2.5D TSV enablement of a multiple processor SiP

By Deborah Patterson, Mike Kelly, Rick Reed [Amkor Technology, Inc.]; Steve Eplett [Open-Silicon, Inc.] and Zafer Kutlu, Ramakanth Alapati [GLOBALFOUNDRIES][1]

The movement to 2.5D TSV package architecture has been a methodical undertaking. The skyrocketing cost of next-generation process node adoption has accelerated a shift in mindset away from the routine acceptance of transistor scaling for complex, new system-on-chip (SoC) designs and toward alternative options that maintain competitiveness. In addition, 2.5D TSV system-in-package (SiP) designs are also seeing interest in end use applications that may require the integration of components that are not yet defined well enough for SoC implementation. Interposer-based construction provides the speed, bandwidth, power efficiency and modularity that traditional multi-chip packages cannot implement. Cost reduction, faster time-to-market, yield improvement, component flexibility and re-use, and reduced program risk are additional expected benefits.

Industry drivers for 2.5D through-silicon vias (TSVs)

System-on-chip design costs are forecast around $300M as manufacturing moves into 16/14nm FinFET technologies. Although not as expensive, the adoption of updated tools to work with more sophisticated versions of older process nodes also contributes significantly to cost. Prudent use of existing IP, software, design and manufacturing tools, and other infrastructure can still run $20M-$50M per SoC design. The immense costs and longer timelines required to design and manufacture leading-edge SoCs have resulted in two significant changes: 1) Only the largest applications and markets are now targeted in order to recoup the considerable up-front investment; and 2) The number of companies with resources to design next-generation technology nodes are rapidly shrinking.

For next-generation SoCs, lifetime revenue requirements now demand multi-billions of dollars in return to be economically feasible. **Figure 1** illustrates how "design costs for advanced SoCs have more than doubled, on average, for each node during the last 10 years [2]."

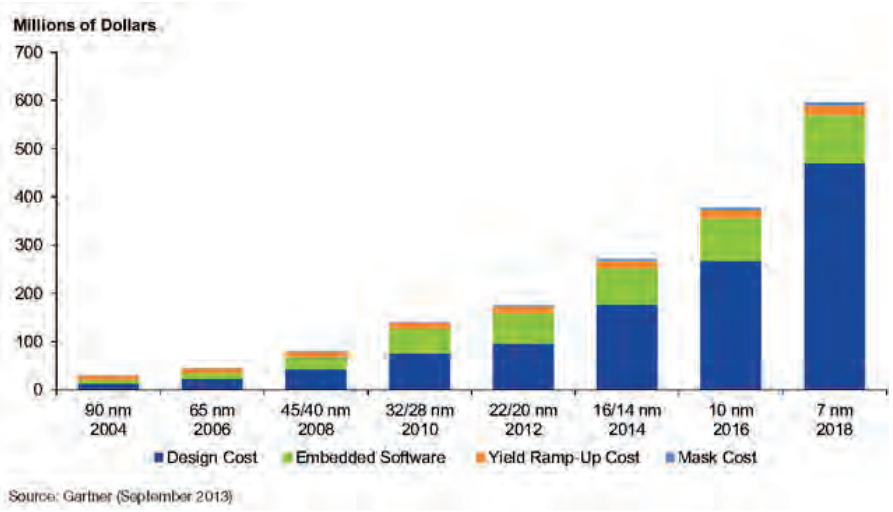

Figure 1: Estimated design costs for next-generation SoCs. SOURCE: Gartner (September 2013)

Because of these trends, more cost effective alternatives such as 2.5D TSV SiP solutions are being evaluated prior to migrating to the next silicon process node. Advanced 2.5D SiP architectures can support leading edge system performance requirements while reducing time-to-market and lowering total cost-of-ownership as compared to new, packaged SoC platforms.

2.5D TSV architecture represents a viable approach to single package system design and allows for benefits such as: 1) Power and performance improvements enabling new applications; 2) Smaller form factor; 3) Silicon layer count reduction for reduced cost and cycle time; 4) Employment of die utilizing the best technology node at the best price and performance (keeping foundry capex down); 5) Die partitioning and optimization for memory, analog, performance, power management, etc.; 6) Higher effective die yields obtained through die recovery; 7) Integration of memory technologies with clear downstream benefits for bandwidth and power; serving as replacements to eDRAM/eFlash; 8) Accelerated time-to-market; and 9) Risk reduction in the schedule.

Several of the benefits listed above offset the costs of an interposer and the more complex 2.5D assembly for an increasing range of applications. In addition, for customers with proven IP that is typically ported to each new process node, IP migration

could be deferred. Third party design houses may choose to offer core IP as a shared resource across multiple users, amortizing development costs or applying the IP across multiple markets. The ability to extend core IP through re-use, or leverage it across multiple platforms, offsets the full NRE costs that individual companies would normally incur, thereby introducing a new IP usage model.

Package design and functional results

The 2.5D TSV SiP integrated two 28nm SoC ARM dual core Cortex™-A9 processors to extend processor function and reduce product risk. As with repartitioned monolithic die, multiple independent ICs are similarly reliant on a silicon node or package interconnect density that can either preserve or limit inter-chip communication.

The silicon interposer was designed to provide a 16GB/sec full duplex data rate between these two devices. A 65nm process technology allowed for finer grain and lower power connectivity, reducing both form factor and overall power budgets. **Figure 2** illustrates the two side-by-side processors connected through the single silicon interposer. A view of the copper pillar bump array on the bottom side of the processors is also shown. **Table 1** defines basic structural elements of the package.

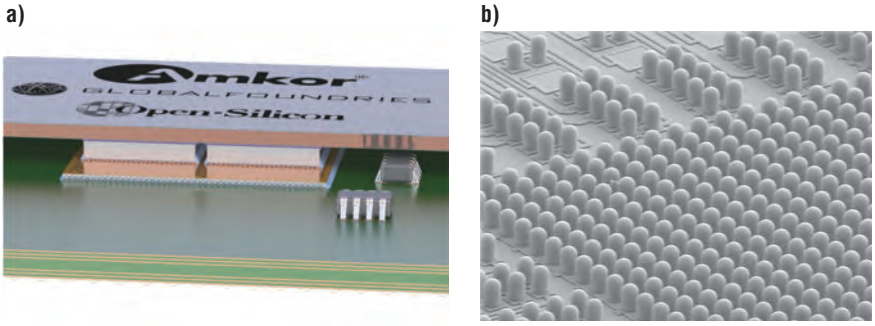

Figure 2: a) (left) Illustration of the two ARM Cortex™-A9 processors connected through b) (right) a silicon interposer by copper pillar micro-bumps. This SiP was designed for low power (although the lid could be employed as a heat spreader if thermal conditions necessitated this use).

Package					
Body Size	27.0 mm x 27.0 mm	**Max. Thickness**	2.78 mm	**Ball Count**	671
Processor					
Die Size	4.1 mm x 4.5 mm	**Thickness**	0.600 mm		
Bump Pitch (Cu Pillar)	40 µm	**Bump Height (Cu Pillar)**	40 µm		
Interposer					
Die Size	10.8 mm x 7.4 mm	**Top Side Pad**	NiAu	**Bottom Side Solder Bump Height**	80 µm
Thickness	~100 µm	**Top Side Pad Pitch**	40 µm	**Bottom Side Solder Bump Pitch**	170 µm

Table 1: 2.5D SiP package structure.

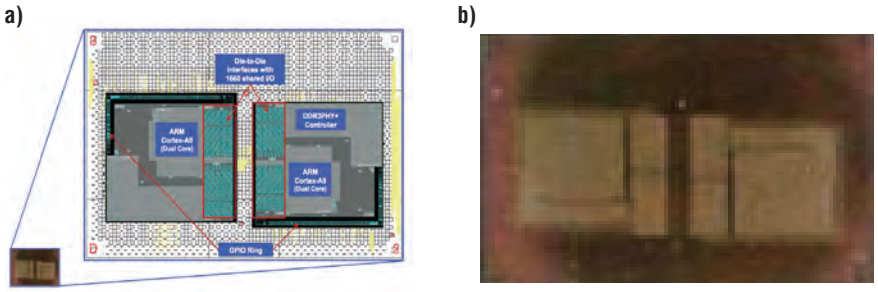

Figure 3: a) Outline of two ARM Cortex™-A9 dual core processors, connected through a 65nm silicon interposer; b) The interposer (shown at the lower left of 3a) with its TSVs and topside metallization.

Interposer design

The most important consideration in joining the 28nm processors to the 65nm interposer was not just routing signals through copper pillar flip-chip interconnects and tight circuit geometries, although these structural elements do enable high performance electrical compliance. The fundamental understanding was that the single device architecture had to be modified to take full advantage of a 2.5D structure. From a functional level, the demonstration showcased how a multiple processor layout could be optimized in ways that would keep die-to-die I/O power low, reduce die-to-die I/O size (while maintaining adequate ESD protection), and optimize functional testing to highlight the competitive merit of multiple die systems.

Facilitating a low power design meant creating a pathway that would support the data rate across both processors through the silicon interposer. The interposer was comprised of four copper metal layers and an aluminum final metal layer with TSVs. The interposer construction used a 65nm back-end-of-line (BEOL) integration process flow at the foundry. The copper interconnection between the chips was designed with minimum 1µm line/space geometry.

The interposer connected 1,660 signals between the two processors. The spacing preserved a die-to-die distance of 0.5mm, though the interface was designed to support up to a 4mm maximum signal length between the die. After leaving the foundry, the interposer was thinned to 100µm to expose the 10µm diameter TSVs. **Figure 3** shows the layout of the two processors on the silicon interposer with a photo of the interposer displaying its topside pad metallization at the lower left.

A single redistribution layer (RDL) on the backside of the interposer was used to connect the TSVs to lead-free solder bumps at a 170µm array pitch. These lead-free solder bumps joined the interposer to the 27mm x 27mm high-density build-up (HDBU) package substrate. The substrate comprised a 4-2-4 layer stack and 0.4mm thick core. **Figure 4** shows a cross section of the die/interposer/substrate construction and representative copper pillar and solder bump joints.

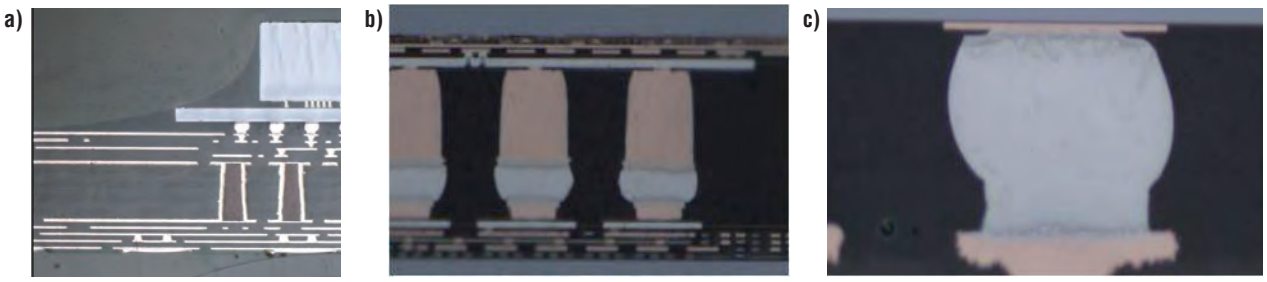

Figure 4: a) Cross section of the 2.5D TSV SiP, b) representative Cu pillar micro-bumps at 40µm pitch, and c) representative SnAg flip-chip solder bump at 170µm pitch.

Concurrent IC and package design produces optimized systems

Whereas the interposer employs 65nm node technology, the processors were manufactured using a 28nm super low power (28nm-SLP) process. The goal was to design the processor as small as possible, reduce the power, and ensure that the distances between the I/O drivers were short.

Driving power down, and speed up. Power is costly to manage and limits performance. In mobile applications, a shutdown of the processor will occur if maximum skin temperature is reached. In computing, the power required to drive external memory reduces graphics and processor performance. In networking, over 50% of data center cost is due to cooling. At the device level, a junction temperature (T_j) in excess of 80°C will increase the DRAM refresh rate, increasing both power and heat. Conversely, high-performance memory that uses massively parallel I/Os keeps power down, but requires very dense routing (that can also be provided in the form of interposer support). Therefore, demonstrating low signal I/O power was a fundamental requirement. Fortunately, it is also a feature that favors the 2.5D interposer system architecture, yielding performance advantages.

Besides the low power offering, the intra-die processor interconnect supports a very wide, higher speed data bus. A data bus is like a highway that connects the processor to the memory. The wider the bus (ex: 64-bit as compared to 32-bit), the more data that can be moved down the highway at the same time, allowing for faster access. To create a high-speed data bus between the two processors, copper pillar micro-bumps were positioned underneath the standard cell logic region of the processor interconnect at a 40µm pitch.

The custom placement of the copper pillar micro-bumps produced several benefits. First, the system designer placed the output (TX/transmit) and input (RX/receive) cells within the logic region next to each other. The copper pillar micro-bumps were then positioned over each of these adjacent cells. This placement allowed for an efficient way to test the processor interconnects – not easy with over 1600 fine-pitch bumped I/Os. Because most wafer probe design-for-test (DFT) solutions reuse ASIC test methodology, wafer probe cards and ATE equipment suitable to test thousands of I/Os can be prohibitively expensive. Instead of a more traditional wafer probe solution, the chip designers embedded a test loopback capability into each of the adjacent input (RX) and output (TX) I/Os. This provided a way to test the wafer through electrical boundary scan. It allowed for nearly complete coverage at wafer probe using an extremely cost-effective probe card. **Figure 5** illustrates a top-down view of the logic area within the processor along with the concurrently designed copper pillar micro-bump layout.

The value-add of test. The importance of test for 2.5D applications cannot be overstated. The sheer number of I/Os between die, interposer and substrate is unforgiving, and a cost-effective 2.5D solution must produce high yields on final assembled parts. Wafer-level test provides full coverage and identifies known-good-die (KGD). The yielding components are then chosen for multi-die assembly to proactively address downstream fallout. For multiple integrated devices (SiP), independent access to each die is required and all interconnects are tested to confirm functional integration.

Once each die is characterized, it is possible to mix and match high and low leakage providing that both die are fully functional. Additionally, if architected into production chips, devices with failed sectors (akin to binning) could be offset against a functional mate with a net result of "recovering" die through test, providing a value-added service to this process (effectively raising yield).

Shrinking the die area through effective supply chain collaboration. A reduction in electrostatic discharge (ESD) protection for the die-to-die I/Os allowed for a reduction in I/O area. This is particularly significant when I/O counts increase into the thousands because area savings can now decrease proportionally. In addition, reducing ESD charge helps reduce capacitive loading, thereby increasing the line bandwidth for the data exchange between receiver and transmitter I/Os.

If a discharge occurs in the 2.5D assembly process, it would most likely arise when a bumped logic die first comes into contact with the metal bond pad of the interposer. Several different I/Os with varying ESD protection schemes were designed. Upon testing, it was found that the ESD requirement for the die-to-die I/Os could be relaxed while still doubling ESD protection targets. This contribution to a more efficient I/O design translated into improved data bandwidth, power, and area simultaneously.

The I/O cell for die-to-die communication occupied an area of only 18µm x 36µm and fit handily within the 40µm pitch of the copper pillar micro-bumps. I/O power was estimated to be 0.5-0.6 picojoule/bit.This low I/O power is a critical requirement for reaching higher bandwidths in products with tight overall power budgets. It is an example of a design element that directly translates into the interposer structure.

The adjacent input (RX) and output (TX) cells fitting under the individual 40µm copper pillar interconnects resulted in a 78% area reduction as compared to general purpose I/O placement. The die-to-die bridge area was 50% smaller and

fully characterized. The large number of I/Os supported by the interposer made for a robust power grid and provided for very low power per I/O.

Although a silicon interposer was chosen for this application, silicon interposers, organic interposers and fine grain substrates represent a continuum of price, performance and density. The right solution per SiP will shift as the capabilities of each improve over time. The design effectively combined the 28nm-SLP die with the 65nm interposer to support size and power reductions, efficient test methodology, and high speed and bandwidth design.

Assembly

Final assembly utilized advanced TSV packaging technologies. Prior to assembly, both the processor dice and interposer go through the bumping process in wafer form. The processor wafers emerge with copper pillar micro-bumps and the interposer wafers have SnAg solder bumps.

The assembly flow is straightforward with the substrate going through an initial bake followed by chip capacitor attach. Interposer placement onto the substrate through solder bump reflow follows. The processors are then placed on top of the interposer and undergo a mass reflow process to join the die to the interposer. Underfill is applied to both the copper pillar micro-bumps and the SnAg flip-chip solder bumps at the same time. Lid attach is then followed by laser marking and ball attach (BGA) as the final step in the assembly flow. **Figure 6** shows how small the interposer footprint actually is compared to the area required by the substrate to fan out the ball grid array. The substrate footprint is defined by the total solder ball count which, in this case, is 27mm x 27mm with depopulated corners for a total of 621 balls at a 1.0mm pitch.

SiP performance

The IC design and 2.5D TSV architecture enabled considerable bandwidth while maintaining high speed and low power operation. The system showcased 1GHz ARM dual core processor operation utilizing 28nm-SLP process technology, supporting memory (DDR3 at 1333Mbps), a boot-ROM (memory chip that allows a workstation to be booted from a server or other remote location), software that

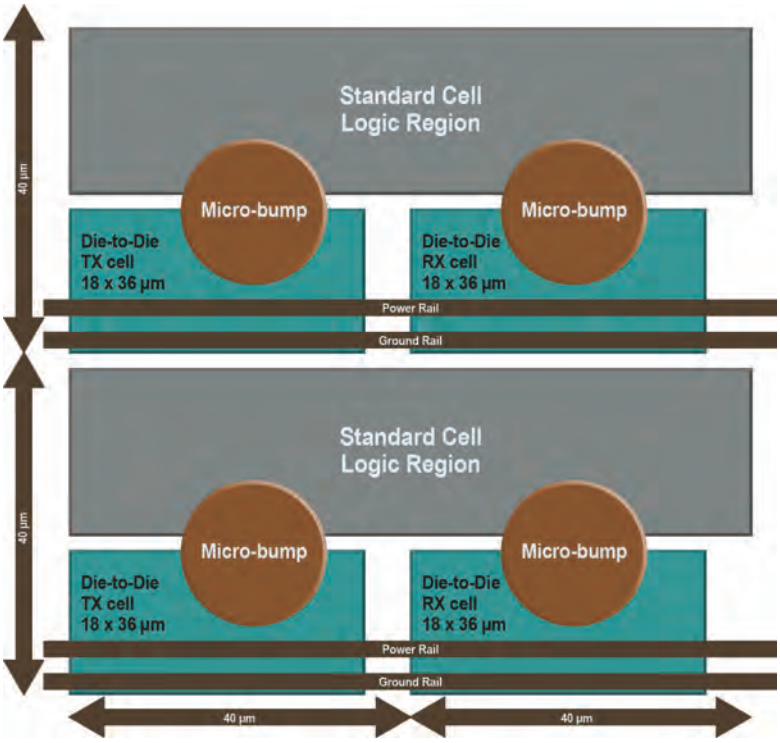

Figure 5: Top down view of a small region of the die-to-die interface (ref. Fig. 2). Relative positioning and size of the I/O cell layout with Cu pillar micro-bump placement for die-to-die interconnection through the 2.5D interposer.

controlled die-to-die communication, general purpose input/output (GPIO) signaling, peripheral devices (i.e., on-board monitor), and an assortment of test functions. The functional die bridge that enabled high bandwidth communication between the processors was successfully characterized. During validation, the interface worked immediately at its (conservative) 500MHz design target. Dedicated I/O test structures were characterized and they appear capable of performance of at least 2GHz.

Applications

One application segment for the Cortex™-A9 processor SiP includes low power mobile devices and servers. Another application targets home gateways that require both low power and massive die-to-die bandwidth.

In general, perhaps the most common application for 2.5D packaging is SoC plus memory with half of all silicon IP targeting this area. Interposers support massive wide parallel buses between memory and logic devices, improving speed and significantly reducing power consumption. High-bandwidth memory (HBM) and Wide I/O memory are clear candidates for 2.5D integration.

The limitations of SoC are strongly seen in some general purpose ASIC designs with large serial memory interfaces that utilize external memory. Silicon node shrinks result in increasing SerDes data that is bumping up against die area constraints. By porting SerDes and large on-chip memories onto discrete die, "the need to move to the latest silicon node can be reserved for only the highest performance functions" [3], i.e. processors. 2.5D materials and designs for extraction models supporting a SerDes interface have since been characterized. Efficient use of 2.5D SiP architecture can reduce die size, power usage, and cost.

High-performance products such as smart TVs, high-end cameras (DSC, SLR, video) and computers will benefit from this type of package construction by having very high speed CPUs connect to differentiated accelerator chips through (high bandwidth) silicon interposers – further reducing cost if a custom CPU die is not needed.

Other applications include SoC plus complex analog, RF, and high-speed interface for industrial, medical, testing, high-end commercial, and networking/telecom ASICs. SoC plus FPGA is a specialized application for fast time-to-market and customized requirements,

while SoC plus other logic will enable IP re-use to reduce total cost-of- ownership.

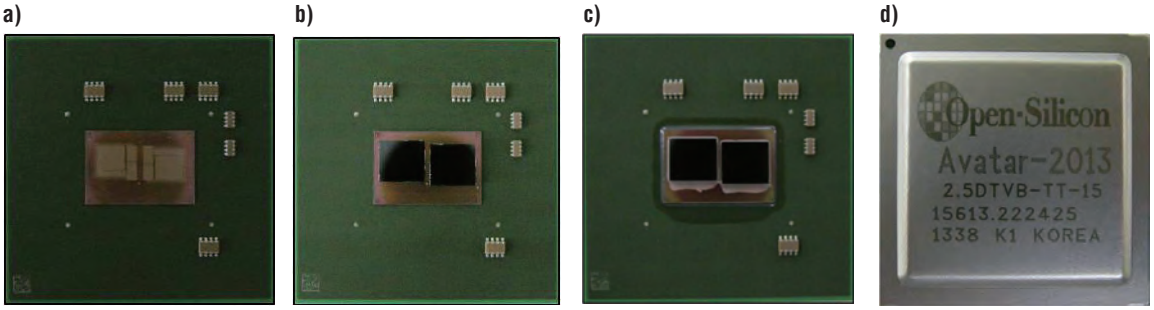

Figure 6: (left to right) a) Interposer placement on the high-density build-up substrate, and b) the two processors mounted on the interposer, c) after underfill application, and d) after lid attach.

Summary

The program affirmed the collaborative business model and swift responsiveness of the supply channel. Investment in infrastructure and process can take years to develop. The program defined and vetted the entire ecosystem from system designer, to foundry, to assembly and test provider. The emphasis on concurrent processor and interposer design optimized performance and design-for-cost advantages.

Foundry and system-level value adds included characterization of materials and validation of a process design kit (PDK). An enhanced EDA reference flow for 2.5D TSV design was created by the foundry to address top level interposer design and floor planning, TSV usage, front- and back-side interconnect and redistribution (routing). Software execution through a dual processor system included design-for-test (DFT) strategies that were shown to both validate performance and increase yields.

The program showcased a supply chain that came together to produce a "first time right" fully functional product demonstration. These competencies will help promote adoption of 2.5D SiP packaging for a broad range of leading edge designs.

Acknowledgements

The authors thank Abu Eghan and Abhishek Chhajer, responsible for substrate and interposer designs, respectively, at Open-Silicon. Acknowledgments also go to: Ravi Gutala on PDK support, Manjunath Prabhu's ESD design guidance. Jens Oswald, Christian Goetze, Samuel George, Juan Boon Tan, Wei Liu, Shun Qiang Gong, and Adam Beece from GLOBALFOUNDRIES, similarly contributed to the successful outcome of the program. The authors thank and acknowledge WonChul Do, EunHo Park, YoungRae Kim, DongHoon Han, and Yonglae Ko of the Amkor K1 R&D team located in Seoul, Korea, for their timely support and dedication. Recognition goes to David McCann, VP of Packaging Technology at GLOBALFOUNDRIES, and Ron Huemoeller, SVP of Advanced Product Development at Amkor Technology, for supporting this project.

References

1. The results of this paper were presented at the IMAPS Device Packaging Conference, March 2014.
2. B. Johnson, G. Ramamoorthy, M. Reitz, S. Tuan Want, "Market trends: soaring design costs threaten economic viability of Moore's Law," Gartner, Sept. 10, 2013.
3. GLOBALFOUNDRIES news release, "Open-Silicon and GLOBALFOUNDRIES demonstrate custom 28nm SoC using 2.5D technology," Nov. 20-21, 2013.

Biographies

Deborah Patterson received her BS in Electrical Engineering from the U. of California, San Diego and BS in Biology from Furman U., and is Senior Director of Product and Technology Marketing at Amkor Technology, Inc.; email Deborah.Patterson@amkor.com

Mike Kelly received his MS in Chemical Engineering from the U. of Idaho and MS in Mechanical Engineering from the U. of Washington and is Senior Director Advanced Product Development at Amkor Technology, Inc.

Rick Reed received his BS in Electrical Engineering from the U. of Illinois, Urbana-Champaign and his MS and PhD in Electrical Engineering from North Carolina State U. He is Director of Advanced Product and Platform Development at Amkor Technology, Inc.

Steve Eplett received his MS in Electrical Engineering from Rice U. and is Manager of Design Technology and Automation at Open-Silicon, Inc.

Zafer Kutlu received his MS and PhD in Mechanical Engineering from Stanford U. and is Principal Member of Technical Staff at GLOBALFOUNDRIES.

Ramakanth Alapati received his MS in Chemical Engineering from the U. of Kansas and is Director of Package Architecture and Customer Technology Group at GLOBALFOUNDRIES.