Budapest, Hungary, 17-19 September 2007

Development of a Prototype Thermal Management Solution for 3-D Stacked Chip Electronics by Interleaved Solid Spreaders and Synthetic Jets

Donavon Gerty[1], David W. Gerlach[1], Yogendra K. Joshi[1], and Ari Glezer[1]
[1]G. W. Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, GA, 30332
Phone: (404) 894-2446, Fax: (404) 894-8496
Email: dgerty@gatech.edu

Abstract- A design for cooling 3D stacked chip electronics is proposed using solid heat spreaders of high thermal conductivity interleaved between the chip layers. The spreaders conduct heat to the base of an advanced synthetic jet cooled heat sink. A prior computational study [1] showed that for moderate power dissipations, 5 W in each 27 × 38 mm layer, a 250 µm thick copper heat spreader would conduct heat adequately. However, the mismatch in coefficient of thermal expansion between copper and silicon required large holes through the copper layer for electrical vias. The current study investigates the design of a thermal prototype for experimental testing. Each active layer will incorporate a thermal test die to simulate an FPGA and a smaller one to simulate a DRAM (Fig. 2). The spreader layer will be silicon with no via holes. The heat sink will contact only three of the stack sides to allow wirebond connections on the fourth side (Fig. 3). The effect of the power dissipated and the heat transfer coefficient applied to the peripheral surface are studied. In order to remove the heat from the edges of a multi-layer stack and transfer it to the ambient air, a novel active heat sink design has been implemented using a matrix of integrated synthetic jets. In previous synthetic jet heat sink designs, cooling air is entrained upstream of the heat sink and is driven along the length of the fins. In the new design, synthetic jets emanate from the base of the fins so that the induced jets and entrained (cooling) ambient air flow along the fin height. The velocity field of the active heat sink is mapped using particle image velocimetry (PIV). Thermal performance is characterized using a surrogate heater and embedded thermocouple sensors. The thermal performance of identical heat sinks cooled by the two synthetic jet approaches is compared. An improved third heat sink solution is introduced and compared to previous results [2].

I. INTRODUCTION

Thermal issues exist over a wide range of power dissipation levels (ITRS[3]), from handheld devices that dissipate a few Watts to high-performance microprocessors dissipating over 100 W. While these steep cooling requirements have prompted the development of advanced two-phase and pumped liquid cooling techniques, consumer-oriented systems still focus on air cooling approaches due to their simplicity and relative ease of implementation (Bar Cohen [4]). 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 [3]. Effective methods must be devised to transfer the heat from the core of the stack to the exterior of the device and its ultimate rejection to the ambient environment.

Most 3-D electronics designs use conduction perpendicular to the layers through the device to a conventional finned heat sink-fan combination, placing a limit on the number of layers that can be stacked. It has also been proposed to insert microchannels between the circuit layers [5], thus removing the limit on the number of stacked layers. However, it is inherently more complex than conduction based systems. In the design developed in this work, the heat is conducted from the chips to the stack edge by solid heat spreaders of high thermal conductivity interleaved between the chip layers. The heat is then conducted through a finned metal heat sink that surrounds the stack edges (Fig. 1), and is rejected to the ambient air by synthetic jets blowing across the heat sink fins.

Although the connection between the stack edges and the heat sink base has not been designed or modeled in detail, a significant safety factor allows for the interface resistance. The resistance from static edge to the surroundings used in the stack modeling is 8.0 K / W. The resistance of the heat sink is calculated as 1.25 K / W. Therefore, the allowable interface resistance could be as high as 6.75 K / W. Simple calculations for joints thicknesses of 25 to 100 µm using solder or thermally conductive adhesives (k = 1-60 W / m K) give resistances of less than 2 K / W.

Review of Heat Conduction in 3-D Architectures

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 [6] 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 ºC.

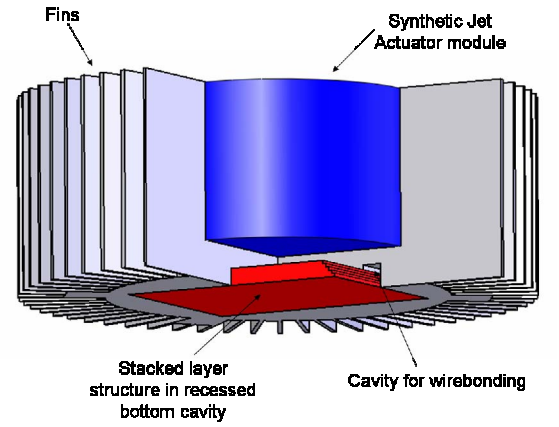

Fig. 1: View of design concept with bottom front corner cut away to show the position of the stacked layer in the heat sink base and wire bond cavity

They conclude that higher thermal conductivity dielectrics must be developed. Kleiner et al. [7] 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. [8, 9] used thermal resistor networks to model stacked Pentium III processor chips. Because the heat was also conducted vertically through the chips and SiO_2 insulation between them, the maximum processor temperature increased with the number of layers. This would limit the number of stack layers to less than six. Goplen and Sapatnekar [10] 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 individual layer is removed through its own individual perimeter area, then the stack height will not be limited thermally and the same planar IC could be used in each layer.

Review of synthetic jets

Forced convection cooling using air is typically based on the use of various configurations of fans and blowers that can fit in medium and large scale enclosures and are used both for global air circulation and for local heat transfer augmentation. In order to achieve increased local power dissipation levels with fan-heat-sink configurations, designers are using copper heat sinks as well as larger fans driving higher flow rates. Although fans can supply ample volume flow rates, they only support relatively low pressure drop and are hindered by noise and low thermal effectiveness.

Synthetic jets offer an attractive solution for highly efficient localized cooling of integrated circuits. These jets are formed by time-periodic, alternate suction and ejection of fluid through an orifice bounding a small cavity, by the time periodic motion of a diaphragm that is built into one of the walls of the cavity. Unlike conventional jets, synthetic jets are "zero net mass flux" in nature and produce fluid flow with finite momentum with no mass addition to the system and without the need for complex plumbing (Smith and Glezer [11]). Because of their ability to direct airflow along heated surfaces in confined environments and induce small-scale mixing, these jets are ideally suited for cooling applications at the package and heat sink levels. While there is extensive literature on cooling with steady and unsteady conventional jets (e.g., Jambunathan et al. [12]), the concept of using synthetic jets for heat transfer is relatively new. It was first implemented by Thompson et al. [13] who demonstrated a 250% increase in power dissipation over natural convection for direct normal impingement cooling of a 49-element MCM using a single synthetic jet having a diameter of 1.6 mm. In a later investigation, Russell [14] showed that the inherent coupling between a local synthetic jet and global air flow (driven by a conventional fan) can be exploited for enhanced heat transfer at the package level at substantially reduced global air flow. Mahalingam and Glezer [15] developed an integrated active heat sink based on synthetic jets for heat dissipation at power levels over 100 W at flow rates of about 3-4 CFM. Mahalingam and Glezer [16] also demonstrated a low-volume active heat sink based on an impinging synthetic jet for moderate power dissipation requirements. A later study [17] demonstrated augmentation of heat transfer from a heat sink by implementing a synthetic jet to reduce the flow bypass of a global fan driven flow.

II. HEAT CONDUCTION WITHIN A STACKED LAYER STRUCTURE

Model Geometry and Parameters

The temperature profiles in the stacked layers of a 3D

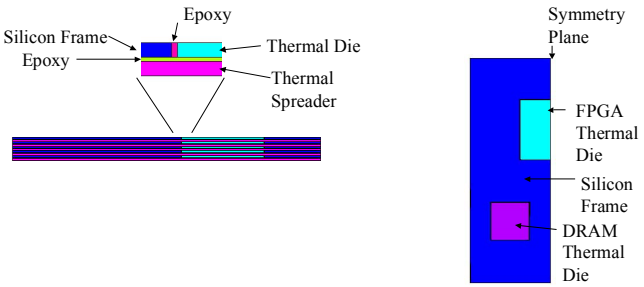

Fig. 2: Left: Mid-sectional side view of stack with an expanded schematic of a layer section of 3D stack; Right: top view of half of active layer

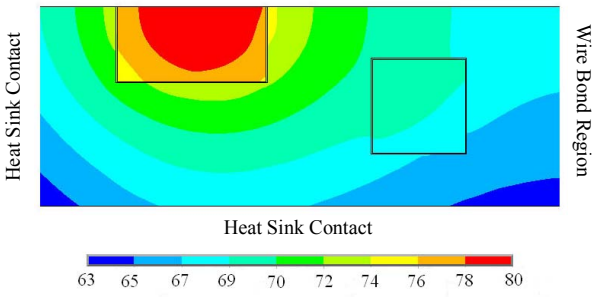

Fig. 3: Contour plot of temperatures on the upper surface of the chip layer in the base case of the parametric study

electronics design (Fig. 2) were computed using a commercial 3D finite element code (ANSYS). The heat output of the processor and the heat transfer coefficient at the layer edges were varied. Table 1 in the Appendix has a list of the parameters used in the runs.

Each tier in Fig. 2 consists of an active silicon layer, a silicon thermal spreader layer, and the thermal grease interface between them. The active layer and heat spreader are both 200 µm thick silicon. The size of each layer will vary in order to allow access for wirebonding as shown in Fig. 1. Because the size variation will be in the region not bonded to the heat sink, the total tier size for this study was held constant at 27 mm × 38 mm. In addition, only one layer was simulated and conduction vertically through the stack was ignored. Two DRAM ICs and one field programmable gate array (FPGA) IC are epoxied into cutouts in the active layer silicon as seen in the top view in Fig. 3. Because the final design with active chips will be symmetrical, only half of the layer was modeled. However, the thermal prototype will include only one DRAM thermal test die. Assuming symmetry and modeling both dies is conservative in this case.

The thermal conductivities of all materials is taken as constant. The conductivity of the silicon is taken as 145 W /m K. The epoxy is 0.3 W /m K. The thermal grease/adhesive between the layers has a conductivity of 4.4 W / m K.

The heating from the chips is 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 DRAM puts out 0.2 W. The temperatures were also calculated with FPGA powers of 2.5 and 10 W. The DRAM power is not varied. The Joule heating in the interconnects is ignored.

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 75 ºC and the ambient at 40 ºC, with power dissipation uniformly spread over the surface area. This yielded a heat transfer coefficient of 0.0026 $W / mm^2 K$. This is a resistance to the environment of 6.5 K / W, which is considerably higher than that of the advanced air cooled heat sinks described below. Heat transfer coefficients of 0.00068, 0.0014, 0.0043, 0.0068, and 0.034 $W / mm^2 K$ were also used in the parametric runs. These values correspond to thermal resistances of 25, 12.5, 4, 2.5, and 0.5 K / W. The ambient air temperature was 40 ºC for all runs.

Modeling Results

The temperature contours for the base case are shown in Fig. 3. The maximum temperatures in the FPGA and DRAM are listed for each run in Table 1. The maximum temperature of the thermal test die simulating the FPGA is shown in Fig. 4 as a function of the layer total power dissipation for several averaged heat sink thermal resistances. Fig. 5 shows similar results for the DRAM maximum temperature. The maximum temperature limit of each die is noted on the graphs with a horizontal dashed line. The 6.5 K / W resistance and 5.4 W power dissipation condition is eliminated by the lower temperature limit of the DRAM and limits the maximum power dissipation and thermal resistance. At combinations of higher powers and thermal

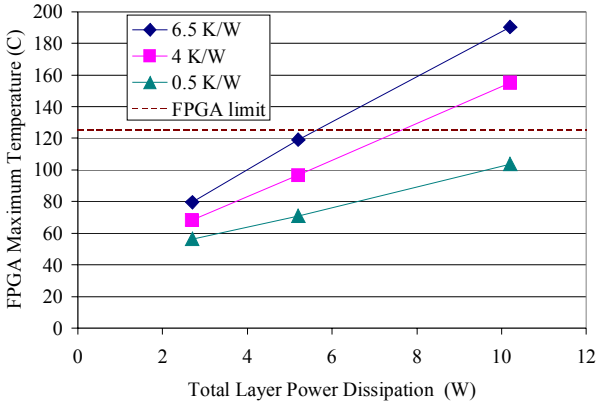

Fig. 4: The effect of power dissipation on the maximum temperatures predicted on the FPGA thermal die for thermal resistances from the stack edge to the environment of 6.5, 4 and 0.5 K / W

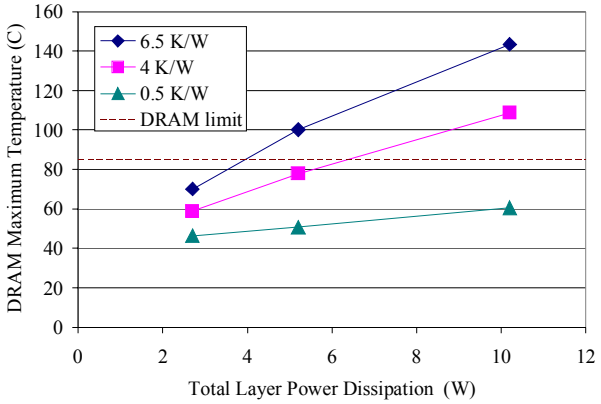

Fig. 5: The effect of power dissipation on the maximum temperatures predicted on the DRAM thermal die for thermal resistances from the stack edge to the environment of 6.5, 4 and 0.5 K / W

resistances, both the FGPA and DRAM dies are over their temperature ranges. Resistances over 6.5 K / W are not included in the graphs, because the die temperatures are too high for all power levels that were simulated.

III. SYNTHETIC JET MODULE DESIGN AND TESTING

To ensure that the thermal resistance from the stack edges to the ambient is sufficiently low, a synthetic jet cooled heatsink was selected. Fig. 6 shows the basic operation and jet implementation for two radically different approaches for integrating synthetic jets with heat sinks. Previous work has focused on placing the jets at the end of the fins and using them to induce air flow longitudinally down the length of the fin as shown in Fig. 6a. A primary array of jets entrains and drives air flow through the channels created by the fins. The new heat sink design allows the primary jet to emanate directly through the heat sink base and entrain and expel air laterally over the height of the fins through a matrix of synthetic jets as shown in Fig. 6b. This forms cellular pockets of recirculating air between each fin, with fresh air entrained during one half-cycle of the actuator operation and the heated air expelled laterally during the second half.

There are several advantages to inducing air flow over the fins as shown in Fig. 6b. First, the air travels across the short length of the fins laterally, rather than longitudinally down the long length. The air moves a shorter distance, making it possible to entrain fresh, cool air easier and faster, making it more effective for cooling. This entrainment of fresh air is important because the air heats up as it moves along the fin, reducing the temperature difference between the fin and air, resulting in lower convective heat transfer. For the new setup, since the jet nozzles are located inside the fin channels, the entrained air is heated not only on the ejection stroke, but also on the suction stroke of the cycle, allowing the air to remove more heat per cycle. Also, the jets experience a lower pressure drop since they flow along the shortest length of the fins. Finally, the entire fin surface area is uniformly inundated with a high velocity stream of air resulting in a high heat transfer coefficient.

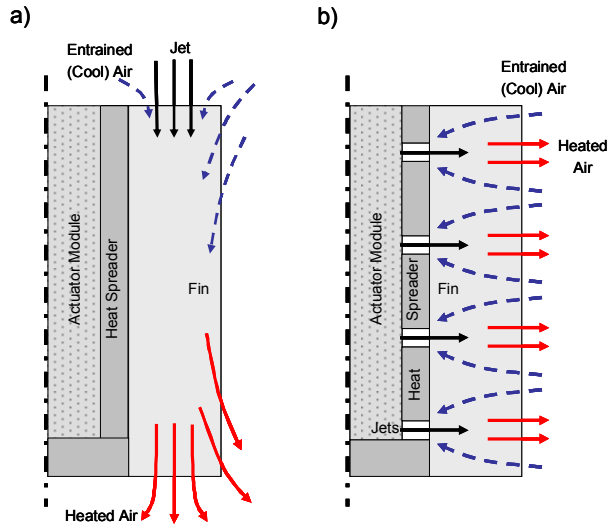

Fig. 6: Basic principle of operation of a) original heat sink with longitudinal air flow and b) newly designed heat sink with lateral air flow (due to symmetry, only half of heat sink shown)

Experimental Apparatus and Testing

In order to make a direct comparison between the two different flow configurations shown in Fig. 6, two heat sinks are designed and the thermal performance characterized. The heat sinks are designed in a U-shape allowing for an onboard location to place the module that contains electromagnetic actuators and also minimizes the total footprint of the integrated heat sink and actuator. The actuators are mounted in the inserted module and route the airflow to the fins. The envelope dimensions are 5.08 cm wide, 4.75 cm tall and 4.0 cm deep, with overall envelope volume of 96.5 cm^3. Omitting the volume for the actuator module, the total volume is 67.6 cm^3. All dimensions of the two heat sinks are identical.

Due to the design of the actuator module, the heat sink has nozzles that allow for air flow created from both the front side and back side of the electromagnetic actuators. Stereolithograph models of the heat sink were created to vary the nozzle exits' size and the flow rate was measured through the fins using Particle Image Velocimetry (PIV). The velocity profile out the air exiting the end of the fins is shown in Fig. 7. Near the base of the heat sink where the heated chip is placed, there is a large region of entrained flow through the fins. There are four distinct peaks where a jet exit a nozzle through the base. The velocity peaks are similar owing to the pressure across the diaphragm being equal. The synthetic jet is a zero net mass flux device so the time-averaged volume flow rate is zero, but a volume flow rate of 3.92 × 10^{-4} m^3 / s (0.83 CFM) is measured by

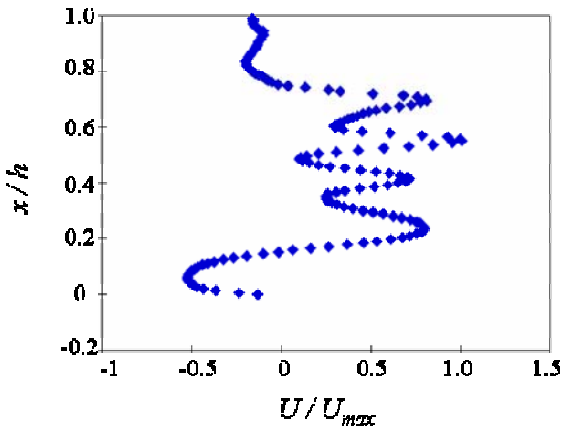

Fig. 7: Normalized velocity profile at fin tip along height ($C_p = 0$), ($U_{max} = 0.82$ m / s)

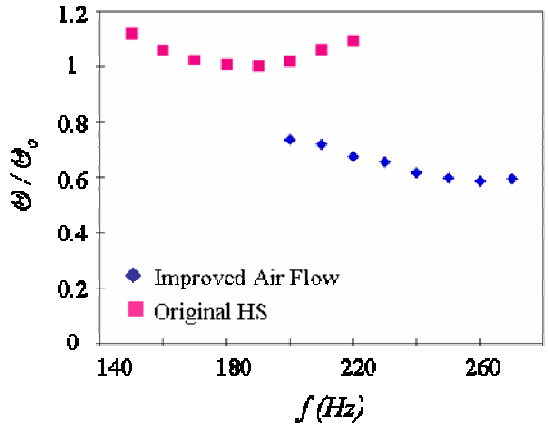

Fig. 8: Comparison of thermal resistance for both heat sinks at various frequencies (actuator voltage held constant in all cases)

calculating only the amount of air exiting the fin channels. Increasing the frequency also affects the volume flow rate obtained with the module. The maximum volume flow rate achieved with this device is $Q_{max} = 5.52 \times 10^{-4}$ m^3/s (1.17 CFM).

After maximizing the volume flow rate, the thermal resistance of the two heat sinks can now be compared to determine the improved performance of the new heat sink design. The difference in thermal resistance between the longitudinal blowing heat sink and the new design is shown in Fig. 8. Identical electromagnetic actuators were used in the two heat sinks, with input voltage held constant. The optimal thermal resistance obtained with the original heat sink is shown. The new heat sink's thermal resistance was normalized with the lowest value obtained with the original case, $\Theta_o = 2.61$ K / W. For both heat sinks, the frequency of the actuator was varied to minimize the thermal resistance. The maximum thermal performance of the new heat sink design shows a thermal resistance of approximately 60% ($\Theta = 1.53$ K / W) of that obtained with the original heat sink design. Determining the convective resistance results in an overall heat transfer coefficient of 45 W / m^{2}K for this device. The original heat sink design achieves a heat transfer coefficient of 22 W / m^{2}K.

The electromagnetic actuator voltage can be varied to increase the amplitude of deflection of the diaphragm, providing a higher flow rate, and therefore higher heat transfer. Fig. 9 shows the thermal resistance for the newly designed heat sink at a constant frequency for various input voltages. As the input voltage is increased, the thermal resistance decreases and levels out as the maximum input to the actuator is reached. The thermal resistance of the new heat sink design is approximately 50% ($\Theta = 1.25$ K / W) of the resistance of the original design. This results in an overall heat transfer coefficient of 55 W / m^{2}K.

An improved heat sink has been designed that no longer uses the U-shape of the previous heat sinks. The heat sink is circular, allowing for greater scalability for inclusion of any number of electromagnetic actuators. This heat sink still uses the improved method air flowing laterally across the short width of the fins as in Fig. 6b. This design also allows for fins to be placed around the entire perimeter of the heat sink. The circular heat sink is ~25% larger in total envelope volume, but decreased Θ / Θ_o by 0.2 ($\Theta = 0.76$ K / W) as shown in Fig. 10.

IV. CONCLUSIONS

A combined systems approach for removing heat from a 3D stacked chip design including both processors and memory has been presented.

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. It is only limited by the ability of the next stage in the heat flow path to remove the heat from the stack edges. In a stack with processors and memories, the memories limit the operation temperature.

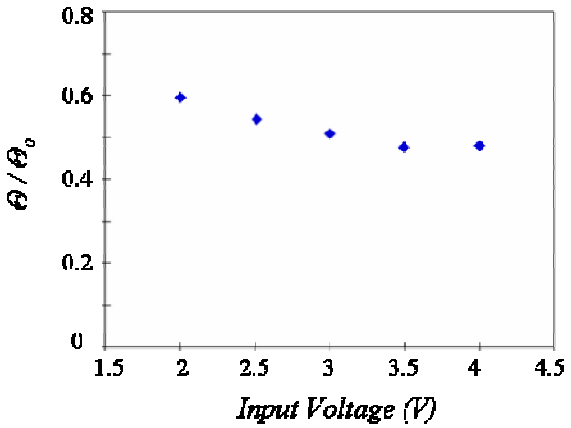

Fig. 9: Thermal resistance for increasing speaker input voltage ($\Theta_o = 2.61$ K / W)

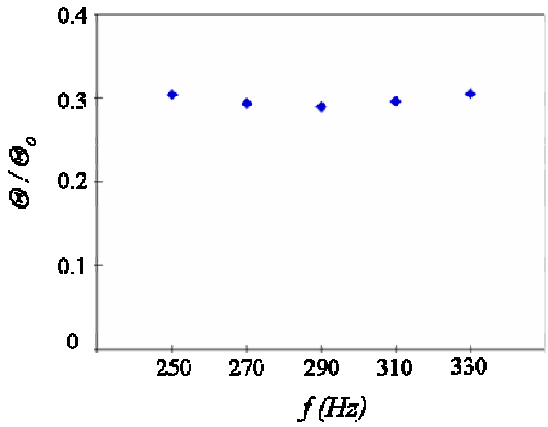

Fig. 10: Normalized thermal resistance of circular heat sink ($\Theta_o = 2.61$ K / W)

A heat sink that uses synthetic jets to induce flow laterally across a heat sink fin moves heat from the stack edges to the ambient air. The design was optimized and compared to a similar heat sink with air moving longitudinally down the length of the fin. This new concept provides greater convective heat transfer for several reasons, including shorter entrainment paths for fresh cool air and induction of even air flow through the entire fin channel length. Velocity profiles were captured with PIV for this case to show the uniformity of the jets. The actuation frequency was varied to determine the effect on volume flow rate. For identical heat sinks, the new air flow concept results in a thermal resistance approximately 60% of the lowest measured with the original design. Finally, an updated design using this air flow concept was able to further reduce the thermal resistance by 20%.

ACKNOWLEDGEMENTS

This work has been funded by DARPA Microsystems Technology Office through the 3D MINT program, Irvine Sensors Corporation, and the Georgia Institute of Technology.

REFERENCES

[1] Gerlach, David W. and Yogendra K. Joshi. "Parametric Thermal Modeling of 3D Stacked Chip Electronics with Interleaved Solid Heat Spreaders." Proceedings of ITHERM 2006, May 30-June 2, 2006, San Diego, CA.

[2] Gerty, D., Mahalingam, R., et al. (2006). "Design and Characterization of a Heat Sink Cooled by an Integrated Synthetic Jet Matrix." Proceedings of ITHERM 2006, May 30-June 2, 2006, San Diego, CA.

[3] International Technology Roadmap for Semiconductors, 2001 Edition, Semiconductors Industries Association Report.

[4] A. Bar-Cohen, "Computer-Related Thermal Packaging at the Millennial Divide", Electronics Cooling, v. 6, n. 1, pp. 32-40, Jan. 2000.

[5] J. Koo, S. Im, L. Jiang, and K. E. Goodson, "Integrated Microchannel Cooling for Three-Dimensional Electronic Circuit Architectures," Journal of Heat Transfer, v. 127, pp. 49-58, Jan. 2005.

[6] 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.

[7] 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.

[8] 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.

[9] 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, v. 52, no. 11, pp. 2395 – 2403, Nov. 2005.

[10] 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.

[11] B. L. Smith, and A. Glezer, "The Formation and Evolution of Synthetic Jets", Physics of Fluids, v. 10, no. 9, pp. 2281-2297, Sep. 1998.

[12] K. Jambunathan, E. Lai, M. A. Moss, B. L. Button, "A Review of Heat Transfer Data for Single Circular Jet Impingement", International Journal of Heat and Fluid Flow, v. 13, pp. 106-115, 1992.

[13] M. R. Thompson, D. L. Denny, W. Z. Black, J. G. Hartley, and A. Glezer, "Cooling of Microelectronic Devices using Synthetic Jet Technology", 11th European Microelectronics Conference, Venice, Italy, pp. 362-366, 1997.

[14] G. B. Russell, "Local- and System-Level Thermal Management of a Single Level Integrated Module (SLIM) using Synthetic Jets", M. S. Thesis, Georgia Institute of Technology, Atlanta, GA, 1999.

[15] R. Mahalingam, and A. Glezer, "An Actively Cooled Heat Sink Integrated with Synthetic Jets", Proceedings of 35th National Heat Transfer Conference, pp. 59-64, June 2001, Anaheim, CA.

[16] R. Mahalingam, and A. Glezer, "Synthetic Jet Based Impingement Cooling Module for Electronic Cooling", Proceedings of the IMAPS 2001 Symposium, pp. 201-206, October 2001, Baltimore, MD.

[17] R. Mahalingam, and A. Glezer, "Forced Air Cooling with Synthetic Jet Ejectors", Proceedings of InterPACK 2005, pp. 199-204, July 2005, San Francisco, CA.

Table 1

Parameter values and results for stack conduction simulations

Run	Thermal resistance from ambient to stack edge 1 K/W	heat transfer coefficient W/K-mm2	FPGA power W	DRAM power per DRAM W	total power in runs W	FPGA max temperature C	DRAM max temperature C
2	25.0	6.84E-04	2.5	0.2	2.7	145	136
3	12.5	1.37E-03	2.5	0.2	2.7	99	90
4	6.5	2.64E-03	2.5	0.2	2.7	80	70
5	4.0	4.27E-03	2.5	0.2	2.7	68	59
6	2.5	6.84E-03	2.5	0.2	2.7	63	53
7	0.5	3.42E-02	2.5	0.2	2.7	56	46
8	25.0	6.84E-04	5	0.2	5.2	251	231
9	12.5	1.37E-03	5	0.2	5.2	159	140
10	6.5	2.64E-03	5	0.2	5.2	119	100
11	4.0	4.27E-03	5	0.2	5.2	97	78
12	2.5	6.84E-03	5	0.2	5.2	85	67
13	0.5	3.42E-02	5	0.2	5.2	71	51
14	25.0	6.84E-04	10	0.2	10.2	453	404
15	12.5	1.37E-03	10	0.2	10.2	276	228
16	6.5	2.64E-03	10	0.2	10.2	190	143
17	4.0	4.27E-03	10	0.2	10.2	155	109
18	2.5	6.84E-03	10	0.2	10.2	133	88
19	0.5	3.42E-02	10	0.2	10.2	104	61