Revolutionary Innovation in System Interconnection: A New Era for the IC

Muhannad S. Bakir, Paragkumar A. Thadesar, Calvin King, Jesal Zaveri,
Hyung Suk Yang, Chaoqi Zhang, Yue Zhang

School of Electrical and Computer Engineering, Georgia Institute of Technology

Abstract: This paper describes novel microscale electrical, optical, and fluidic interconnect networks to address off-chip interconnect challenges in high-performance computing systems as well as to enable 3D heterogeneous integration of CMOS and MEMS/sensors.

I. Introduction

The information revolution has been the most important economic event of the past century, and its most powerful driver has been the silicon integrated circuit (IC). Over the past fifty years, the migration from bipolar-junction-transistor (BJT) to complementary metal-oxide-semiconductor (CMOS) transistor technology combined with transistor scaling has produced exponential benefits in microchip productivity and performance. However, the role that ***interconnects*** play in influencing the performance, power dissipation, and cost of an IC and an electronic system has been increasing in the past two decades. For example, by the 1990s, it was well recognized that on-chip interconnects (i.e., the wires that interconnect the transistors on an IC) present grand challenges to minimizing the power consumption, to increasing the clock frequency, and to reducing the cost of high-performance microprocessors [1-6]. To address this problem, the semiconductor industry responded with a number of innovations for on-chip interconnects that included, for example, the transition from aluminum to copper wires (which began in 1997) to reduce wire resistivity and improve electromigration performance, the use of low-k interlayer dielectrics to reduce parasitic capacitance, the use of repeaters, and the adoption of interconnect-centric architectures among other things.

As the clock frequency of single-core processors continued to increase up to the early 2000s, power dissipation and cooling, along with design complexity, emerged as the principal limiters to continued performance gains from high-performance processors. Multi-core, and eventually many-core, processors with multithreading have emerged as the architecture of choice for improved power efficiency in the era of silicon nanoelectronics [7-8]. While these processors are more power efficient, they are projected to dissipate ~150 W at ~0.7 V supply in the next 15 years according to the International Technology Roadmap for Semiconductors (ITRS) [9]. This presents challenges in cooling and power delivery. Moreover, it is projected that as the number of cores increases, so will the demand for off-chip bandwidth [7-8, 10-12]. Achieving the multi-terabits-per-second of aggregate bandwidth soon to be needed by multicore processors using conventional low-density, high-power, and low-quality off-chip interconnects will not be possible. And thus, ***this introduces a new interconnect problem in the 2000s*** **for silicon nanoelectronics** – ***the off-chip interconnect problem***: we have essentially reached the limits to what we can do with current silicon ancillary (support) technologies in the areas of off-chip signaling, cooling, and power delivery - and the solutions to which will have to be revolutionary [13-16]. While three-dimensional (3D) stacking of nanoelectronics

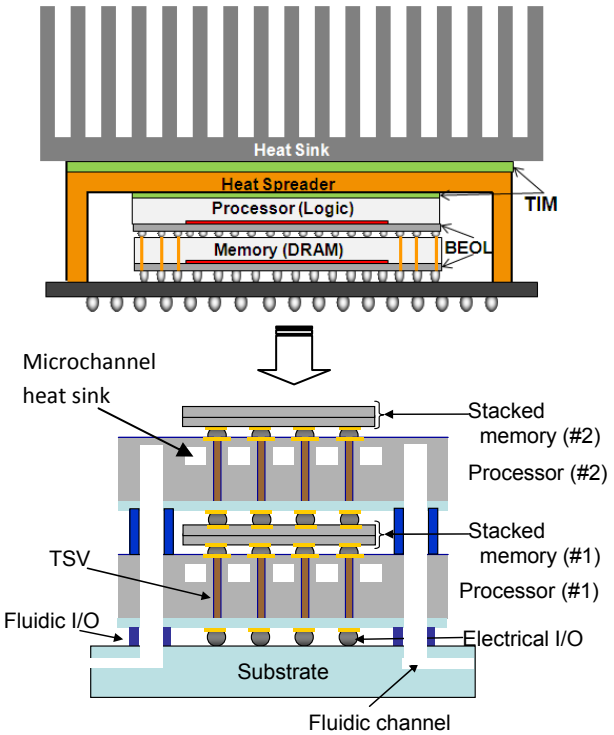

Figure 1: Schematic illustration of the objective of the research. The technology and models developed will enable stacking of multiple processors and memory chips.

Reliability, Packaging, Testing, and Characterization of MEMS/MOEMS and Nanodevices X,
edited by Sonia Garcia-Blanco, Rajeshuni Ramesham, Proc. of SPIE Vol. 7928, 792803

can greatly enhance the signaling between ICs (larger bandwidth, lower latency, and lower energy per bit), it also presents a number of challenges and the solutions for which do not exist. Aside from issues relating to manufacturing, power delivery [7, 15, 17] and cooling [9, 18-19] of a stack of logic and memory chips presents many challenges and *will be the key limiters to 3D IC stacking*.

Historically, in order to maintain constant junction temperature with increasing power dissipation, the size of the heat sink used to cool a microprocessor has been increasing (along with improved heat spreaders and thermal interface material (TIM)), and thus imposing limits on system size, chip packing efficiency, and interconnect length between chips. The cooling of hot-spots (power density up to 500 W/cm^2) greatly exacerbates the complexity of cooling and it is becoming increasing challenging to meet thermal performance needs of silicon nanoelectronics. Challenges in cooling become magnified in 3D high-performance nanoelectronics because of increased power density. Moreover, the cooling technology used has large implications on the electrical system design. As shown in Figure 1, if an air-cooled heat sink is used in a processor-memory stack, then stacking the processor on the top level (i.e., closest to the heat sink) provides the best thermal solution for the processor. However, using this approach makes it challenging to deliver power to the processor through the memory chip because of the large number of power/ground TSVs that must be routed through the memory chips, which can reduce the density of the memory chip, for example. Moreover, the number of memory chips that can be stacked will also be small because of the processor power dissipation and thermal resistance path for the memory chips; thermal crosstalk between the two chips can also be a challenge. Recently, the use of liquid cooling within a 3D stack of ICs has been proposed by a number of researchers [19-23]. This cooling approach has the benefit of enabling large heat flux removal from *within the stack* as well as being a microscale cooling solution (same order of magnitude as size of electrical chip I/Os). The technologies proposed in this paper leverage this approach and form the basis for the system under consideration, as shown in Figure 1. Through such integration, we can now move the processor to the lower level and place the memory chips on top of the processor and in close proximity to the microchannel heat sink in the processors. It is envisioned that multiple processors and memory chips can be stacked, as shown in Figure 1. Challenges in designing the power delivery network for low supply noise are reported in [17, 23]

But, there is more to the story. With the advent of next generation electronics such as smartphones, patient-side health care systems and motion-enabled video gaming systems, integration of sensors/MEMS with CMOS ICs is becoming ever more relevant and important. The use of 3D system integration provides an ideal platform for the heterogeneous integration of those two critical components [24]. This is because almost all MEMS chips require an interaction with a CMOS IC for processing raw signals (signal conditioning, amplification, processing, actuation, etc.) from individual MEMS devices; high density and low parasitic interconnects common to most 3D integration schemes mean that even raw signals from a large area-array of devices and sensors can be processed simultaneously in parallel using low parasitic interconnects. Such capabilities are critical for applications that require simultaneous capturing of the data from large and dense array of sensors and devices. Also, vertical integration through chip stacking means that microsystems with smaller form factor can be achieved – an increasingly important feature as we enter the "more than Moore" era.

The objective of this paper is to provide an overview of some of the interconnect technologies being developed by our lab to address the grand challenges described above. With respect to interconnect technologies for computing systems, electrical, optical, and microfluidic ('trimodal') interconnect networks are demonstrated both at the IC level and at the package level using a silicon interposer. With respects to photonics, in this paper, we focus on optical interconnects operating at 850nm. An overview of these technologies is presented in Section II. This paper also reports novel interconnect technologies to enable the heterogeneous integration of a large array of MEMS/sensors with CMOS electronics in Section III. Finally, Section IV is the conclusion.

II. Interconnect Opportunities for Computing Systems

A. Overview of a Trimodal Interconnect Network

A schematic illustration of a trimodal based interconnect network is shown in Figure 2. At the die level, memory chips are stacked above a chip multiprocessor with an integrated microliquid cooled heat sink. Of course, in order to implement a microliquid cooled heat sink, a number of innovations must occur, including innovation in process integration of microchannels with CMOS ICs, delivery of a liquid coolant to each chip in the stack (i.e., within the stack), and full compatibility with electrical TSVs and chip I/Os. A discussion of how the TSVs and microliquid heat sink are

integrated at the IC level has been reported previously [22, 24-25]. In this paper, with respect to cooling, we focus on novel C4-based microfluidic interconnects that can deliver a coolant to the embedded microliquid heat sink in the 3D stack. This is discussed in Section II.B.

The substrate on which the 3D ICs are stacked is of critical importance, and we propose the use of a silicon interposer with integrated electrical, optical, and microfluidic interconnect networks. The optical interconnects would not only enable within interposer optical communication, but incorporating optical TSVs within the interposer enables off-package optical interconnection. Depending on the size and design of the optical TSVs, they behave either as an optical-waveguide or simply as an optical aperture. The silicon interposer technology under consideration also features copper-based electrical TSVs that contain a relatively thick polymer liner, which helps in CTE-mismatch related challenges as well as provide opportunity to form, for example, coaxial-like TSVs with improved high-frequency electrical characteristics [26]. Within interposer integrated microfluidic interconnections are also developed to enable the routing of a coolant to all ICs bonded on the interposer. The fluidic inlets/outlets in the silicon interposer can either be interconnected to microfluidic channels on the printed-wiring board (as shown in the figure) or interconnected to free-standing fluidic 'tubes' from a local pump. Key aspects of the silicon interposer technology are described in Section II.C.

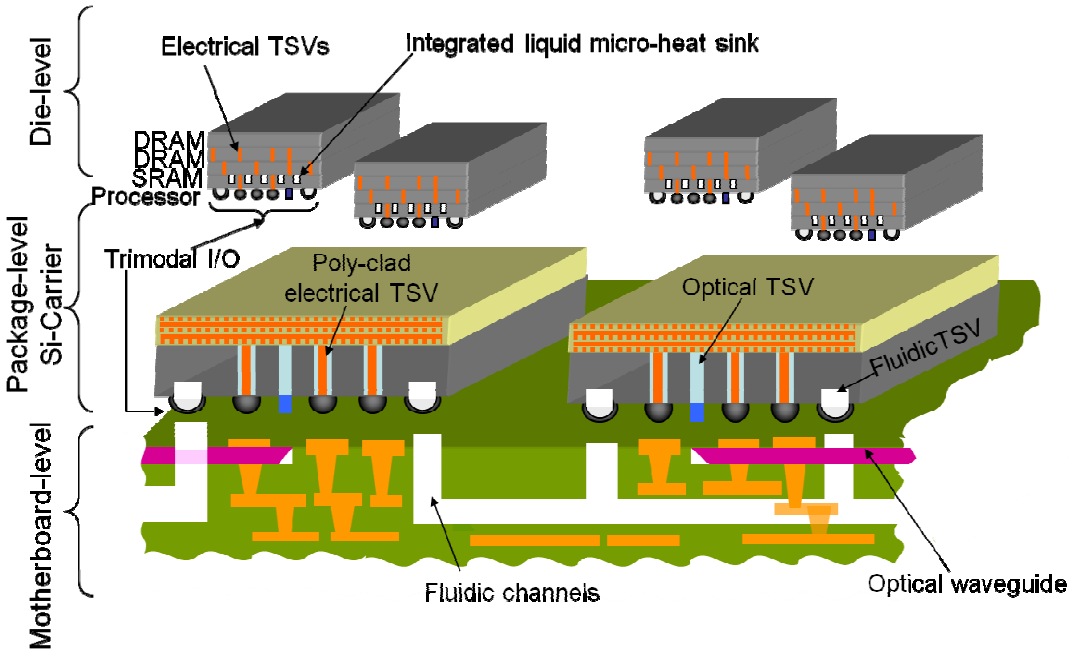

Figure 2: Schematic illustration of a trimodal based interconnect network.

B. 3D Stacking Using Microfluidic Chip I/O Interconnects

In the 3D cooling scheme outlined in Figure 1, microscale fluidic interconnections between the chip and substrate (and between strata when needed) is enabled by microfluidic chip I/O interconnects using solder technology [27]. Development of a microscale solder-based fluidic interconnect technology allows simultaneous batch fabrication of electrical and fluidic I/Os, making electrical and fluidic I/O integration seamless as well as making chip reworkability possible. Figure 3 shows SEM images of ring-like solder-based fluidic I/O interconnects adjacent to electrical C4 bumps. The density of the electrical I/Os shown is ~1600/cm^2 (pitch of 240μm, although smaller pitches are certainly possible). The adjacent fluidic I/Os have a pitch of

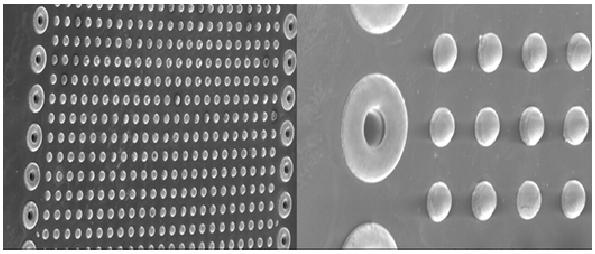

Figure 3: SEM image of C4 electrical and fluidic I/Os after reflow.

480μm. The fabrication of electrical and fluidic I/Os yields features with good uniformity and a standard deviation in feature size of <1μm.

In an alternate approach, 'air-gap C4 I/Os," which are essentially solder bumps with an air-filled depression, can also be fabricated [27]. X-ray image of an array of air-gap C4 I/Os is shown in Figure 4. One key advantage of the air-gap fluidic I/O technology is that it is transparent to the flip-chip assembly process. Furthermore, the air-gap C4 technology is compatible with the use of no-flow underfill in applications for which it is required. When the air-gap C4 fluidic I/Os are assembled to the substrate (Figure 5), the solder wets ring-pattern copper pads on the substrate. Upon contact, the air-gap C4 fluidic I/Os transform in shape forming a solder pipe-like fluidic I/O interconnects. This is verified by x-ray images taken of the I/Os after assembly (Figure 5). The final pipe-like pattern is of course critical to enable coolant passage into the chip-level microliquid cooled heat sink.

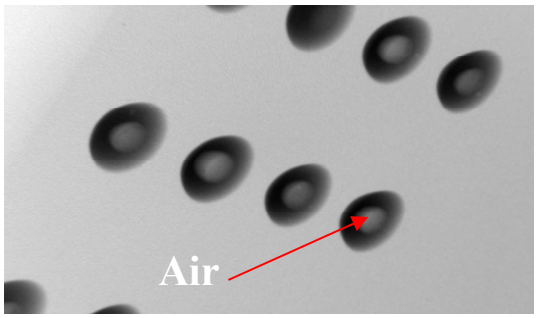

Figure 4: X-ray image of air-gap C4 fluidic I/Os after solder reflow.

To test the C4 fluidic I/Os, two chips were assembled to a silicon substrate with fluidic vias (Figure 6). A mechanical pump and fluid inlet pipe were attached to the bottom side of the silicon substrate in the 3D stack and used to pass fluid through the chip stack. Flow-rates up to 100 ml/min were tested; the pumped fluid passed through the stack as depicted in the figure verifying functionality of the novel microfluidic chip I/Os.

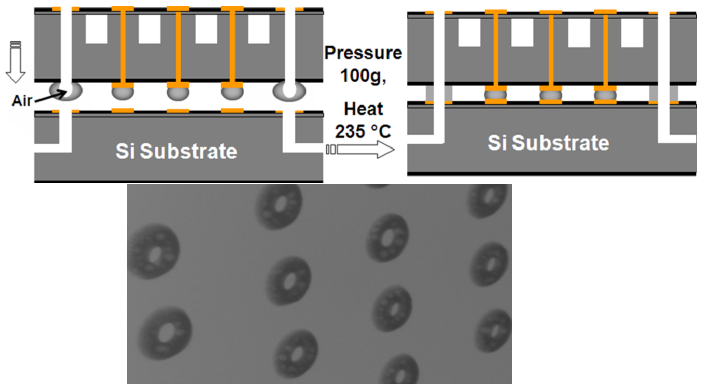

Figure 5: Assembly of solder-based fluidic I/O interconnects. X-ray image of air-gap C4 fluidic I/Os after assembly is also shown.

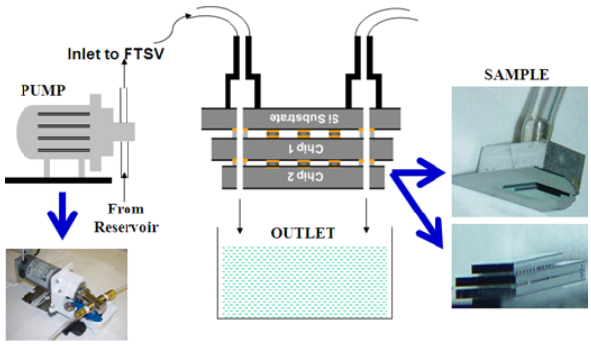

Figure 6: Experimental setup for testing the C4 fluidic I/Os.

C. Silicon Interposer with Electrical, Optical, and Microfluidic Interconnects

Silicon interposer technology, commonly referred to as 2.5D integration technology, has recently emerged as an attractive method of increasing chip-to-chip interconnection density. For example, instead of interconnecting two packaged chips using large-pitch and lossy motherboard-level interconnects, a silicon interposer provides interconnect pitches comparable to on-chip global wiring level pitches thereby affording significant off-chip bandwidth density. In this work, a silicon interposer with electrical, optical and microfluidic TSVs has been developed to not only support these high density electrical interconnects but to also enable off-package optical interconnection and liquid coolant delivery to the stacked ICs (Figure 2). The fabrication process for the silicon interposer with electrical, optical, and fluidic TSVs is shown in Figure 7. The process begins with the formation of vias in the silicon wafer followed by the formation of a silicon-dioxide mesh layer [28] at the base of the vias (Figure 7a). The mesh layer serves multiple functions: 1) it simplifies the formation of the seed layer at the base of the via, which is necessary for bottom-up plating,

and 2) it improves the polymer filling process of the vias (which is the subsequent process). An optical image of the silicon-dioxide mesh layer following silicon via formation is shown in Figure 7a. Next, a photodefinable polymer film is spin coated on the wafer to fill the vias (Figure 7b); an optical image of a 400-µm thick silicon interposer following this process step is also shown. Next, vias are photodefined through the polymer filled vias such that a polymer liner remains on the sidewalls of the vias, as shown in Figure 7c. Cross-sectional and top view images are shown in Figure 7c; the liner thickness in this case is ~20µm. The polymer liner is formed such that it does not block the silicon-dioxide mesh pattern. Finally, a copper seed layer is sputter coated on the back side of the interposer (on the mesh side) followed by the copper electrodeposition to pinch-off the cavities within the mesh [28]. Once this is complete, the vias containing the polymer liner are copper electroplated, after which CMP is used to planarize the vias. An optical image of the polymer-clad copper TSVs following this process step is shown in Figure 7d. The losses, which are minimal, in the optical TSVs were characterized; Figure 8 is an optical image of a single optical via being probed with an optical fiber. A photodetector was mounted on the opposing side of the interposer to measure the transmission losses.

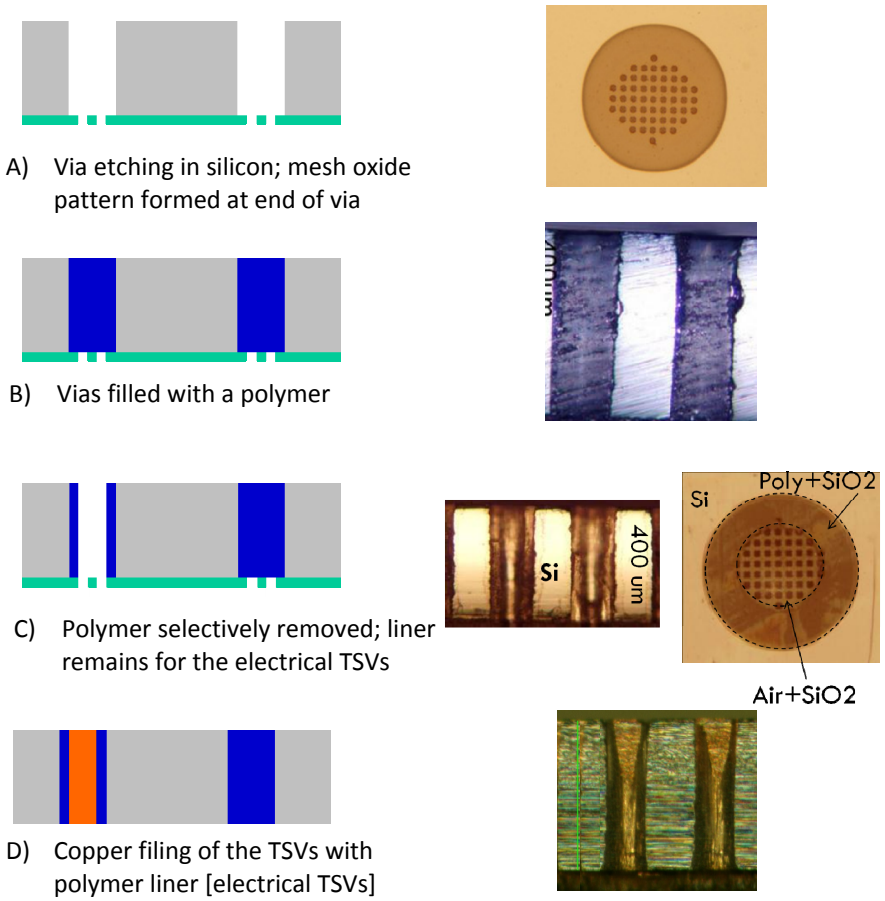

Figure 7: Process flow for the fabrication of a silicon interposer with polymer-clad copper TSVs with optical TSVs. TSVs are approximately 400 µm deep.

Figure 9 illustrates schematic and SEM images of a silicon interposer containing embedded microfluidic channels. The fluidic channels are used to route a coolant into and out of flip-chip bonded ICs (single or 3D) on the interposer. This greatly eliminates the complexity in delivering a liquid coolant to a large number of microprocessors. The hydraulic diameter of the microfluidic channels is chosen such that they represent a small fraction of the pressure drop in the cooling solution (i.e., they should be a small fraction of the on-chip pressure drop). This helps to minimize the overall pressure drop in the system and limit pump power.

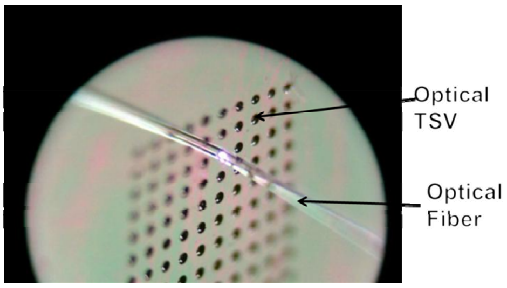

Figure 8: Test setup for the optical TSVs

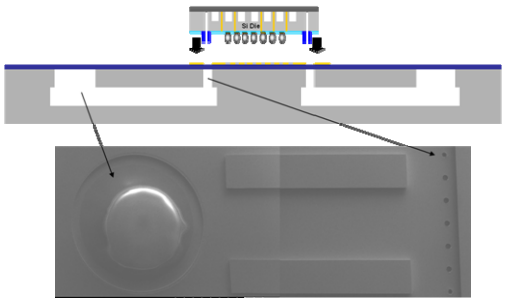

Figure 9: Schematic and SEM image of a silicon interposer with microfluidic interconnects.

III. High Density Mechanically Flexible Chip I/O Interconnects

Three-dimensional (3D) heterogeneous integration of CMOS ICs and MEMS (or sensors) chips offers an enormous design freedom for microsystem designers. With the growing number of MEMS and innovative ways to fabricate such devices [2], it is becoming increasingly attractive to find an integration solution that is not specific to an application or a MEMS device, but an integration method that can become a generic platform in which arbitrary MEMS devices can be integrated with a state-of-the-art CMOS chip, i.e. enabling the marriage of CMOS and MEMS.

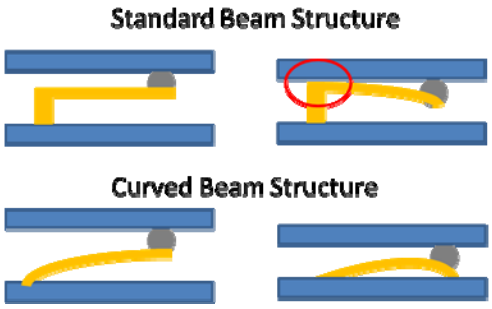

Figure 10: Conventional beam (top) and curved beam (bottom) under loading. Conventional beam design has a limited vertical range of motion due to the vertical post.

3D heterogeneous integration has the potential to be such method; what is needed though, is a 3D ancillary technology, namely an interconnect technology that can provide both high performance electrical and mechanical characteristics between the two dissimilar chips in the stack.

The use of mechanically flexible interconnects (MFIs) can offer a number of benefits in such integration [29]. Flexible interconnects' ability to reduce thermomechanical stress has been demonstrated previously, and if the technology is used correctly between MEMS and the package substrate, it can significantly reduce the thermomechanical stress from propagating to MEMS devices. If needed, flexible interconnects can also be used between CMOS ICs and the package substrate to reduce thermomechanical stress in CMOS ICs as well. Beyond this, MFIs may be used to interface between a large silicon interposer and the motherboard, i.e., enable the bonding of the silicon interposer on the motherboard shown in Figure 2. MFIs can also allow assembly of chips that may not have a perfectly planar surface by applying sufficient load during the assembly, and MFIs can even make contact to surfaces inside a cavity or surfaces on top of a tall feature.

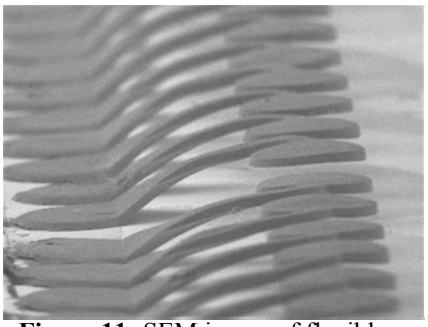

Figure 11: SEM image of flexible interconnects with large range of vertical motion.

Flexible interconnects also have other potential benefits; MFIs can be used to make low-force and low-resistance temporary electrical connections with a bare-die, enabling at-speed testing of chips before they are bonded to the final substrate. MFIs can also be used to make a disposable sensor system where only the potentially contaminated sensor chip is replaced while the "expensive" CMOS IC is reused [30].

The use of MFIs has also been recently proposed in the construction of a nanophotonic interconnected 'macrochip' [11]. In such systems where the precise alignment between chips is critical, flexible interconnects represent an elegant solution for making electrical interconnection; its lateral and vertical compliance allow the alignment to be adjusted and fined tuned in all six degrees-of-freedom even after chips have been brought close together for alignment and interconnections have been made.

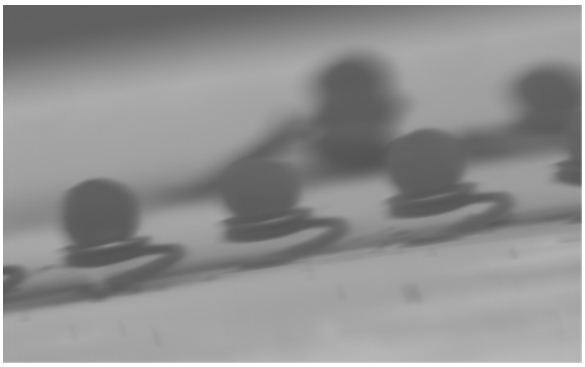

Figure 12: MFIs after solder reflow.

However, in order to use MFIs for such purposes in addition to using MFIs to reduce thermomechanical stress, it is essential that MFIs can take advantage of most of the vertical standoff height without being damaged or going through a significant plastic deformation, which could reduce the available standoff height and degrade the capability of MFIs.

In order to allow 100% of the stand-off height to be utilized, it is necessary to diverge from conventional cantilever design, as shown in Figure 10. With such a design, the range-of-movement would be restricted to the height of material

deposited on the tip of the beam. By having a curved beam design, as shown in Figure 11, this problem can be avoided and it is the design used for the MFIs in this work. This design is especially critical if the interconnect pitch is to be scaled down while the vertical standoff height is kept constant. When permanent interconnection is needed, solder may be deposited on the tip of the leads, as shown in Figure 12.

In order to determine the extent of plastic deformation, a single 12µm thick 200µm x 100µm MFI was indented multiple times. The maximum vertical displacement achieved was 4µm, which is limit of the equipment used in the indentation. Results show no sign of plastic deformation as both the loading and unloading curves were an identical linear curve, as shown in Figure 13. Even after the twentieth indentation, the loading and unloading curves were matched and identical to the first indentation. In another set of measurements, the full vertical range of motion of MFIs with a 20µm gap has been measured [29].

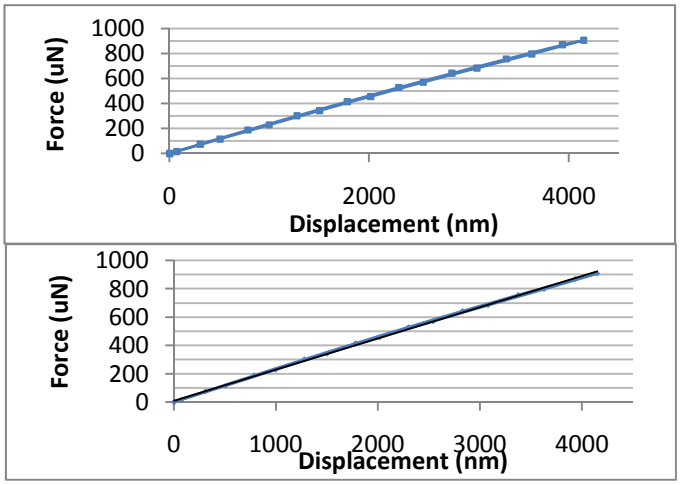

Figure 13: Force vs. Displacement of 12µm thick MFI a) first indentation b) after twenty indentations. It has a linear loading and unloading profile that is matched meaning that no plastic deformation has occurred.

IV. Conclusion

This paper presents an overview of various interconnect technologies that collectively can address grand challenges in future high performance computing systems as well as enable 3D heterogeneous integrated systems. A trimodal based interconnect network is proposed for silicon interposers to enable power delivery, cooling, and off-package optical communication. This paper also describes the use of novel mechanically flexible interconnects to enhance system connectivity.

Acknowledgment

This work has been carried out in part under a Semiconductor Research Corporation (SRC) GRC-Texas Instruments Custom Funded Project 13066P8, National Science Foundation Grant No. 0701560, Microelectronics Advanced Research Corporation (MARCO), its participating companies, and DARPA under contract 2003-IT-674, and an award from the Department of Defense.

References:

[1] M. T. Bohr, "Interconnect scaling-the real limiter to high performance ULSI," in *Proc. IEEE Int. Electron Devices Meeting*, 1995, pp. 241-244.

[2] M. Bohr. (2007) A 30 Year Retrospective on Dennard's MOSFET Scaling Paper. *IEEE Solid-State Circuits Society Newsletter*. 11-13.

[3] J. D. Meindl, "Beyond Moore's Law: the interconnect era," *IEEE Computing in Science & Engineering,* vol. 5, pp. 20-24, 2003.

[4] J. A. Davis*, et al.*, "Interconnect limits on gigascale integration (GSI) in the 21st century," *Proceedings of the IEEE,* vol. 89, pp. 305-324, 2001.

[5] J. A. Davis and J. D. Meindl, *Interconnect Technology and Design for Gigascale Integration*: Kluwer Academic Publishers, 2003.

[6] R. Ho*, et al.*, "The future of wires," *Proceedings of the IEEE,* vol. 89, pp. 490-504, 2001.

[7] P. G. Emma and E. Kursun, "Is 3D chip technology the next growth engine for performance improvement?," *IBM J. Res. & Dev.,* vol. 52, pp. 541-552, Nov 2008.

[8] B. Casper*, et al.*, "Future microprocessor interfaces: analysis, design and optimization," in *Proc. IEEE Custom Integrated Circuits Conference*, 2007, pp. 479-486.

[9] S. Borkar, *et al.*, "Microprocessors in the era of terascale integration," in *Proc. IEEE Design, Automation & Test in Europe Conference*, 2007, pp. 1-6.
[10] G. Balamurugan, *et al.*, "A scalable 5-15 Gbps, 14-75 mW low-power I/O transceiver in 65 nm CMOS," *IEEE J. Solid-State Circuits,* vol. 43, pp. 1010-1019, 2008.
[11] A. V. Krishnamoorthy, *et al.*, "Computer Systems Based on Silicon Photonic Interconnects," *Proceedings of the IEEE,* vol. 97, pp. 1337-1361, 2009.
[12] J. C. Eble, "Off-chip signaling," in *Integrated Interconnect Technologies for 3D Nanoelectronic Systems*, M. S. Bakir and J. D. Meindl, Eds., ed Boston: Artech House, 2009.
[13] M. S. Bakir, *et al.*, "Revolutionary nanosilicon ancillary technologies for ultimate-performance gigascale systems," in *Proc. IEEE Custom Integrated Circuits Conf.*, 2007, pp. 421-428.
[14] M. S. Bakir and J. D. Meindl, *Integrated Interconnect Technologies for 3D Nanoelectronic Systems*. Boston: Artech House, 2009.
[15] J. U. Knickerbocker, *et al.*, "Three-dimensional silicon integration," *IBM J. Res. & Dev.,* vol. 52, pp. 553-569, Nov 2008.
[16] L. Polka, *et al.*, "Package technology to address the memory bandwidth challenge for tera-scale computing," *Intel Technol. J.,* vol. 11, pp. 197-205, 2007.
[17] G. Huang, *et al.*, "Power delivery for 3D chip stacks: physical modeling and design implication," in *Proc. IEEE Conf. on Electrical Performance of Electronic Packaging*, 2007, pp. 205-208.
[18] S. M. Sri-Jayantha, *et al.*, "Thermomechanical modeling of 3D electronic packages," *IBM J. Res. & Dev.,* vol. 52, pp. 623-634, Nov 2008.
[19] T. Brunschwiler, *et al.*, "Interlayer cooling potential in vertically integrated packages," *Microsystem Technologies,* vol. 15, pp. 57-74, 2009.
[20] Y. Aibin, *et al.*, "Fabrication of Silicon Carriers With TSV Electrical Interconnections and Embedded Thermal Solutions for High Power 3-D Packages," *IEEE Trans Components and Packaging Technol.,* vol. 32, pp. 566-571, 2009.
[21] A. Bar-Cohen, *et al.*, "Direct liquid thermal management of 3D chip stacks," in *Proc. Electronics Packaging Technol. Conf.*, 2009, pp. 119-130.
[22] D. Sekar, *et al.*, "A 3D-IC Technology with Integrated Microchannel Cooling," in *Proc. Int. Interconnect Technol. Conf.*, 2008, pp. 13-15.
[23] M. S. Bakir, *et al.*, "3D heterogeneous integrated systems: liquid cooling, power delivery, and implementation," in *Proc. IEEE Custom Integrated Circuits Conf.*, 2008.
[24] D. Bing, *et al.*, "Integrated Microfluidic Cooling and Interconnects for 2D and 3D Chips," *IEEE Trans. Advanced Packaging,* vol. 33, pp. 79-87, 2010.
[25] J. Zaveri, *et al.*, "Wafer level batch fabrication of silicon microchannel heat sink and electrical through-silicon vias for 3D ICs," in *Int. Symp. on Microelectronics*, 2009, pp. 579-584.
[26] H. Soon Wee, *et al.*, "High RF performance TSV silicon carrier for high frequency application," in *Proc. Electronic Components & Technol. Conf.*, 2008, pp. 1946-1952.
[27] C. R. King, *et al.*, "Electrical and fluidic C4 interconnections for inter-layer liquid cooling of 3D ICs," in *Proc. Electronic Components & Technol. Conf.*, 2010, pp. 1674-1681.
[28] J.-H. Lai, *et al.*, "A 'mesh' seed layer for improved through-silicon-via fabrication " *J. Micromech. Microeng,* vol. 20, 2010.
[29] H. S. Yang and M. S. Bakir, "3D integration of CMOS and MEMS using mechanically flexible interconnects (MFI) and through silicon vias (TSV)," in *Proc. Electronic Components & Technol. Conf.*, 2010, pp. 822-828.
[30] R. Ravindran, *et al.*, "Biochemical sensing with an arrayed silicon nanowire platform," in *Proc. Electronic Components & Technol. Conf.*, 2010, pp. 1015-1020.