

# PARAMETRIC THERMAL MODELING OF 3D STACKED CHIP ELECTRONICS WITH INTERLEAVED SOLID HEAT SPREADERS

David W. Gerlach and Yogendra K. Joshi  
G. W. Woodruff School of Mechanical Engineering  
Georgia Institute of Technology  
771 Ferst Dr.  
Atlanta, Georgia, 30332-0405  
Phone: (404) 385-1878  
Fax: (404) 894-8496  
Email: david.gerlach@me.gatech.edu

## ABSTRACT

Effective methods must be devised to transfer the heat from the core of 3D stacked chip electronics to the exterior of the device. The use of solid heat spreaders of high thermal conductivity interleaved between the chips was investigated parametrically through computational modeling. The effect of the power dissipated, the applied heat transfer coefficient, the spreader thickness, spreader thermal conductivity, and the shape of via holes in the spreader were modeled. Results show that for moderate power dissipations, 5 W in each 27x38 mm layer, a 250  $\mu\text{m}$  thick copper heat spreader would conduct heat adequately.

**KEY WORDS:** thermal management, three dimensional, heat conduction, simulation, IC, integrated circuit

## INTRODUCTION

3D stacked chip electronics pose a serious cooling challenge due to the increase in volumetric heat generation coupled with the limited heat removal surface area [1]. Effective methods must be devised to transfer the heat from the core of the stack to the exterior of the device. Solid heat spreaders of high thermal conductivity could be interleaved between the chips. One of the primary goals of 3D stacked chip architectures is the reduction of signal path length to decrease the time for signal propagation. As such, the layer-to-layer spacing must be minimized.

However, 3D architectures that rely on the conduction of heat perpendicular to the stacking planes are inherently limited in scalability. As the stack gets taller, the thermal load and resistance both increase. Im and Banerjee [2] modeled 3D ICs that were vertically stacked and glued together with polyimide. Because all of the heat was removed from the bottom of the stack, the heat had to flow through repeated layers of dielectric with low thermal conductivity. This led to a maximum die temperature in the range of 380-400°C. They conclude that higher thermal conductivity dielectrics must be developed. Kleiner et al. [3] modeled a similar geometry and verified it experimentally. They found that the measured thermal resistance of the polyimide layer was 10 times that

predicted by the bulk conductance value, due to the interface resistance between the layers. Akturk et al. [4, 5] used thermal resistor networks to model stacked Pentium III processor chips. Because the heat was also conducted vertically through the chips and SiO<sub>2</sub> insulation between them, the maximum processor temperature increased with the number of layers. This would limit the number of stacks to less than six.

Goplen and Sapatnekar [6] presented an algorithm for moving cells in a 3D IC away from high temperature regions, thus reducing temperature while minimally increasing the wire lengths. However, this involves redesigning the ICs and possibly having different electrical designs in different layers.

A more elegantly scalable solution is for the heat to be conducted out of the sides of the stack. If the heat produced in each layer is removed in the area of the layer perimeter, then the stack height will not be limited thermally and the same planar IC could be used in each layer.

## MODEL GEOMETRY AND PARAMETERS

The temperature profiles in the stacked layers of a 3D electronics design were computed using a commercial 3D finite element code (ANSYS). The heat output of the processor, thermal conductivity of the spreader layer, its thickness, the heat transfer coefficient at the layer edges, and the shape of via cutouts in the thermal layer were varied. Table 1 in the Appendix has a complete list of the parameters used in the runs. The design space was studied by varying one parameter at a time in a cross shaped matrix, as opposed to a full factorial analysis.

Each tier consists of an active layer, a copper thermal spreader layer, and the interface between them (Figure 1). The active layer is 200  $\mu\text{m}$  thick silicon. Two DRAM ICs and one field programmable gate array (FPGA) IC are epoxied into cutouts in the active layer silicon (Figure 2). In addition, several blocks of decoupling capacitors on silicon are epoxied into the layer. The total tier size for this study was held constant at 27 mm x 38 mm. For reasons of symmetry, only half of the layer was modeled.

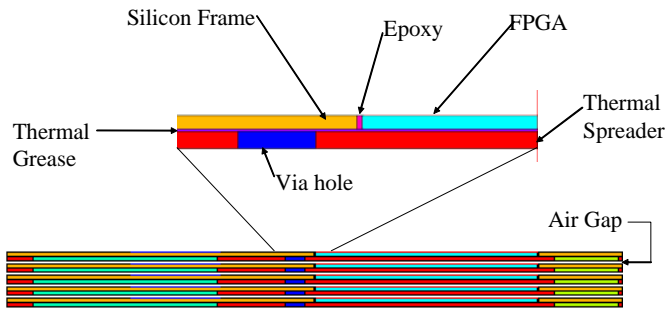


Fig. 1 Side view of stack with expanded schematic of a layer section.

The layer lateral dimensions are primarily determined by the size of the FPGA, DRAM, and capacitor blocks. However, the decoupling capacitor regions were removed from the model due to their low power consumption to decrease the model complexity. Because they were composed of silicon with a small region of epoxy around them they did not greatly affect the temperature profile in the silicon layer (Compare Figure 2 and Figure 3). The original layer size was retained at 27x38 mm to reflect the space needed for the capacitors without explicitly modeling them.

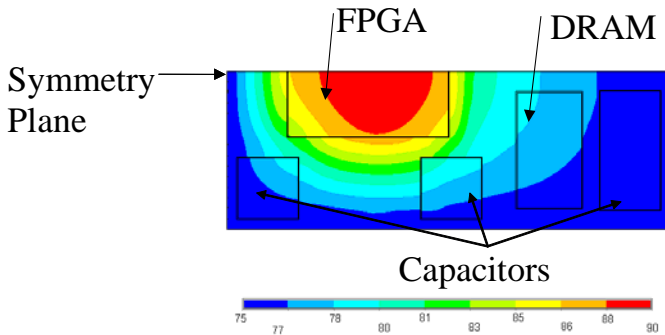


Fig. 2 Top view of computed temperature profiles including the decoupling capacitors. Note: this calculation was performed with an isothermal boundary condition

The base case thermal layer thickness was 250  $\mu\text{m}$ . Higher and lower values of 500  $\mu\text{m}$  and 50  $\mu\text{m}$  were used in the parametric runs.

A significant feature of this design is that holes must be cut in the thermal management layer for the layer-to-layer interconnects. These holes are major obstacle to heat flow from the chips to the layer edge. The vias extend through an air space to allow for variations in thermal expansion between the silicon and copper layers. Because the total number of vias needed between layers is determined by the electrical design, the area taken up by the via holes is fixed. However, this area may be arranged to create better thermal pathways to the edges. The fraction of the FPGA perimeter that was occluded by the vias was varied parametrically. The fraction of the distance the vias took up between the FPGA and the layer edge was then calculated to keep the via hole area constant for all runs (Figure 4). The size of the via hole between the DRAMs is not varied.

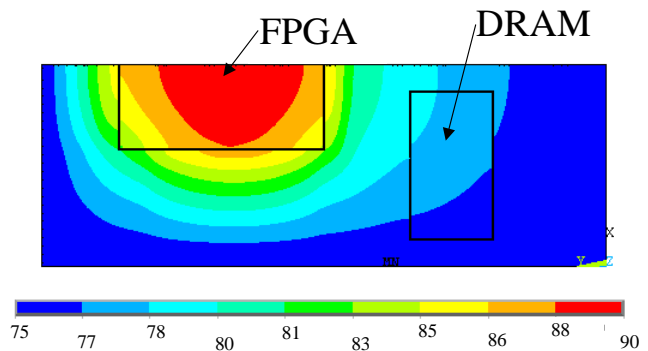


Fig. 3 Top view of computed temperature profiles with the decoupling capacitors removed and isothermal 75°C layer edges

The thermal conductivities of all materials were taken as constant. The conductivity of the silicon was taken as 145 W/mK. The epoxy is 0.3 W/mK. The thermal grease/adhesive between the layers has a conductivity of 4.4 W/mK. The via hole air has a conductivity of 0.01 W/mK. The base case spreader was made of copper with conductivity of 390 W/mK. The other cases are aluminum with  $k=230$  W/mK and diamond with  $k=1000$  W/mK.

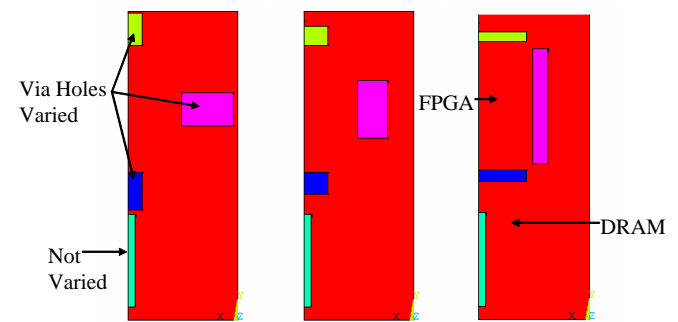


Fig. 4 Left: Low perimeter, Middle: base case, Right: High perimeter. These schematics show the bottom views

The heating from the chips was applied as a uniform heat flux on the top surface of the chip. The FPGA puts out 5 W for the base case and the DRAMs put out 0.4 W. The temperatures for the base configuration were also calculated with FPGA powers of 1 and 10 W. The DRAM power was not varied. The Joule heating in the interconnects was ignored. The heating from the chips was held constant in all runs.

The heat transfer coefficient at the edges of the layers is specified. The base case heat transfer coefficient was derived by assuming the layer edge to be isothermal at 70  $^{\circ}\text{C}$  and the ambient at 40  $^{\circ}\text{C}$ , with power dissipation uniformly spread over the surface area. This yielded a heat transfer coefficient of 0.0024 W/m<sup>2</sup>K which is achievable with advanced air cooled heatsinks. Heat transfer coefficients an order of magnitude higher (0.02 W/m<sup>2</sup>K) and lower (0.0002 W/m<sup>2</sup>K) were also used in the parametric runs. The air temperature was 40  $^{\circ}\text{C}$  for all runs.

## IC TEMPERATURE LIMITS

The temperature limits of memory and processor dies are driven by different constraints. The processor limit is driven by reliability concerns. The maximum temperature limit commonly used is 85 °C on the package surface. However, the temperature of the junction may be considerably higher. In this work the limit on the FPGA die is taken as 125 °C. The upper temperature limit of the memory is constrained by the increase in memory refresh rate as the memory becomes more volatile with increasing temperature. This reduces the memory's availability. Therefore, the maximum temperature for any point on the memory is taken as 85 °C. Conveniently, the higher power component has the higher temperature limit. However, the silicon active layer and thermal spreader layer both easily conduct heat from the FPGA to the memory. Consequently the memory temperature is the limiting factor.

Because the processor consumes more power and has a higher temperature limit, thermally isolating the memory and processor dies could allow them to operate at different temperatures. Operating the processor at a higher temperature than the memory could also allow the average temperature of the stack edge to be higher. The higher driving temperature difference to ambient would allow for a smaller system for rejecting the heat to the ambient through a heatsink or fluidic device such as a thermosyphon.

## MODELING RESULTS

The temperature contours for the base case are shown in Figure 5. The maximum temperatures in the FPGA and DRAM and the lowest temperature in the layer are listed for each run in the Table 2 in the Appendix. The FPGA and DRAM temperatures are quite close to each other in all the runs, due to heat conduction through the silicon.

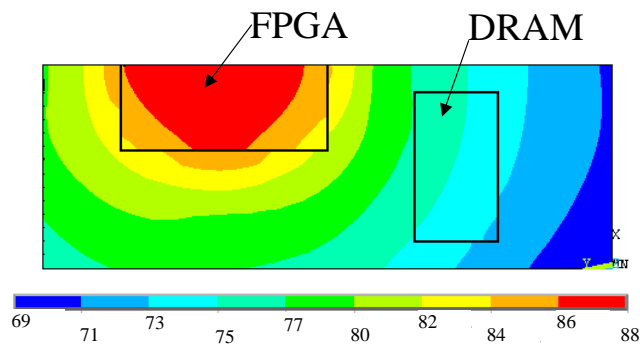


Fig. 5 Temperature profiles of base case

The power output of the FPGA dominates over the smaller DRAM dissipation. Therefore the FPGA die temperatures are linearly related to the power (Figure 6). Therefore a thermal resistance from the FPGA to ambient can be calculated by dividing the power output of the FPGA (5W) by the difference between the highest temperature on the FPGA and the ambient temperature (Table 2 Appendix).

The heat transfer coefficient applied to the stack has a dramatic effect on the temperatures in the layers (Figure 7).

Considerable thought must be given to ensuring that the layer edge temperature is held sufficiently low.

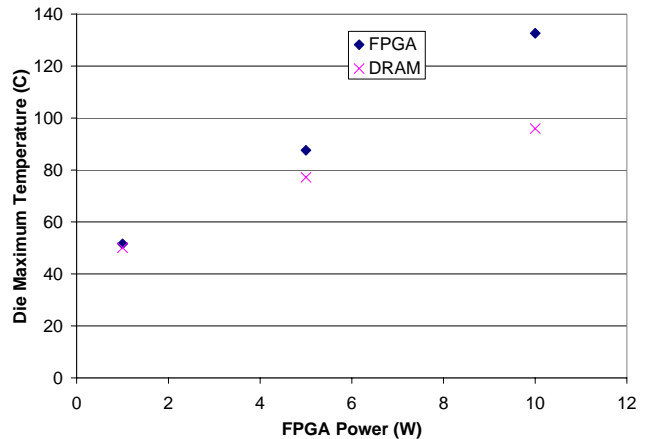


Fig. 6 The effect of FPGA power output on the die temperatures

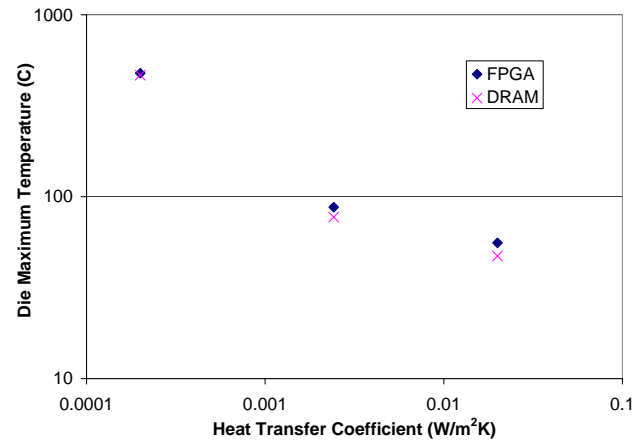


Fig. 7 The effect of external heat transfer coefficient on IC temperatures

Figure 8 shows the effect of the shape of the via holes on the die temperature. As the holes increasingly occlude the heat pathway to the edge, the FPGA temperature increases as expected. The DRAM temperature stays nearly constant because the via holes do not prevent its power dissipation from reaching the edge. However, the presence of holes in the thermal layer does not prevent the heat from the FPGA reaching the memory, because the silicon layer has a fairly high thermal conductivity. In order to isolate the DRAM from the FPGA it is necessary to place a thermal barrier in the silicon active layer. Arranging the vias such that they leave a large part of the die perimeter open but increase the average distance from the die is better from a thermal perspective. However, this may increase the length of the electrical signal paths unacceptably. In addition, increasing the width or length of the layer to move the vias may not decrease the thermal resistance.

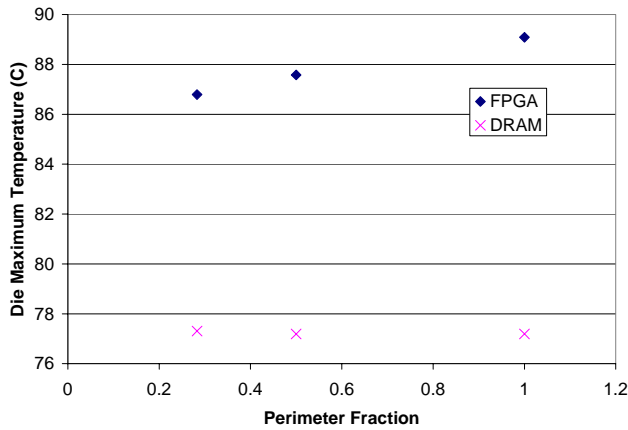


Fig. 8 Effect of via hole shape on die temperature

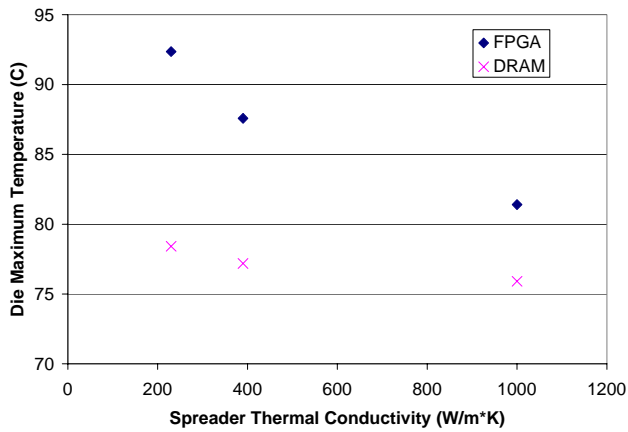


Fig. 9 Effect of thermal conductivity on die maximum temperature.

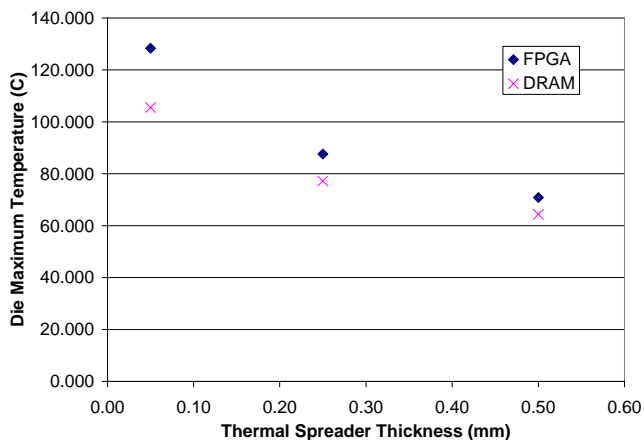


Fig. 10 The effect of thermal spreader thickness on IC temperatures.

As expected, the die temperatures decrease with increasing spreader thermal conductivity and thickness (Figures 9 and 10). The temperatures of the processor and memory also converge as the heat flow pathway between them increases

with conductivity and thickness. The differences between the high and low temperatures in the runs when thermal layer conductivity and thickness values are varied are plotted against the conductivity multiplied by the thickness (Figure 11). As a result, the data collapses into a simple curve. Because the thickness corresponds to the cross sectional area in the direction of heat flow, the value of conductivity times thickness is related to the layer thermal resistance. This shows the direct tradeoff between layer thickness and conductivity.

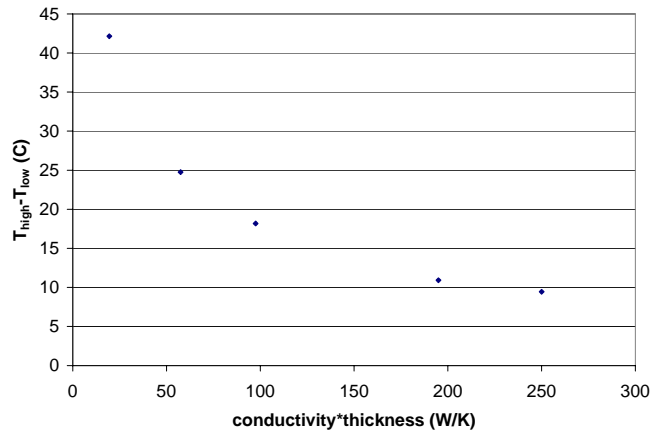


Fig. 11 The effect of the thermal conductivity and the thermal layer thickness on the temperature rise in the active layer.

## CONCLUSIONS

Finite element simulations show that solid heat spreaders can be used to conduct the heat from the interior of stacked chip electronics to the periphery if the heat transfer coefficient to the ambient is high enough. Conduction of heat out of the sides of the stack produces a very scalable design. In a stack with processors and memories the memories limit the operation temperature. Further work on reducing the thermal conductance from the processor to memory would ease this restriction.

## Acknowledgments

The authors would like to acknowledge the DARPA Microsystems Technology Office for support through the 3D MINT program.

## References

- [1] J. A. Davis, R. Venkatesan, A. Kaloyeros, M. Beylansky, S. J. Souri, K. Banerjee, K.C. Saraswat, A. Rahman, R. Reif, and J.D. Meindl, "Interconnect limits on gigascale integration (GSI) in the 21st century," Proceedings of the IEEE, vol. 89, no. 3, pp. 305-324, Mar. 2001.
- [2] Sungjun Im and Kaustav Banerjee, "Full Chip Thermal Analysis of Planar (2-D) and Vertically Integrated (3-D) High Performance ICs," International Electron Devices Meeting 2000 Technical Digest, pp. 727-30, 2000.
- [3] M.B. Kleiner, S.A. Kuhn, P. Ramm, and W. Weber, "Thermal Analysis of Vertically Integrated Circuits" International Electron Devices Meeting 1995 Technical Digest, pp. 487-90, 1995.

[4] A. Akturk, N. Goldsman, and G. Metze, "Coupled simulation of device performance and heating of vertically stacked three-dimensional integrated circuits," 2005 International Conference on Simulation of Semiconductor Processes and Devices, pp. 115-18, 2005.

[5] A. Akturk, N. Goldsman, and G. Metze, "Self-consistent modeling of heating and MOSFET performance in 3-D integrated circuits," IEEE Transactions on Electron Devices, vol. 52, no. 11, pp. 2395 – 2403, Nov. 2005.

[6] B. Goplen and S. Sapatnekar, "Efficient thermal placement of standard cells in 3D ICs using a force directed approach" International Conference on Computer Aided Design , pp. 86-89, 2003.

## Appendix

Table 1 Parameter values for simulations

Run	heat transfer		conductivity	perimeter fraction		FPGA power
	coefficient	thickness		fraction	to edge	
	W/m <sup>2</sup> *K	mm	W/m <sup>2</sup> *K			W
1	0.0024	0.250	390	0.50	0.51	5
2	0.0002	0.250	390	0.50	0.51	5
3	0.0200	0.250	390	0.50	0.51	5
4	0.0024	0.050	390	0.50	0.51	5
5	0.0024	0.500	390	0.50	0.51	5
6	0.0024	0.250	230	0.50	0.51	5
7	0.0024	0.250	1000	0.50	0.51	5
8	0.0024	0.250	390	0.28	0.90	5
9	0.0024	0.250	390	1.00	0.24	5
10	0.0024	0.250	390	0.50	0.51	1
11	0.0024	0.250	390	0.50	0.51	10

Table 2 Results of simulations

Run	FPGA high	DRAM high	layer low	total
	temperature	temperature	temperature	resistance
	C	C	C	K/W
1	88	77	69	9.15
2	477	466	457	84.02
3	56	47	42	3.03
4	128	106	86	16.99
5	71	64	60	5.92
6	92	78	68	10.07
7	81	76	72	7.96
8	87	77	69	9.00
9	89	77	69	9.44
10	52	50	48	9.63
11	133	96	111	9.08