techniques to keep temperature in safe range for stacking more layers into a simple chip.**

***Keywords – thermal; three dimensional (3D) chip structures; Through Silicon Via (TSV); integrated circuits (ICs); two layer; three layer; COMSOL***

## I. INTRODUCTION

Three dimensional stacked integrated circuit (3D IC) is a promising technology where two or more active layers are stacked in a single chip. 3D ICs present a number of benefits over 2D circuits: (1) Increased packing density attained by stacking a number of IC layers; (2) An opportunity to integrate heterogeneous layers into a single chip; and (3) Smaller average and maximum length of interconnects by using vertical interconnects, thereby reducing wire delays, which translates into increased data bandwidth. Potential applications of 3D ICs include processor stacking (CPU and various levels of caches) [1 – 2], memory stacking (SRAM, DRAM, and/or Flash) [3 – 5] and wireless sensing [6]. Through Silicon Vias (TSVs) provide a communication link for dies in the vertical direction to achieve 3D integration. Although alternatives like wire bonding and micro-bumping exist, TSV structures achieve higher interconnection density and better performance [7 – 10].

Temperature has become a major concern and constraint for high performance packaged integrated system design as more devices are integrated on a chip. This is because: (1) the heat flux generated by stacked multifunctional chips in miniature packages is extremely high due to increased thermal resistance; (2) 3D circuits increase total power generated per unit surface area; (3) chips in 3D stacks may get overheated if the cooling is not adequately provided; (4) significant thermal gradients across different layers and within a die may exist due to increased thermal paths and resistances; and (5) thin chips may create extreme conditions for on-chip hot spots. Cooling solutions, thermal management and related design problems continue to be identified by the Semiconductor Industries Association Roadmap [20] as one of the five key challenges during the next decade for achieving the projected performance goals of the industry. Thus, accurate and efficient thermal modeling and analysis is vital for the thermal-efficient circuit, chip architecture and structure and package designs to improve performance, reliability, power reduction, and online temperature regulation techniques [12 – 14].

In this paper, we study the impact of TSVs on thermal performance of 3D stacked ICs. First, we build realistic models of a 3D integrated circuit using an accurate modeling tool, COMSOL 4.1 software [15]. The software provides a tool to build realistic models that accurately capture the characteristics of the 3D stacked chip structures with practical material properties. We built two 3D stacked chip structures: (1) Two layer chip structure consisting of two silicon layers, each of dimensions 1mm x 1mm and thickness of 100um; and (2) Three layer chip structure consisting of three silicon layers, each of dimensions 1mm x 1mm and thickness of 100um. The chip structures in both the models are placed on a printed circuit board (PCB). Each silicon layer has 16 TSVs, distributed evenly on the chip. The impact of the TSVs thermal performance is then studied for two layer and three layer chip structures and the results are compared with each other. The Wattch tool [16] is used to provide a realistic input power profile to the chip in order to generate the heat source. Our study shows that there is a significant temperature gradient across the stacked dies for both the two layer and three layer structures. It is seen that the three layer chip structure experiences a much larger thermal gradient than the two layer chip structure under the same TSV density. In the case of three layered chip structures, the presence of TSVs has a limited impact on reducing the temperature gradient across

the different layers. It is also seen that the temperature gradients increase from 6 degree Centigrade to 18 degree Centigrade from two-layer to three-layer structures with the same power and TSV densities and distributions. This can be quite significant as it has an impact on thermal sensitive reliability issues such as thermal cycling, stress migration and negative bias temperature instability (NBTI) [21]. This trend is also observed for maximum temperature where the temperature difference between two layer and three layer structures is around 9 degree Centigrade. As a result, stacking of active layers will not be scalable as the maximum temperature can rapidly reach the 105 degree Centigrade limit for CMOS technology. Advanced cooling, low power and thermal management and new architecture techniques are hence required as temperature sensitive reliability issues are a major challenge for 3D stacked ICs.

This paper is organized as follows: Section II discusses the integrated 3D stacked chip model built using COMSOL 4.1 software, in detail. The experimental results are discussed in Section III of the paper and the conclusion is presented in Section IV of the paper.

## II. INTEGRATED 3D STACKED MODEL

The model is built using COMSOL 4.1 software. The dimensions of the silicon layers are 1mm x 1mm with a thickness of 100um. Fig. 1 shows the two layer model. It consists of two silicon layers with 16 TSVs placed on each of the silicon layers in the form of a 4 x 4 matrix. They are placed in such a way that they are equidistant to each other. The TSVs are made of a conductive material like copper and are surrounded by an insulation layer made of silica glass. The complete model consists of a total of 48 bumps – 16 on top of the top chip, 16 in between the two chips and 16 at the bottom of the bottom chip. The bumps are made of copper. An FR4 circuit board is in contact with the bottom layer of bumps. The top layer of bumps is in contact with the heat sink [19]. The mesh grid for the two layer TSV model is shown in Fig. 2.

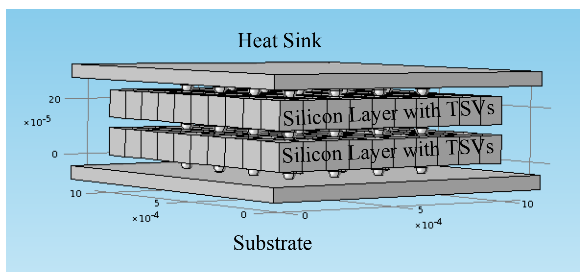

Figure 1. Two layer TSV stack model built using COMSOL 4.1 software

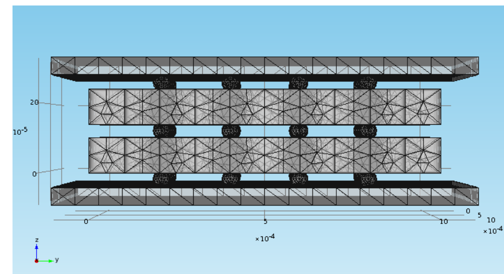
Figure 2. Mesh grid for two layer TSV model simulated using COMSOL 4.1 software

The three-layer model shown in Fig. 3 consists of three silicon layers with TSVs. A total of 48 TSVs and 64 bumps are used in the three layer model. The bottom layer of bumps is in contact with the FR4 circuit board and the top layer of bumps is in contact with the heat sink through which the heat gets dissipated. The mesh grid for the three layer TSV model is as shown in Fig. 4. The TSVs and bumps in both the two layer and three layer models are equidistant to each other.

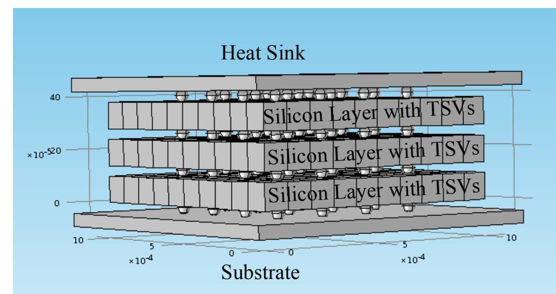

Figure 3. Three layer TSV stack model built using COMSOL 4.1 software

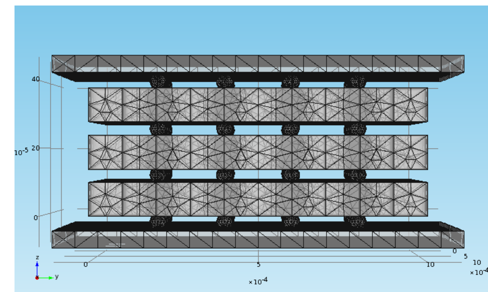
Figure 4. Mesh grid for three layer TSV model simulated using COMSOL 4.1 software

The convective boundary on top of the heat sink models the convective cooling from the fan placed above the processor. A convection coefficient of 450 (W/(m2-K)) is used to model the convective cooling effect from the cooling fan.

The silicon die in the model is partitioned into a four 2 x 2 matrices of power grids as shown in Fig. 5. Each grid represents a different function block. To model the power consumption of these function blocks, input power sources are placed in these power grids and the temperature is measured in each square of the power grid.

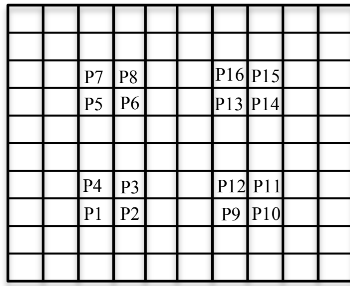


Figure 5. Partitioned die area with power grids

The simulation is run from 0.1 seconds to 19.7 seconds in steps of 0.1 seconds. Wattch power trace is used as the input to the thermal simulation. Finite-element based method of analysis is performed using COMSOL to get accurate transient thermal profiles for both the two layer and three layer chip structures.

## III. EXPERIMENTAL RESULTS

Experiments are performed using COMSOL 4.1 software on a Linux server with 1.6GHz quad-core CPU and 16GB memory. Wattch program is used to generate the power information with SPEC 2000 benchmarks [17]. The runtime power estimator is built using two important performance counts for each functional block of power. The parameter of each performance count is obtained through a linear regressor with samples from several benchmarks similar to [11], [18].

The thermal response was obtained by the finite-element method in COMSOL 4.1 software under the input power maps we generated. Fig. 6 shows the 3D view of the steady state temperature distribution for two layer TSV model. The temperature distribution shows that both the silicon layers get heated only in regions where the heat source is present. The surrounding areas do not get heated due to the presence of the TSVs. The difference between the hottest and coolest portion of the chip is around 6 degree Centigrade and the maximum temperature is around 73 degree Centigrade. Fig. 7 shows the top view of the chip which shows the localization of the heating to the areas on the chip with the heat source.

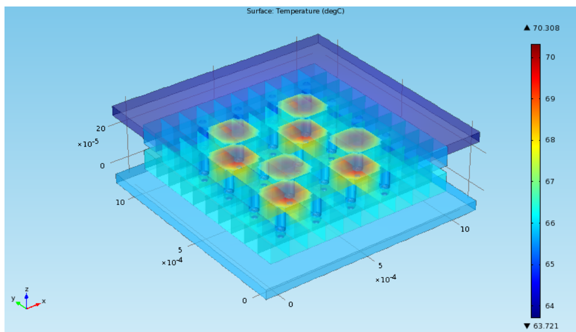


Figure 6. 3D view of steady state temperature distribution for two layer chip with TSVs

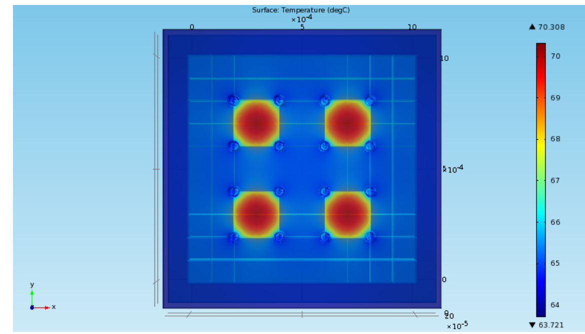


Figure 7. Top view of steady state temperature distribution for two layer chip

Fig. 8 shows the steady state temperature distribution under a given power input on the constructed package for the three layer chip and Fig. 9 shows the top view of the chip package. For the same input power trace, it is seen that the maximum temperature is around 82 degree Centigrade and the temperature difference between the hottest and coolest portions of the chip is around 18 degree Centigrade. Even with TSVs going through the silicon layers to the heat sink, the bottom layer gets heated throughout, thereby becoming the hottest layer in the chip structure. This implies that given the same TSV density, power density and power distribution for each layer, when stacked together, the cross-layer temperature gradients grow rapidly, showing that TSVs have a limited capacity to remove heat. Different TSV densities for different layers may be required but at a higher cost and overhead to chip design and manufacture processes.

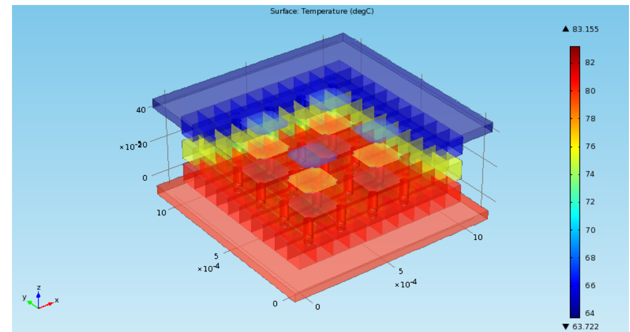


Figure 8. 3D view of steady state temperature distribution for three layer chip with TSVs

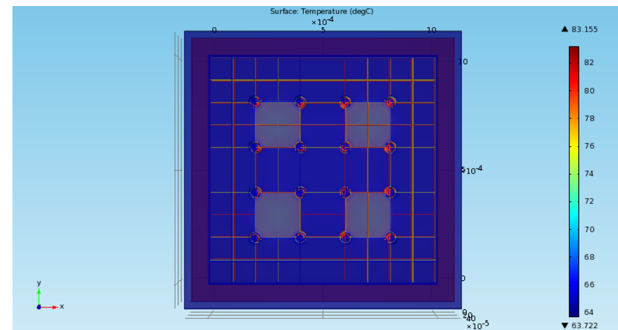


Figure 9. Top view of steady state temperature distribution for three layer chip

The maximum temperature difference between two-layer and three-layer structure sees a similar trend of around 9 degree Centigrade. This suggests that by stacking the active layers without reducing the power densities, the temperature can grow rapidly to reach the maximum allowed temperature of 105 degree Centigrade for CMOS technology. As a result, temperature remains to be the key challenge for TSV based 3D stacked ICs.

## IV. CONCLUSION

In this paper, we studied the thermal impact and characterization of TSVs based on the thermal performance of 3D stacked ICs using a realistic model using COMSOL 4.1 software. The thermal properties are studied for two layer and three layer stacked ICs. The heat source is modeled on all the silicon layers in certain areas of the chip and the impact of placing TSVs around the heat source is analyzed. Experimental results show that there is a significant temperature gradient across the stacked dies for both two layer and three layer structures. The cross-layer temperature is seen to grow rapidly from two layer structures to three layer structures with the same power and TSV densities. This trend is also observed for maximum temperature. As a result, stacking of active layers will not be scalable as the maximum temperature can rapidly reach the 105 degree Centigrade limit for CMOS technology. Advanced cooling, low power, thermal management and new architecture techniques are hence required as temperature sensitive reliability issues are a major challenge for 3D stacked ICs.

## V. ACKNOWLEDGEMENT

This work is supported in part by NSF grant under No. CCF-0902885, in part by NSF Grant under No. OISE-1130402, in part by Semiconductor Research Corporation (SRC) grant under No. 2009-TJ-1991.

## REFERENCES

[1] G.-H. Loh, Y. Xie, and B. Black, "Processor design in 3D die-stacking technologies," IEEE Computer, Vol. 27, No. 5, May 2007, pp. 31–48.

[2] C.-C. Liu, I. Ganusov, M. Burtscher, and S. Tiwari, "Bridging the processor-memory performance gap with 3D IC technology," IEEE Design & Test of Computers, Vol. 22, Nov.-Dec. 2005, pp. 556–564.

[3] K.-T. Park, D. Kim, S. Hwang, M. Kang, H. Cho, Y. Jeong, Y. Seo, J. Jang, H.-S. Kim, S.-M. Jung, Y.-T. Lee, C. Kim, and W.-S. Lee, "A 45nm 4Gb 3-dimensional double-stacked multi-level NAND flash memory with shared bitline structure," in Dig. Tech. Papers Int. Solid-State Circuits Conf., Feb. 2008, pp. 510–632.

[4] P. Batude, M.-A. Jaud, O. Thomas, L. Clavelier, A. Pouydebasque, M. Vinet, S. Deleonibus, and A. Amara, "3D CMOS integration: introduction of dynamic coupling and application to compact and robust 4T SRAM," in Proc. Int. Conf. Integrated Circuit Design Tech., June 2008, pp. 281–284.

[5] U. Kang, H.-J. Chung, S. Heo, S.-H. Ahn, H. Lee, S.-H. Cha, J. Ahn, D. Kwon, J.-H. Kim, J.-W. Lee, H.-S. Joo, W.-S. Kim, H.-K. Kim, E.-M. Lee, S.-R. Kim, K.-H. Ma, D.-H. Jang, N.-S Kim, M.-S. Choi, S.-J. Oh, J-B. Lee, T.-K. Jung, J.-H. Yoo, and C. Kim, "8Gb 3D DDR3 DRAM using through- silicon-via technology," in Dig. Tech. Papers IEEE Int. Solid-State Circuits Conf., Feb. 2009, pp. 130–131.

[6] P. Ramm and A. Klumpp, "Through-silicon via technologies for extreme miniaturized 3D integrated wireless sensor systems (e-CUBES)," in Proc. Int. Interconnect Tech. Conf., June 2008, pp. 7–9.

[7] B. Kim, C. Sharbono, T. Ritzdorf, and D. Schmauch, "Factors affecting copper filling process within high aspect ratio deep vias for 3D chip stacking," in Proc. Electronic Components Tech. Conf., May-June 2006.

[8] P. Benkart, "3D chip stack technology using through-chip interconnects," IEEE Design & Test of Computers, Vol. 22, Nov. 2005, pp. 512-518.

[9] M. Motoyoshi, "Through-silicon via (TSV)," in Proc. IEEE, Vol. 97, No. 1, Jan. 2009, pp. 43–48.

[10] G. Campardo, G. Ripamonti, and R. Micheloni, "Scanning the issue: 3-D integration technologies," Proc. IEEE, Vol. 97, No. 1, Jan. 2009, pp. 5–8.

[11] W. Wu et al., "A systematic method for functional unit power estimation in microprocessors," in Proc. Design Automation Conf. (DAC), June 2006, pp. 554–557.

[12] Yu, A., N. Khan, G. Archit, D. Pinjala1, K. Toh, V. Kripesh1, S. Yoon, and J. H. Lau, "Development of silicon carriers with embedded thermal solutions for high power 3-D package", IEEE Proceedings of Electronic, Components & Technology Conference, Orlando, FL, May 27-30, 2008, pp. 24-28.

[13] Kumagai, K., Yoneda, Y, Izumino, H., Shimojo, H., Sunohara, M., Kurihara, T., "A silicon interposer BGA package with Cu-filled TSV and multi-layer Cu-plating interconnection," IEEE Proceedings of Electronic Components and Technology Conf, Orlando, FL, May 2008, pp. 571-576.

[14] Premachandran, C. S., M. Chew and W. Choi, A. Khairyanto, K. Chen, J. Singh, S. Wang, Y. Xu, N. Chen, C. Sheppard, M. Olivo, and J. H. Lau, "Influence of Optical Probe Packaging on a 3D MEMS scanning micro mirror for Optical Coherence Tomography (OCT) applications", IEEE Proceedings of Electronic, Components & Technology Conference, Orlando, FL, May 27-30, 2008, pp. 829-833.

[15] "COMSOL Mutiphysics: User guide," Version 4.1.

[16] D. Brooks, V. Tiwari, and M. Martonosi, "Wattch: A framework for architectural-level power analysis and optimizations," in Proc. Int. Symposium on Computer Architecture (ISCA), 2000, pp. 83–94.

[17] J. L. Henning, "SPEC CPU 2000: Measuring CPU performance in the new millennium," IEEE computer, vol. 1, no. 7, pp. 28–35, July 2000.

[18] M. D. Powell et al., "CAMP: A technique to estimate per-structure power at run-time using a few simple parameters," in Proc. IEEE Int. Symposium on High-Performance Computer Architecture (HPCA), 2009, pp. 289–300.

[19] F. Hameed, M. A. A. Faruque, J. Henkel, "Dynamic Thermal Management in 3D Multi-Core Architecture through Run-Time Adaptation", Date 2011.

[20] "International technology roadmap for semiconductors (ITRS)", 2011," 2011, http://public.itrs.net.

[21] "Failure Mechanisms and Models for Semiconductor Devices," In JEDEC Publication JEP122-A, Jedec Solid State Technology Association, 2002.