Chip Scale Review®

ChipScaleReview.com

The Future of Semiconductor Packaging

Volume 29, Number 4

July • August 2025

Lithography and bonding: Holistic scaling solution for next-generation 3D devices

- A path to high-density front and backside wafer connectivity
- Bridging performance and yield: The evolving role of interconnect technologies in HBM
- Advanced ion beam technologies for full-surface etching and dimensional correction of wafers
- WLP, PLP, FOWLP, FOPLP, CoWoS® (TSV-interposer), CoPoS (TGV-interposer), and advanced packaging

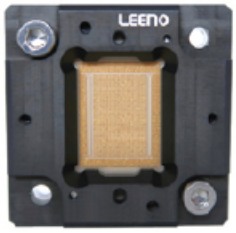
LEENO

www.leeno.com

10K Pins in 7 Days

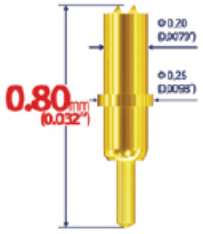
Fine Pitch Probe & Probe Head

Proven Mass Production Capability



High Speed RF

Short Signal Path 0.80mm with Spring Contact Probe!



Through Continuous Non-stop Innovation

75㎛ (0.0029")

80㎛ (0.0031")

2mm

3.45mm

4.10mm

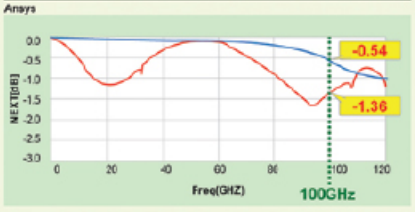
75㎛ (0.0029")

Probe Head

RF-300㎛ Pitch	97㎛ Pitch	92㎛ Pitch	92㎛ Pitch

5G MP 100GHz Coaxial Socket

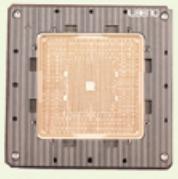
Insertion Loss



Return Loss

Crosstalk

▪ Pitch: 0.35mm, Array



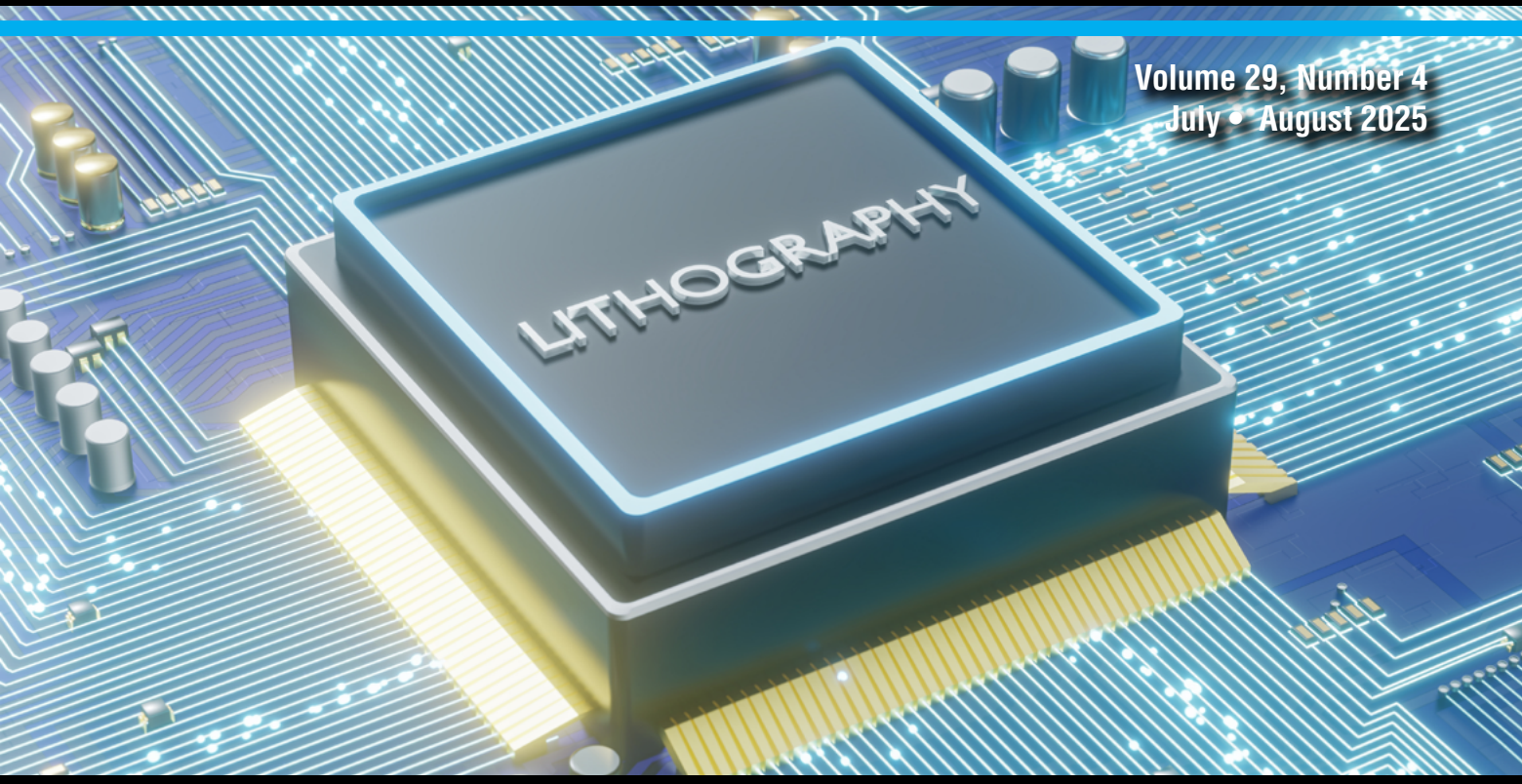
5 Sockets in 10 Days

GLOBAL LEADER LEENO

HEAD OFFICE
10 105beon-gil MieumSandan-ro Gangseo-gu, Busan Metropolitan City, Korea

CONTACT
hskang@leeno.co.kr / +82 10 8737 6561
kimmj@leeno.co.kr / +82 10 5314 6560

CONTENTS



Volume 29, Number 4
July • August 2025

To meet the demand for extreme interconnection density reduction, a holistic approach is needed to handle the challenges of overlay requirements. Combining advanced front-end lithography with state-of-the-art bonding equipment achieves this objective. The feature article on page 11 also discusses how what is learned with respect to tackling overlay challenges can also be used to enable advanced packaging, bonding-assisted 3D stack engineering for thermal management, and D2W hybrid bonding applications.

Cover image courtesy of iStock/koto feja

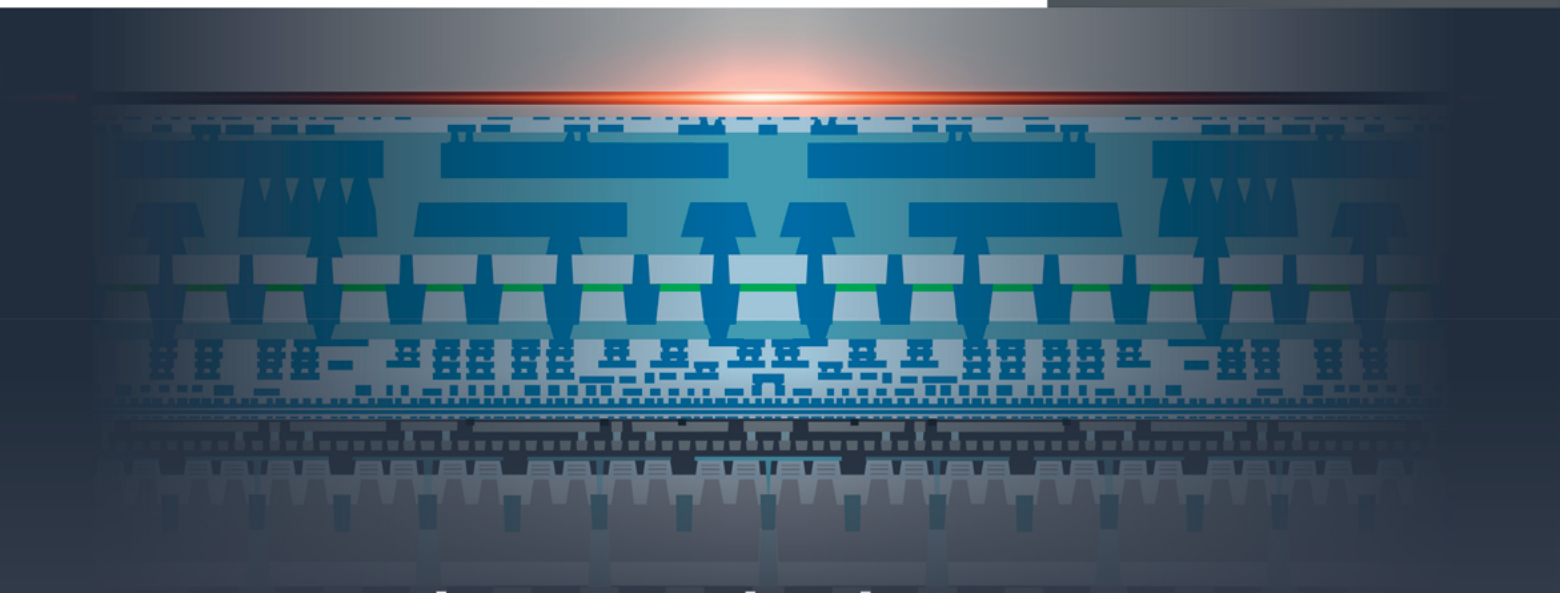
DEPARTMENTS

FEATURE ARTICLES

EVG®

invent
innovate
implement

www.EVGroup.com



IR LayerRelease™ Technology

- Replacement for mechanical debonding thanks to through-silicon IR laser
- Enabling ultra-thin film or layer transfer from silicon carriers
- Nanometer-precision release of bonded, deposited or grown layers
- Front-end compatibility through silicon carriers and inorganic release materials
- Supporting future roadmaps in advanced packaging and transistor scaling

EVG®880 LayerRelease™

GET IN TOUCH to discuss your manufacturing needs
www.EVGroup.com

CONTENTS

Chip Scale Review
ChipScaleReview.com
The Future of Semiconductor Packaging

STAFF
Kim Newman
Publisher
knewman@chipscalereview.com

Lawrence Michaels
Managing Director
Editor-in-Chief
lmichaels@chipscalereview.com

Debra Vogler
Senior Technical Editor
debravogler@me.com

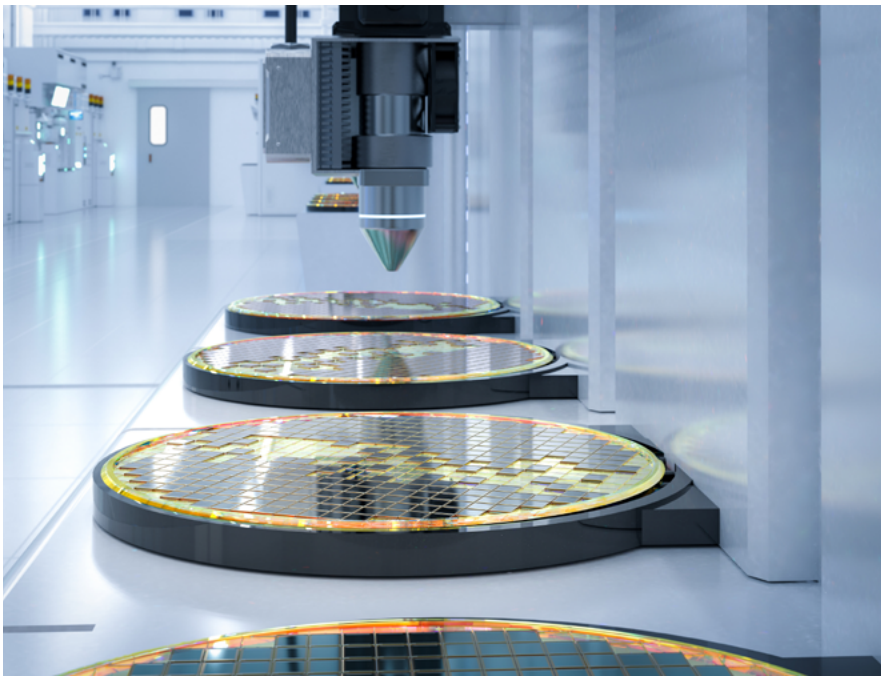
SUBSCRIPTION—INQUIRIES
Chip Scale Review
All subscription changes, additions, deletions to any and all subscriptions should be made by email only to
subs@chipscalereview.com

Advertising Production Inquiries:
Lawrence Michaels
lmichaels@chipscalereview.com

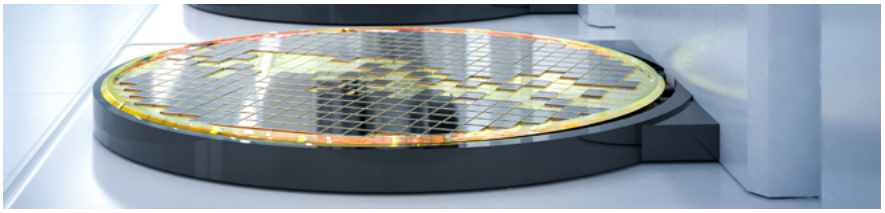
Subscriptions in the U.S. are available without charge to qualified individuals in the electronics industry.
Chip Scale Review is published six times a year with issues in January-February, March-April, May-June, July-August, September-October and November-December.

P.O. Box 2165
Morgan Hill, CA 95038
Tel: +1-408-846-8580
E-Mail: subs@chipscalereview.com

Printed in the United States



FEATURE ARTICLES (continued)

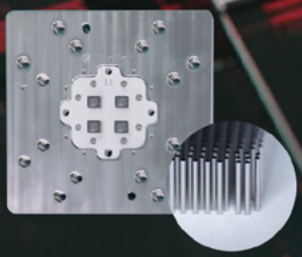


WinWay Technology

sales@winwayglobal.com
+886 7 361 0999 / +886 3 656 8282
www.winwayglobal.com

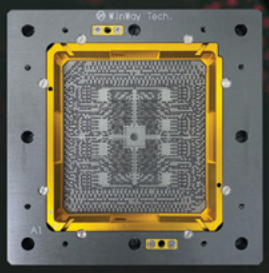
Your Trusted Partner in Semiconductor Testing

Total Solution for Advanced Semiconductor Test



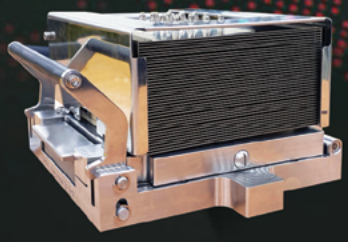
Vertical Probe Card

Minimum pitch : 75 um
High pin count ≤ 35,000pins



Coaxial Socket

SerDes 224Gbps PAM4
Supports up to 30,000 pins

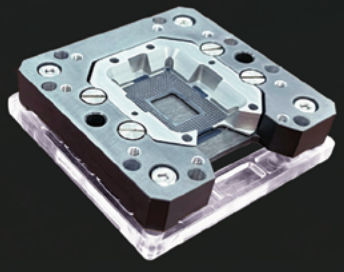


Functional Burn-in

High Power 1000W
High Speed PCIE Gen5 32Gbps

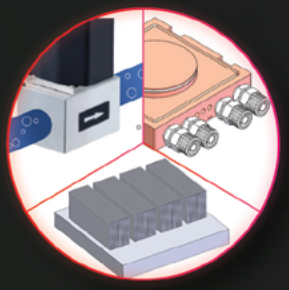
Spring Probe

In-house manufacturing



Optical Socket

SerDes 224Gbps PAM4
Optical alignment



Thermal Product

2000W cooling capacity
Temperature : 20 to 150°C

Advanced ion beam technologies for full-surface etching and dimensional correction of wafers

By Matthias Nestler, Mandy Gebhardt [scia Systems GmbH]

Advanced packaging is no longer simply a manufacturing improvement; it's a strategic differentiator for chipmakers aiming to stay competitive in artificial intelligence (AI), quantum computing, Internet of Things (IoT), mobile, and photonics markets. With Moore's Law approaching both physical and economic limits, advanced packaging has become essential for driving further gains in chip performance, power efficiency, and integration.

A key enabler of these advanced packaging solutions is wafer thinning, which allows for the production of thinner, lighter, and more densely-packed chips. With the proliferation of 2.5D and 3D integration technologies, wafer thinning has become essential for improving interconnect density, heat dissipation, and chip reliability.

Traditional methods of wafer thinning like mechanical grinding, chemical etching, and chemical mechanical polishing (CMP) can create certain defects like micro-cracks and stress damage that reduce yield and chip reliability, involve harmful and costly chemicals that risk contamination, and cause surface variations and uneven thickness that impact the electrical and thermal performance of the chips on a wafer. As the semiconductor industry moves toward increasingly complex, multi-die and 3D architectures, new approaches are needed that not only address these issues, but also provide higher precision, minimal surface roughness (less than 1nm, root mean square [RMS]), and the ability to create nano-scale geometries and intricate patterning to develop high-density interconnects and microstructures.

Among the various thinning methods, ion beam processing is one of the most precise and effective techniques. It provides well-controlled material removal, the lowest thickness variation, good surface quality, and minimized wafer damage. These advantages make it ideal for advanced packaging, 3D integration, and 3D nano-structuring.

Ion beam etching: An overview

Ion beam etching (IBE) (or ion beam milling [IBM]), is ideally suited for precise surface processing. The technology uses a directed beam of high-energy ions to selectively remove material from the surface and create specific patterns or structures. The ion beam process combines physical and chemical etching. Physical etching uses the kinetic energy of fast inert ions bombarding the surface to sputter, releasing atoms from the target surface. That process works on all materials facing the ion beam. Chemical etching utilizes a chemical reaction between the reactive ions and the target surface. The reaction products must, therefore, be volatile.

A broad beam of positively charged ions, typically argon ions, is accelerated onto a substrate. The ions transfer their kinetic energy to the surface atoms, causing them to be ejected, thereby removing the material. The ion beam is typically larger in diameter than the substrate size, ensuring sufficient removal uniformity and throughput. During milling, the wafer substrate can rotate for the best uniformity.

By varying the angle of incidence and the substrate rotation, material removal can be adjusted precisely to achieve a perfect etching structure with superior homogeneity. This allows for creating many different geometries in a broad spectrum of processable materials.

Ion beam trimming

Ion beam trimming (IBT) is a particular type of IBE that uses a small beam of positively-charged ions (e.g., Ar^+) to etch material from a substrate by ion bombardment. A beam width of typically 8-15mm (full width at half maximum [FWHM]) ensures a sufficient lateral resolution and a high throughput. During trimming, a focused, broad ion beam moves in a meander-shaped pattern across the substrate surface. By altering the local dwell time, it is possible to precisely adjust the material thickness and, therefore, device properties like the frequency of acoustic filters. By introducing an additional reactive gas into the ion beam source, a reactive structuring of the surface—the so-called reactive ion beam trimming (RIBT)—is applied.

Advantages of ion beam processes

Ion beam processes offer several advantages. They can be applied to almost all materials. Ion beam machining is always contactless and nondestructive and does not create mechanical stress on the substrate's surface, thereby avoiding subsurface damage. The ion beam current, energy, and etching rates can be independently controlled, allowing for precise material removal and excellent uniformity across the etched surface. Furthermore, ion beam processes offer high resolution, making them suitable for detailed, intricate designs and sidewall shaping through simple sample tilting. These advantages and their process characteristics make ion beam processes best suited for wafer thinning and advanced packaging.

Ion beam technologies for wafer thinning, film thickness correction, and polishing

While ion beam trimming of components has been established in radio-frequency (RF) filter production for more than twenty years, wafer thinning using ion beam trimming is a relatively new process. On the one hand, wafer thinning can mean that the active layer of a wafer is trimmed to a target thickness and a reduced thickness

variation. On the other hand, reducing the total thickness variation of bulk wafers has recently gained importance. For example, this is true for SiC wafers to ensure high-yield components for vertical power semiconductors. **Figure 1** shows the reduction of the total thickness variation (TTV) of a 6-inch SiC wafer by ion beam trimming.

One challenge for such processes is the significantly higher material removal required than in device trimming. In addition, there is much higher cost pressure, as the wafers are still without components at this processing stage. To make the process economically viable, a compromise between throughput and accuracy (spatial resolution) must be found.

A clever machining strategy, therefore, aims for high throughput with sufficient accuracy as a cost-optimized solution. Particular attention must, therefore, be paid to calculating the removal function, which not only calculates the smallest deviation from the target topology as an optimization parameter, but also the fastest possible target achievement while maintaining moderate trimming accuracy. **Figure 2** shows the thickness variation of twelve silicon-on-insulator (SOI) wafers before and after ion beam trimming.

While ion beam trimming is primarily used to reduce the long-wave thickness variations of layers and wafers, ion beam polishing influences the roughness in the medium-to-high spatial-frequency range (M/HSFR). The effectiveness of ion beam polishing depends on the choice of beam parameters and the material to be processed. Both smoothing and roughening, or the formation of nanoscale patterns, can be achieved. Therefore, the parameters for ion beam polishing should be worked out very carefully.

Advanced direct oxide-oxide and Cu-Cu bonding (DBI) is a newly developed hybrid bonding technology that extends the capabilities of Ziptronix's ZiBond® technology. It allows an interconnect pitch of just a few micrometers and can accommodate 1.5 million connections per cm² [1]. According to Fraunhofer ASSID, the benefits of DBI include fine-pitch 3D interconnect (scaling from <10µm to 1µm or less), high bandwidth (enabling increased I/O as needed), improved performance (with enhanced electrical and thermal characteristics due to the elimination of micro-bumps, underfill and solder), improved yield (with minimized warpage during assembly), and low cost (with reduced process steps and a simplified manufacturing process) [2]. DBI requires extremely smooth surfaces at the interface, exposing both SiO_2 insulators and Cu interconnects. Ion beam polishing is a key enabler that can fulfill these requirements effectively. **Figure 3** shows an example of roughness reduction.

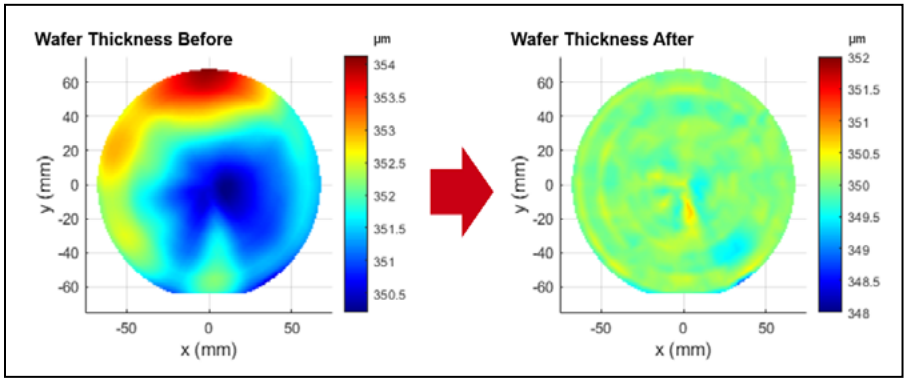


Figure 1: Reduction of the total thickness variation (TTV) of a 6-inch SiC wafer from 3.9µm to 2.4µm by ion beam trimming with 1.8µm average removal, standard deviation reduced from 837nm to 193nm, 50min processing time.

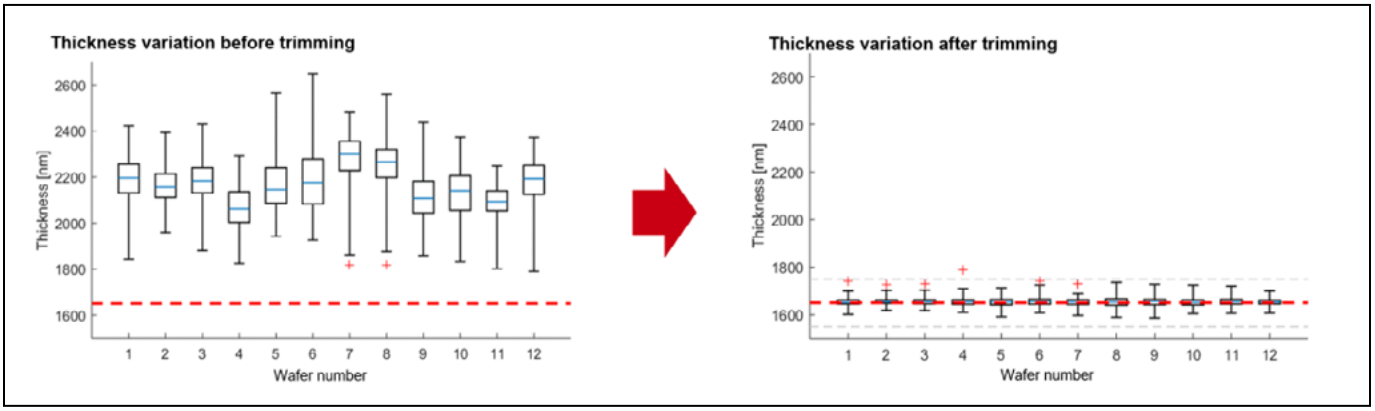


Figure 2: Reduction of wafer-to-wafer-variation (WtW) and within-wafer-variation (WiW) of the silicon layer of a set of twelve SOI wafers.

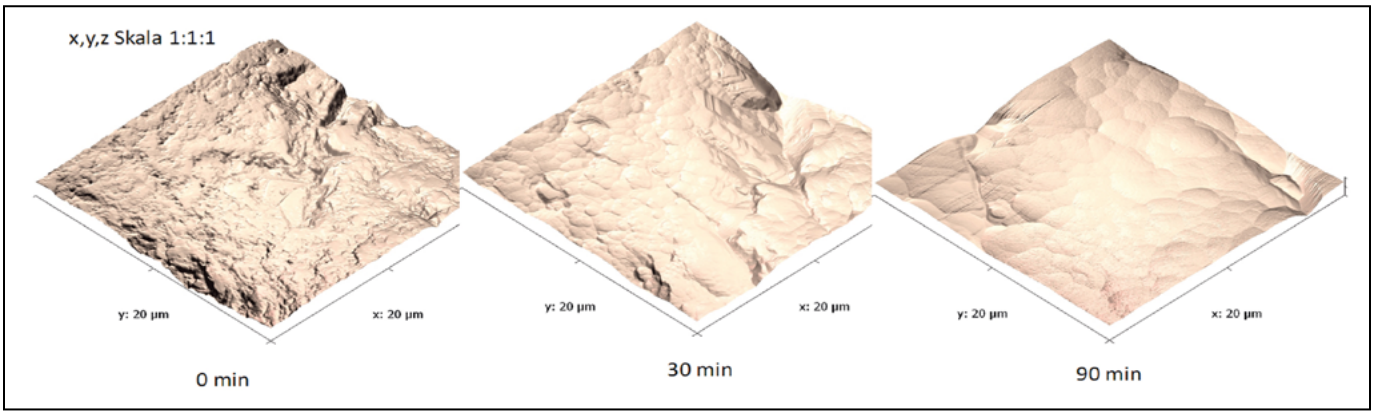


Figure 3: Smoothing of an SiO_2 surface by direct ion beam polishing, scale x:y:z = 1:1:1.

Summary

Ion beam technologies are ideally suited for wafer thinning with nanometer resolution. They can correct preparation errors caused by mechanical polishing or CMP and provide well-controlled material removal, excellent surface quality, and minimized wafer damage.

While IBM can reduce micro-roughness and smooth the wafer surface, IBT is primarily used to correct layer and wafer thickness variations, enabling excellent non-uniformities of less than 1nm. The shown application results demonstrate the high performance needed for advanced packaging applications and help drive the next generation of advanced packaging innovations.

Acknowledgements

The authors wish to thank all involved colleagues at scia Systems for the process and application development of reactive ion beam etching and reactive ion beam trimming, especially Thoralf Dunger, Enrico Loos, and Robert Metzner.

References

1. "Research highlights from Fraunhofer IZM-ASSID," Fraunhofer, https://www.dresden.fraunhofer.de/en/institutes/fraunhofer_izm-assid/highlights.html
2. "Ziptronix and Fraunhofer IZM-ASSID collaborate on development of low-cost 3D integration solutions," Xperi, Oct. 29, 2015; https://investor.xperi.com/news/news-details/2015/Ziptronix-and-Fraunhofer-IZM-ASSID-Collaborate-on-Development-of-Low-Cost-3D-Integration-Solutions/default.aspx

Biographies

Matthias Nestler is Product Development Director at scia Systems, Chemnitz, Germany. He is responsible for the constant optimization of existing products and the development of new fields of application. He began his career as Project Manager for ion beam systems at MicroSystems. Later, he became a Product Manager for that company before joining scia Systems in 2013. He holds a Diplom-Ingenieur (Master's degree) in Physical Engineering from the West Saxon University of Applied Sciences of Zwickau, Germany. Email: m.nestler@scia-systems.com.

Mandy Gebhardt is Head of Marketing at scia Systems, Chemnitz, Germany. Since 2000, she has worked in marketing and public relations in various high-tech companies, and since 2010, as Team Leader of Marketing/PR. She holds a diploma degree in Economic Computer Science from the University of Cooperative Education Glauchau, Germany.

i3 unify™

www.tse21.com

TSE

Enabling Tomorrow's Semiconductor

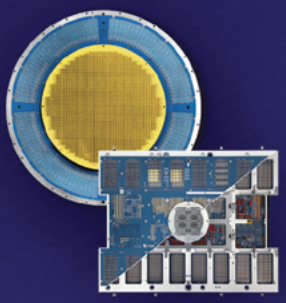
Optimized Advanced InterconneXion Solution

"Innovating Semiconductor Testing"

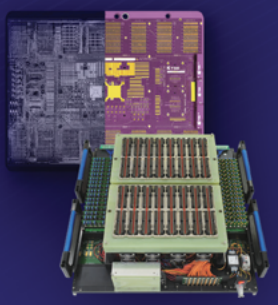
Three-in-One intergrated

InterconneXion

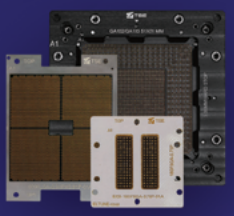
Hardware



Probe Card



Test Interface Board



Test Socket

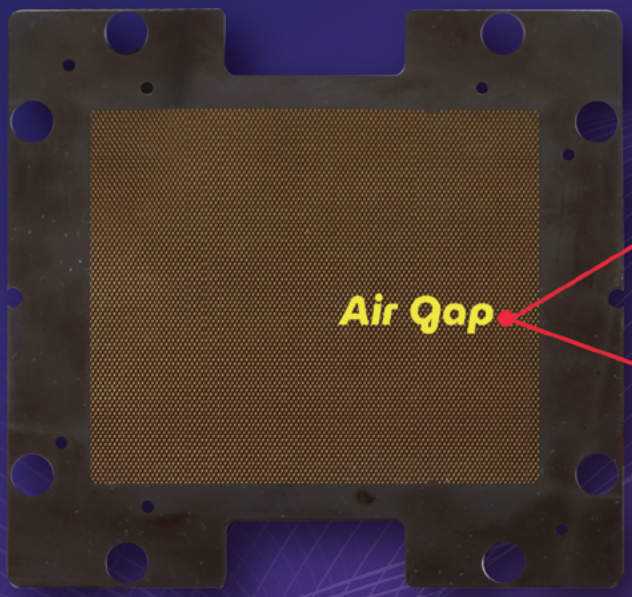
TSE's i3unify architecture integrates Probe Cards, Test Interface Boards, and Test Socket technology, providing an optimal integrated test solution that flexibly responds to various semiconductor chip test requirements. This solution combines integration and flexibility, integrating all components into one efficient system.

A revolutionary three-in-one integrated solution shaping the future of semiconductor testing

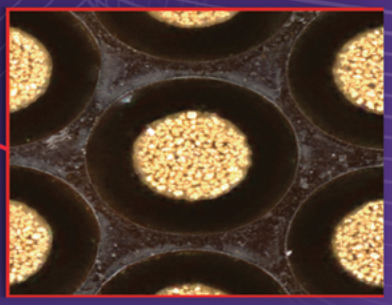
i3 unify™

ELTUNE-air™

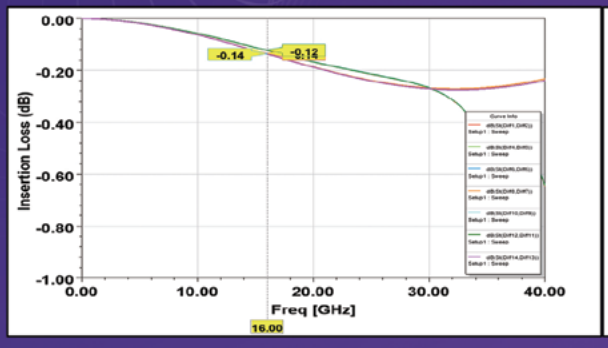
Independent Stroke and Enhanced SI Performance by air gap



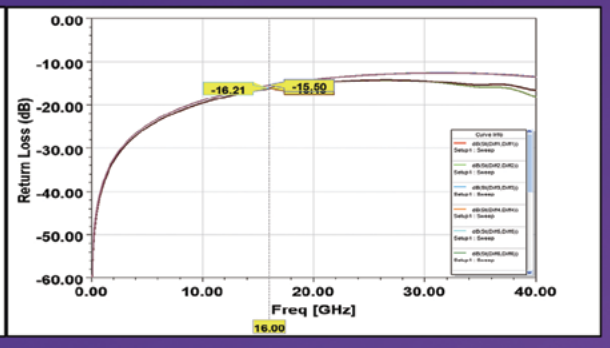
<Section View>



<Top View>



Insertion Loss(dB)



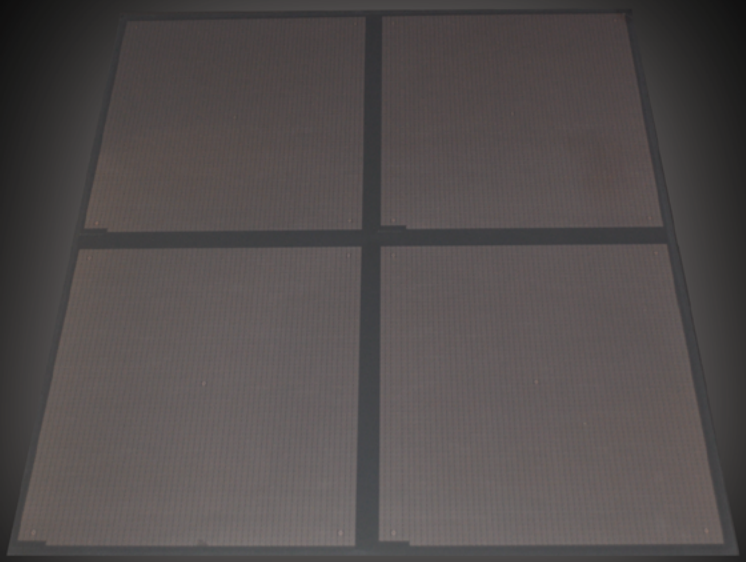
Return Loss(dB)

TSE Co., LTD | www.tse21.com | salesap@tse21.com

DECA

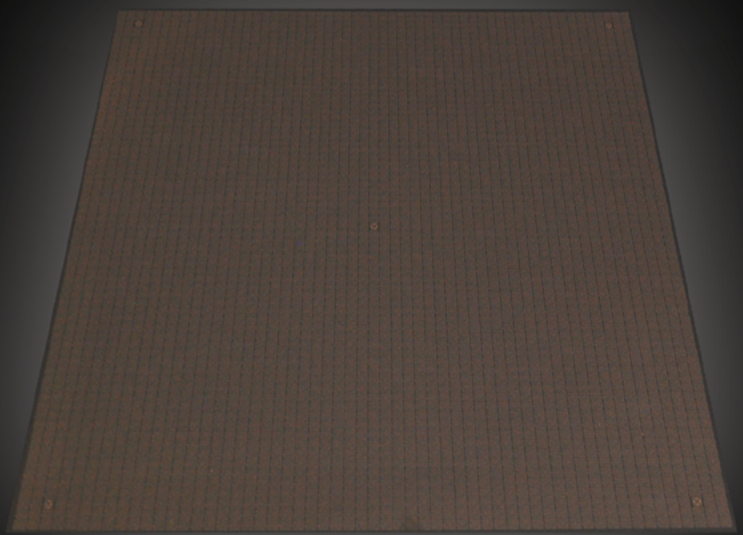
thinkdeca.com

Realizing the full value of semi standard 600mm square panels



Initial format for 300mm test

designed to utilize existing industry
300 mm probe capacity



Current optimized format

full 600mm active area for
maximum utilization

Lithography and bonding: Holistic scaling solution for next-generation 3D devices

By Anton Alexeev, Thomas Plach [EV Group]

Wafer-to-wafer (W2W) bonding is a transformative technology increasingly being adopted in advanced semiconductor manufacturing. It introduces a third dimension to traditional scaling, enabling 3D integration in the front end, similar to back-end advanced packaging trends. As the achievable interconnection pitch decreases, the bonding interface moves closer to the front-end transistor layers. Some researchers predict the emergence of hybrid bonding directly in the front-end-of-line (FEOL). For example, imec proposed a concept of a bonded complementary field-effect transistor (CFET), where bonding connects n- and p-type transistor layers directly. However, the overlay requirements for such a technology are in the single-digit nanometer range—an order of magnitude beyond current state-of-the-art results. Figure 1 illustrates the W2W hybrid bonding roadmap and the main device types benefiting from the technology.

Wafer bonding also allows the integration of chiplets from different process nodes or different substrate materials into a single chip. This significantly shortens electrical paths, reduces power consumption, enhances functionality, and minimizes the physical size of integrated circuit (IC) systems.

Two main W2W bonding techniques for 3D device integration are fusion bonding and hybrid bonding. Fusion bonding connects dielectric surfaces, while hybrid bonding [1] adds metal-to-metal contacts for electrical connections. This article discusses these two techniques in conjunction with lithography—a process crucial for semiconductor manufacturing. We analyze the synergy between lithography overlay process control and bonding.

Achieving high yield in semiconductor manufacturing is critical, and wafer bonding is no exception. Yield can be affected by voids, poor overlay, insufficient bond strength, contact corrosion, and particles. This article focuses on overlay. We break it down to the essential components including alignment error (linear displacement) and nonlinear distortion that is characteristic for high-performance bonding.

Wafer bonding types

In the context of semiconductor device manufacturing, wafer bonding-related overlay can be classified into two major types (Figure 2): W2W bonding overlay and post-bond lithography overlay.

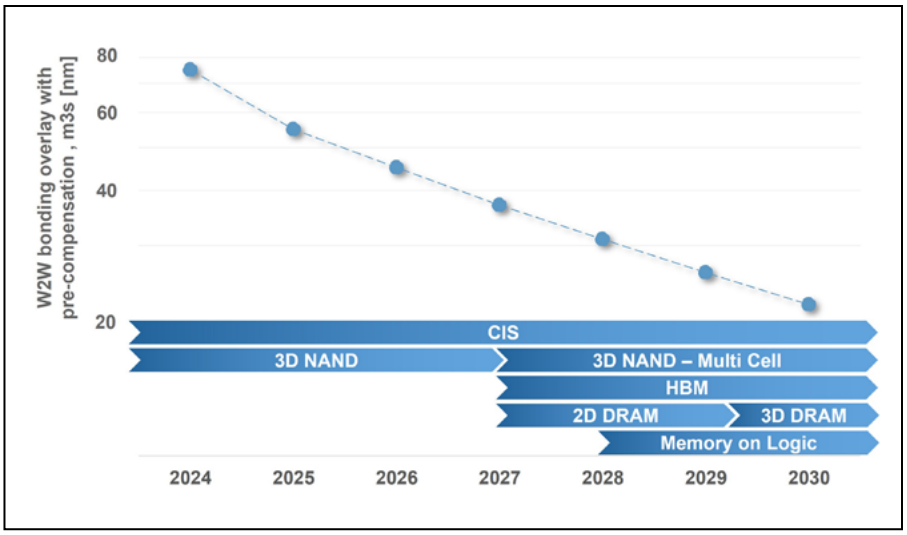


Figure 1: W2W bonding overlay roadmap and major related technologies.

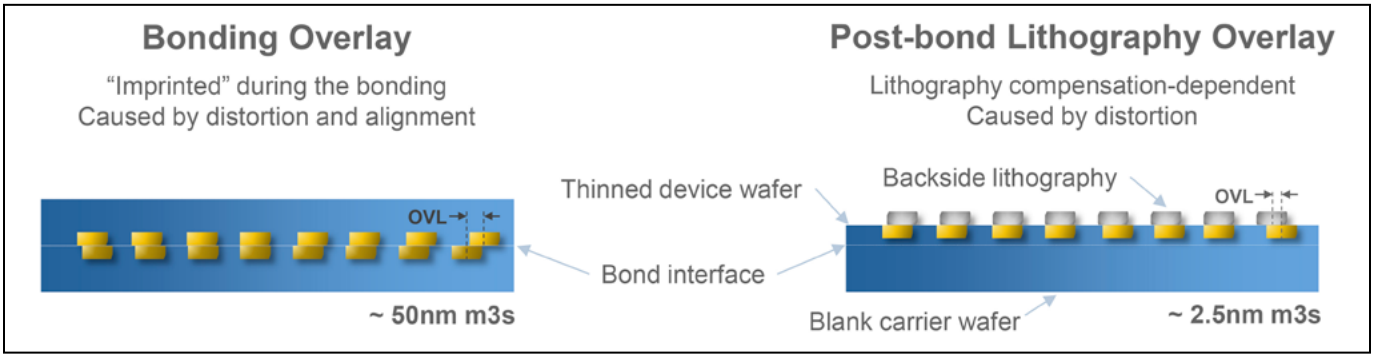


Figure 2: Two main overlay concepts relevant for wafer bonding applications.

Applications where bonding overlay is critical require precise wafer alignment and stable distortion. Low bonding overlay is important for hybrid bonding to ensure physical electrical coupling of metal contacts. Aligned fusion bonding, where the electrical coupling is done via through-silicon vias (TSVs) later, also requires low bonding overlay. Complementary metal-oxide semiconductor (CMOS) image sensors, three-dimensional NAND flash memory (3D NAND), emerging dynamic random access memory (DRAM) designs, advanced packaging, and micro-light-emitting diode (µLED) displays are common applications dependent on the bonding overlay performance.

Post-bond lithography overlay is another critical metric for applications requiring the transfer of a patterned layer onto a blank carrier for further back-side processing. The typical post-bond lithography overlay is an order of magnitude lower than bonding overlay and can be as low as 2.5nm. This process requires wafer edge alignment with relatively low precision—in the tens of micrometers—allowing scanner alignment mark registration in the subsequent lithography step. However, repeatable wafer distortion is crucial because further processing is typically done with lithography. A typical example of technology requiring this bonding type is the backside power delivery network (BSPDN).

Subsequent wafer thinning is required for most bonding applications. Traditionally, the thinning is done by grinding and chemical mechanical polishing (CMP). Grinding destroys one of the wafers resulting in a significant resource waste. An alternative is the usage of infrared (IR) layer release technology combined with silicon-on-insulator (SOI)-like wafers. For integration of this technology, a thin substrate layer is first bonded to a carrier silicon wafer via a nanometer-thin inorganic release layer [2,3]. Next, the substrate layer is processed, and the stack is bonded to the target wafer. The release layer is actuated by an IR laser that passes through the original silicon wafer, which is transparent to the IR laser wavelength.

Wafer bonding overlay

Traditionally, the industry used only linear overlay components (translation, rotation and scale) to characterize the misalignment between bonded wafers. This approach was suitable for µm-level alignment specifications. However, when the overlay decreased down to the sub-500nm range, a new component was revealed: wafer distortion. Wafer distortion is nonlinear deformation that has a complex profile and is dependent on the elastic properties of the wafer and its interaction with the bonder hardware. Two contributions of distortion need to be considered: 1) Pre-process-induced distortion due to deposition, etch, thinning, etc., and 2) Wafer bonding-induced distortion.

The root causes of wafer bonding-induced distortion are found in the physics of wafer bonding. When wafers are aligned, the air trapped between them prevents instant adhesion. To initiate the bonding, the wafers are put in contact in a single point. As a result, a bond wave spreads away, which pushes the air out of the wafer interface. During this process, the wafers slightly deflect from the initial separated position to get in contact with each other. Despite the deflection being extremely small, it causes wafer stretch at the contact front. This results in the distortion of the wafer surface. Wafer mechanical properties like local thickness, crystal orientation, local structural properties, together with the surface uniformity and chuck surface condition, also affect the distortion. Figure 3 depicts the comparison between the measured and simulated distortion overlay components. A quarter-symmetric finite element model simulation was done [4].

Summarizing the relations between wafer alignment, overlay and distortion during the bonding process, we define these concepts as:

- Overlay: A metric for the displacement of wafer (or lithography) patterns after the bonding;
- Alignment: A measure of the accuracy of relative wafer positioning prior to bonding; and
- Distortion: A measure of the wafer in-plane deformation relative to the pre-bonded condition.

In general, the overlay is equal to the sum of the alignment error and the top and the bottom wafer distortion patterns (under the assumption that incoming wafers have a perfect grid).

Repeatability is crucial for bonding overlay minimization during high-volume manufacturing (HVM). A bonder design should, therefore, ensure minimal alignment error and a highly repeatable distortion pattern. Compensation for the repeatable distortion can be effectively accomplished by lithography. Different strategies can be used. For post-bond backside lithography, stability of the distortion enables efficient per lot corrections with advanced process control (APC) lithography feedback loops. For wafer-to-wafer bonding overlay, which is a typical hybrid bonding metric, stable distortion enables effective lithography pre-compensation.

During thinning, which is a mandatory post-bonding process for most of the products, the distortion of the thinned wafer increases, while the distortion of the thicker wafer decreases and often virtually disappears. This is caused by

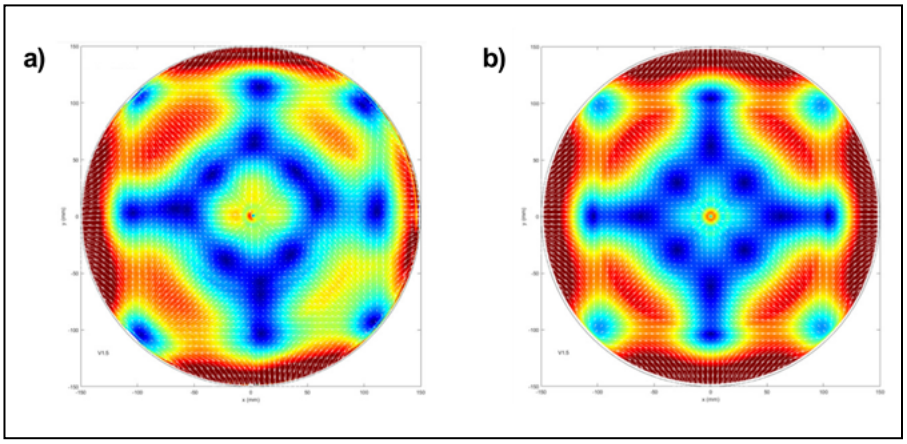
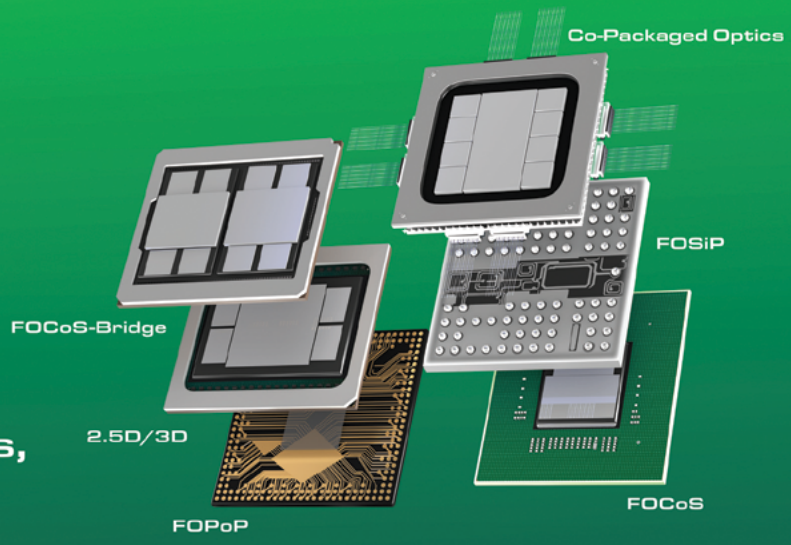


Figure 3: Comparison of: a) (left) The distortion overlay component measurement; and b) (right) The EVG finite element model simulation.

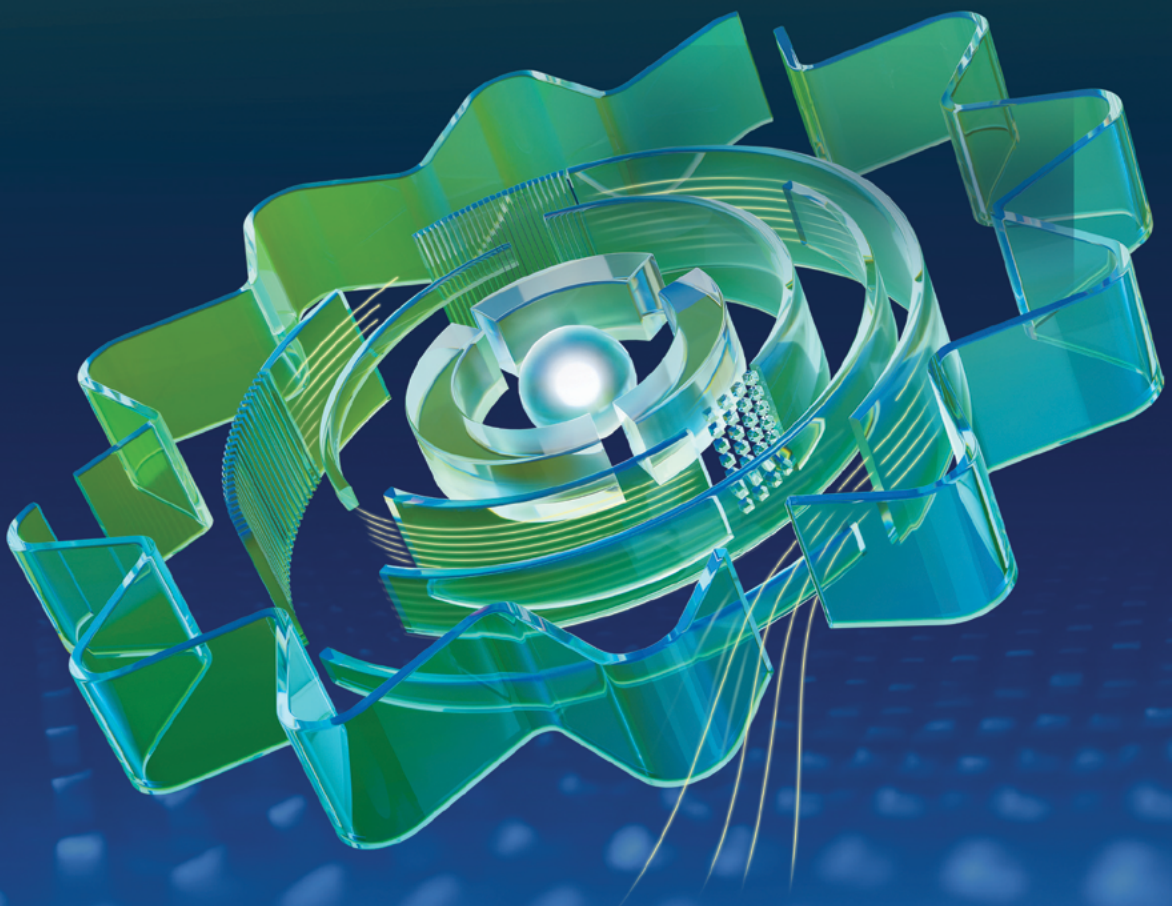
ASE



From electrons to photons,
ASE is innovating for
performance advantage
and energy efficiency.
ASE is enabling the heterogeneous integration
and chiplets era through VIPack™
while delivering sustainable
advanced packaging innovations for:

AI • HPC • DATA CENTER • ROBOTICS
AUTOMOTIVE • IoT • 6G • AEROSPACE • And more.

@aseglobal @aseglobal aseglobal.com



the wafer stack relaxation to the pre-bond shape of the remaining thicker wafer. The overlay between the wafers remains constant because the wafer surfaces do not slip.

Overlay measurements

The classical way to measure the overlay between two wafers is to measure the displacement between two overlay targets. This can be executed by diffraction-based metrology tools or image-based tools. Diffraction-based measurements have a superior accuracy, but require extreme wafer thinning. Image-based overlay measurement can be executed through a wafer using infrared wavelength. The overlay measurements combine both the alignment error and the distortion components.

Distortion isolation

Distortion isolation is more complicated compared to overlay measurements. To accurately derive wafer distortion, the positions of the alignment keys must be measured in the absolute grid before and after bonding. This task can be performed by using the alignment stage of a lithography scanner as a metrology tool. We have demonstrated such a method in our post-bond lithography overlay analysis [5].

An alternative method of distortion estimation is based on measurements of the shape of the incoming wafers before bonding and the resultant shape of the bonded wafer stack. This approach has remarkable sub-millimeter spatial sampling density and high accuracy. Our studies have demonstrated a strong correlation between this method and direct distortion measurements with the scanner alignment stage. We developed a shape model that takes incoming and outgoing wafer shape information to accurately predict the distortion. A comparison between the two methods is shown in Figure 4.

Overlay actuation by bonders and scanners

Besides the alignment, bonders can also actuate a part of the distortion. For this purpose, the top and the bottom bonder chucks have different designs. For example, one of the chucks can be inflated to compensate for wafer scale mismatch. The other chuck is used to control the higher order spatial components of the distortion. The distortion can be optimized by implementing sophisticated sequential release of vacuum of the zones of the top chuck. It enables minimization of the distortion fingerprint for various wafer shapes (e.g., bowl, umbrella or saddle) [6].

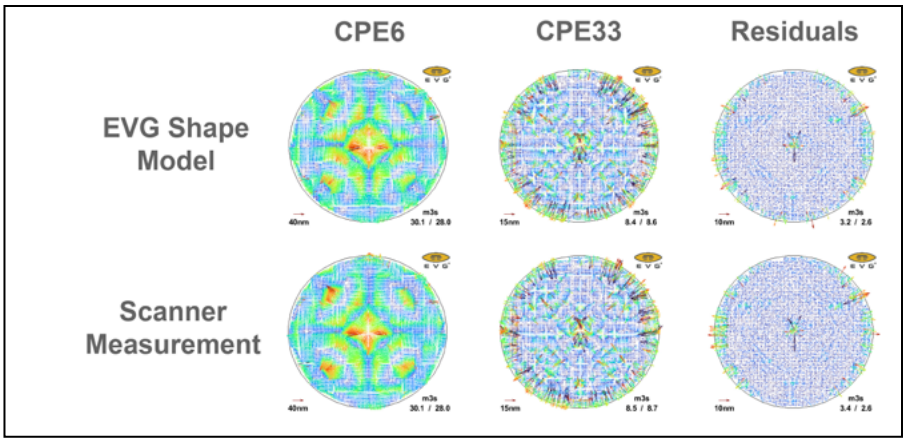


Figure 4: EVG shape model vs. scanner alignment stage data: Distortion measurement accuracy comparision. Data is from imec's STCH wafers. Lithography models content and residuals are shown. CPE models are sequentially fit after simulation of the scanner alignment. The central pin residuals signature signifies the sensitivity of the method and results from the obsolete initiation pin design used in the experiment.

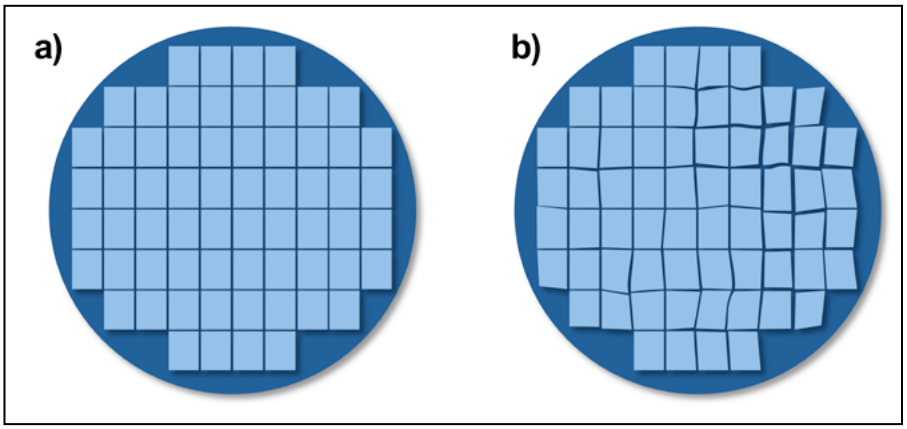


Figure 5: Illustration of exposure actuation per field by a lithography scanner. Each exposure field is depicted as a light blue quadrilateral: a) Exposure with no actuation; and b) Exposure with correction per field enabled.

Modern scanners can actuate the overlay with unprecedented nanometer-level accuracy. This accuracy significantly exceeds the bonder actuation capabilities, yet it cannot be applied directly at the event of bonding. Scanners control dozens of exposure field distortion parameters. Each field can be individually stretched, shifted, rotated, and warped with high-order polynomial profiles (Figure 5). The corrections of each field are defined based on: 1) Wafer alignment data; 2) Feedback from APC; and 3) Input from other systems.

The main purpose of scanner wafer alignment is to precisely locate the wafer on a chuck and characterize its grid. The grid is typically affected by the previous wafer processing and the clamping on the chuck. The feedback from APC loops is based on off-line and in-line overlay metrology, which evaluates stable high-order corrections. Such feedback loops reduce the effects of slowly drifting production processes, such as CMP, etching, and deposition, etc. The information from alignment, feedback loops and other external sources are translated into per-field corrections and actuated by the scanner.

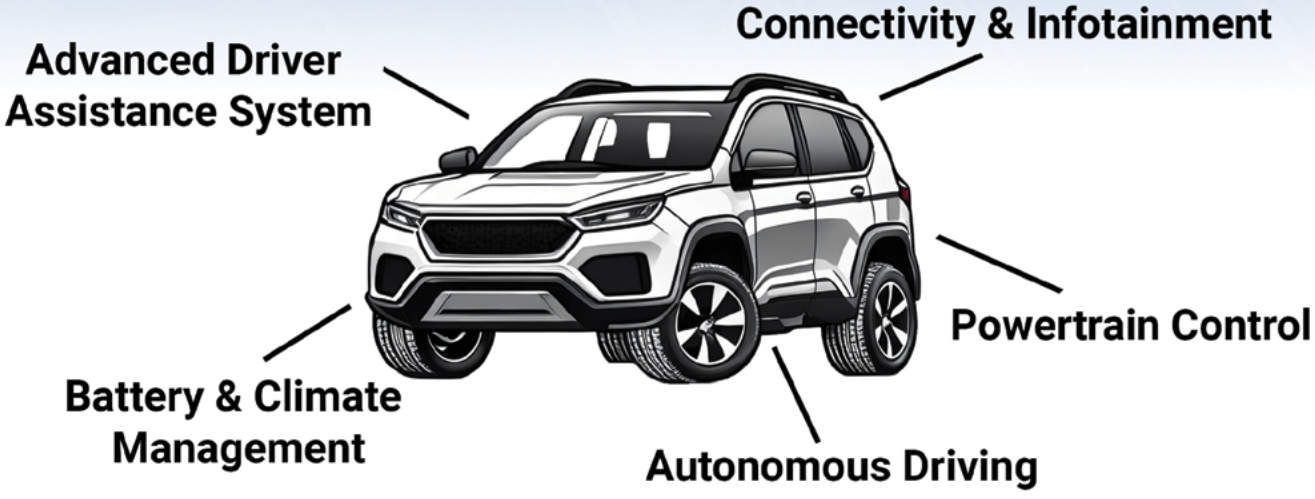
The co-optimization perspective

EVG bonders are designed with the idea of co-optimization with state-of-the-art lithography systems. The induced wafer distortion, therefore, is stable and easy to actuate with the lithography. This enables effective lot-based actuation and pre-compensation strategies, and results in cost savings for our customers.

MICRO CONTROL COMPANY SINCE 1972

Enjoy the Ride When Burn-In Has Been Applied

As automotive systems evolve, semiconductor reliability is critical to the enjoyment, safety, and comfort of the driver and passengers.



Semiconductor devices that complete a burn-in cycle in a Micro Control Company burn-in with test system are significantly more reliable for long-term use.

For the long road ahead, turn to Micro Control Company, The Advanced Burn-In Solution Provider.

LEARN MORE

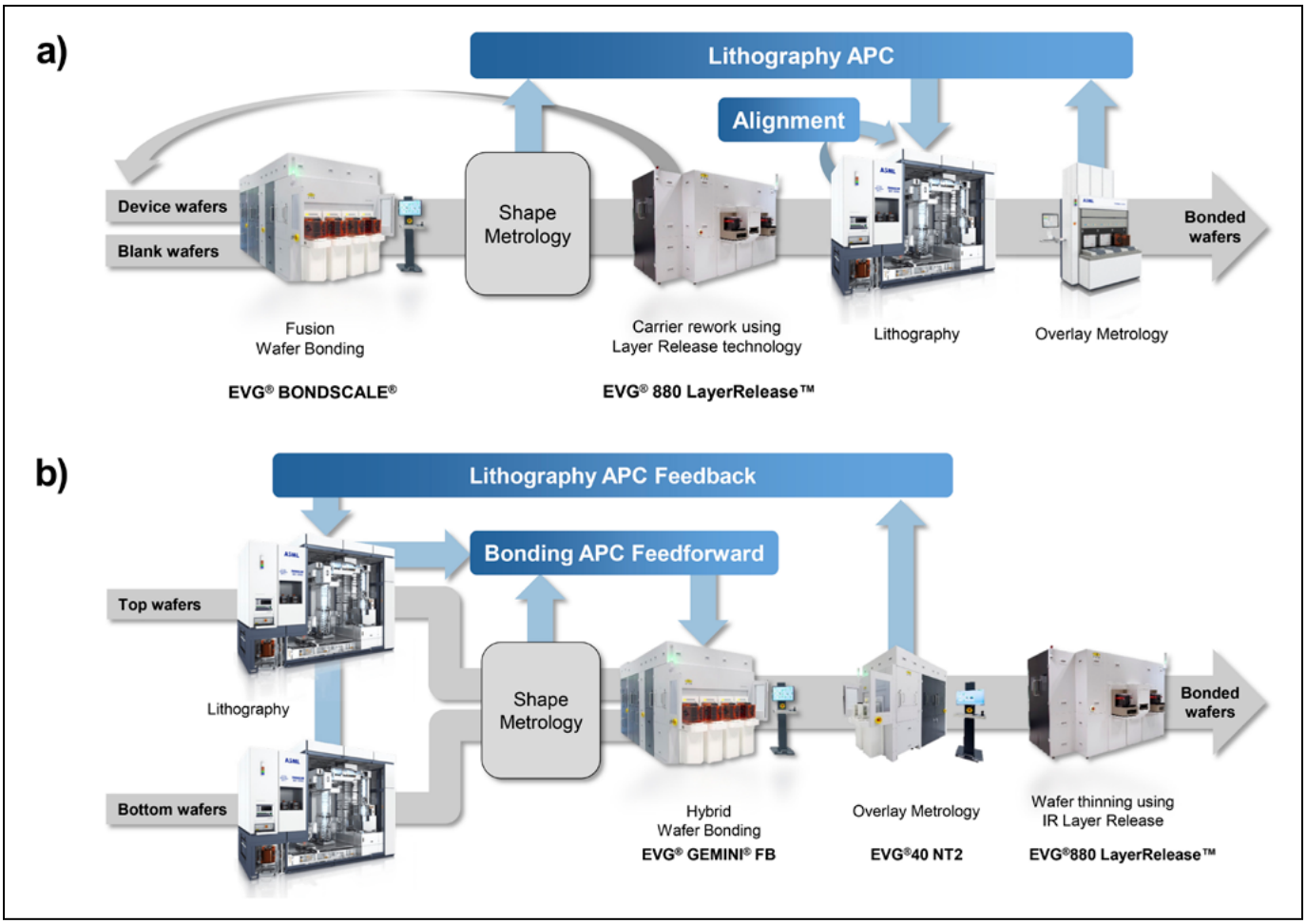


Figure 6: Potential wafer process flows for: a) (top) Thin layer transfer; and b) (bottom) Aligned bonding. Important APC loops are indicated.

We propose two strategies for lithography-based bonding overlay and distortion control in Figure 6. Our analysis first focuses on bonding a patterned wafer to an unpatterned one. In this scenario, overlay between the wafers is not a concern because one wafer is blank. Bonding-induced distortion, however, is crucial. Typically, after bonding, the patterned wafer is thinned to continue processing with lithography from the back side of the device wafer. The stability of the bonder distortion fingerprint enables the per-lot lithography corrections driven by the APC overlay feedback. Moreover, we know that the distortion is also dependent on the stability of the incoming wafers. Wafer shape models can predict the distortion alterations associated with the incoming wafer shape variation. This information can be used as feed-forward data for the scanner corrections.

For the hybrid bonding case, the distortion pattern must be imprinted into the device structure from layer zero long before the bonding event. This is called pre-compensation. This puts additional constraint on the temporal stability of the bonder because the feedback loop is significantly longer in time. Wafer shape metrology can also complement the bonder APC to facilitate overlay correction by the bonder hardware.

Summary

The demand for extreme interconnection density reduction drives overlay requirements and facilitates complex technological solutions. This article demonstrates how combining advanced front-end lithography with state-of-the-art bonding equipment achieves unprecedented bonding overlay. Many learnings are transferable to advanced packaging.

Progressing toward sub-50nm bonding overlay requires a holistic optimization strategy of bonding and bonding-related processes. These considerations include the quality of incoming wafers, the implementation of advanced feedback and feedforward control loops that tightly integrate bonding and lithography systems, and innovations in bonder hardware. We have a complete suite of tools and expertise essential for this purpose. While this article focuses primarily on overlay performance, it is important to recognize that successful bonding also depends on surface activation, cleaning, contamination control, and a deep understanding of both pre- and post-bonding processes. The insights gained throughout this journey are equally vital for enabling advanced packaging, bonding-assisted 3D stack engineering for thermal management, and die-to-wafer (D2W) hybrid bonding applications.

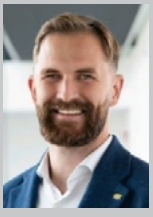
Acknowledgement

The authors and EV Group would like to express our sincere gratitude to imec for its invaluable collaboration

and contribution to our research and development efforts. Their expertise, support, and innovative thinking of the imec team have been instrumental in advancing our work.

References

1. T. Sherwood, et al., "Aggressive pitch scaling (sub-0.5µm) of W2W hybrid bonding through process innovations," IEEE 73rd Electronic Components and Technology Conf. (ECTC), Orlando, FL, USA, 2023, pp. 13-18, doi: 10.1109/ECTC51909.2023.00010.
2. P. Urban, et al., "Novel inorganic IR release process for high temperature W2W and D2W integration," IEEE 74th ECTC (2024), DOI: 10.1109/ECTC51529.2024.00073.
3. T. Sounart, et al., 2024, "IR laser debond from silicon carrier wafers with inorganic thin film release layers for high-density 2.5D and 3D integration," IEEE 74th ECTC (2024), DOI: 10.1109/ECTC51529.2024.00066.
4. C. Mühlstätter, L. Koller, T. Plach, V. Dragoi, M. Wimplinger, "Multiphysics overlay modelling of monolithic 3D fusion and hybrid bonding processes," IEEE 74th ECTC, Denver, CO, USA (2024), pp. 1548-1551, doi: 10.1109/ECTC51529.2024.00252.
5. R. van Haren, et al., "Characterization and mitigation of local wafer deformations introduced by direct wafer-to-wafer bonding," Proc. SPIE 12956, Novel Patterning Technologies (2024) DOI: 10.1117/12.3010477.
6. S. Kang, et al., "Investigation of distortion in wafer-to-wafer bonding with highly bowed wafers," IEEE 74th ECTC, Denver, CO, USA (2024), pp. 386-393, doi: 10.1109/ECTC51529.2024.00069.



Biographies

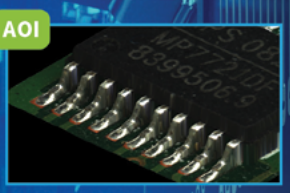
Anton Alexeev is the Business Development Manager at EV Group, based at imec in Belgium. He oversees the hybrid and fusion bonding technologies. Before joining EV Group, he was an expert in front-end lithography overlay at ASML. Anton holds a PhD in Electrical Engineering, and an EngD from Applied Physics Department of Eindhoven U. of Technology (TU/e). Email An.Alexeev@EVGroup.com

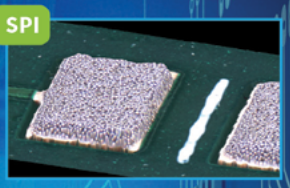
Thomas Plach is Senior Scientist at EV Group, based in St. Florien, Austria where he focuses on direct wafer bonding technologies for a variety of applications. He also leads the Data Analysis and Modeling team, which develops new analysis and advanced process control methodologies for EVG equipment. Thomas holds a PhD in Physics from the U. of Linz

Think Fast. See Small.

SQ7000™+

Multi-Function AOI, SPI & CMM





NEW
5 Micron
Ultra-High Res
MRS Sensor

SQ7000™ + All-in-One Solution

Next-gen multi-process inspector with paramount speed, accuracy and higher resolution for improved yields and processes.

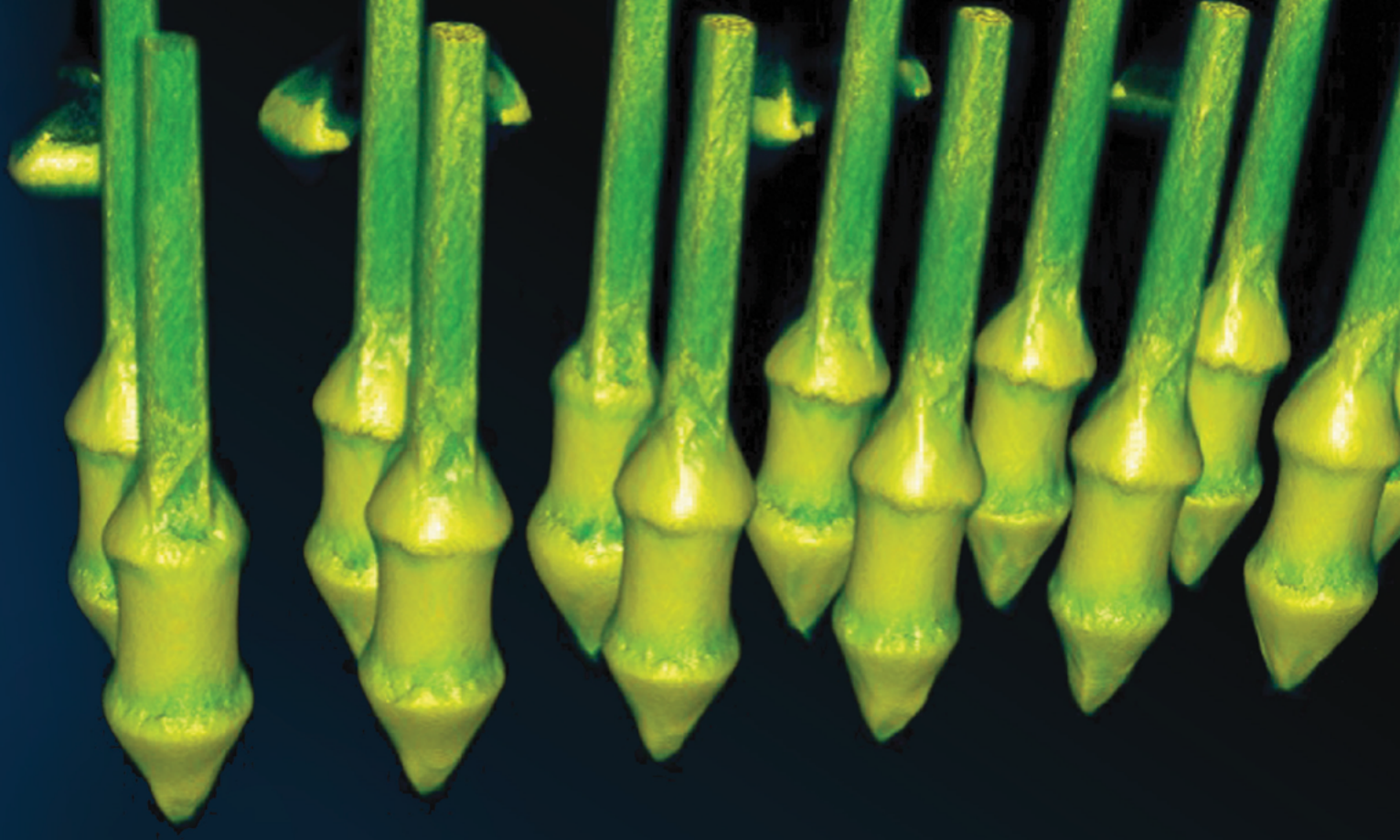
Powered by Multi-Reflection Suppression® (MRS®) sensor technology, the new 5 micron Ultra-High Resolution MRS sensor offers unmatched accuracy by meticulously identifying and rejecting reflection based distortions caused by shiny components and surfaces. The result is ultra-high quality 3D images, high-speed inspection and metrology, and improved yields and processes.

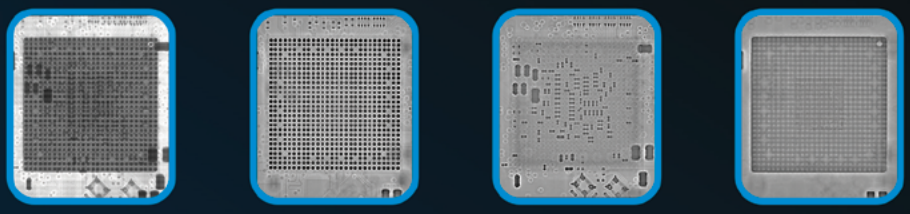
Count on the Nordson SQ7000+ for superior performance for next-generation applications including advanced packaging, advanced SMT, mini/micro LED, 008004/0201 solder paste, and other next-gen applications.

Save Time. Save Expense. Improve Yields.

nordson.com/testinspect | +1 760.918.8471

Nordson
Test & Inspection





Look Past The Impossible.

Dynamic Planar CT

Superior Software for Next Generation 3D Planar X-ray Inspection

Revealing Incredible Details for Nordson's Zero Defects Strategy

Accelerated Image Acquisition and UPH > 2x faster than Planar CT.
Enhanced data quality with clear layer separation and new 3D reconstruction algorithm.
Larger FoV for increased coverage and shorter cycle times.
Reduced complexity and cost with no special clamping needed.

Save Time. Save Expense. Improve Yields.

nordson.com/testinspect | +1 760.918.8471

Nordson
Test & Inspection

A path to high-density front and backside wafer connectivity

By Zsolt Tokei, Eric Beyne, Geert Hellings, Julien Ryckaert [imec]

In 2024, imec introduced complementary metal-oxide semiconductor (CMOS) 2.0, or CMOS 2.0, as a new scaling paradigm to cope with the ever-increasing variety of computational needs associated with the diversification of applications (**Figure 1**) [1-3]. With CMOS 2.0, a system-on-chip (SoC) is partitioned into different functional layers (or tiers) guided by system-technology co-optimization (STCO). Each of the functional layers is built using the technology option that most closely matches the constraints of the functionality.

Advanced 3D interconnect technologies reconnect the heterogeneous tiers of the SoC. This reconnection is reminiscent of an evolution that has already made its way into commercial compute products. For example, think about the 3D stacking of a static random access memory (SRAM) chip on top of a processor. But what differentiates the CMOS 2.0 approach is that the heterogeneity is brought inside the SoC itself. Depending on the needs of the application, for example, CMOS 2.0 envisions even splitting the logic part of the SoC into a high-drive logic layer (optimized for bandwidth and performance) and a high-density logic layer (optimized for logic density and performance/Watt). The high-density layer can be fabricated using the most advanced technologies, including the most scaled transistor architecture.

Another key feature is a backside power delivery network (BSPDN): Part of the active devices are powered from the wafer's backside rather than through conventional frontside power delivery schemes. As such, extreme back-end-of-line (BEOL) pitch patterning will be possible in the tier's frontside without the constraint of voltage drop on the power supplies.

Basically, in this approach, we have modified the device wafer in a very thin front-end-of-line (FEOL) active device layer, wherein on one side (the original "frontside") there is a dense back-end-of-line (BEOL) signal routing layer stack, and on the other side (the original "backside," but now the new frontside) are the power supply and external I/O connections. It is also possible to stack multiples of such thin device layers with dense interconnects on each side. Each layer may integrate different types of devices, e.g., different types of logic or memory, electrostatic discharge (ESD) protection devices, voltage regulating circuits, etc. We call this dense 3D stacking of device layers, "CMOS 2.0."

With this modular approach to system scaling, chip design and manufacturing move away from the general-purpose CMOS technology platform that has

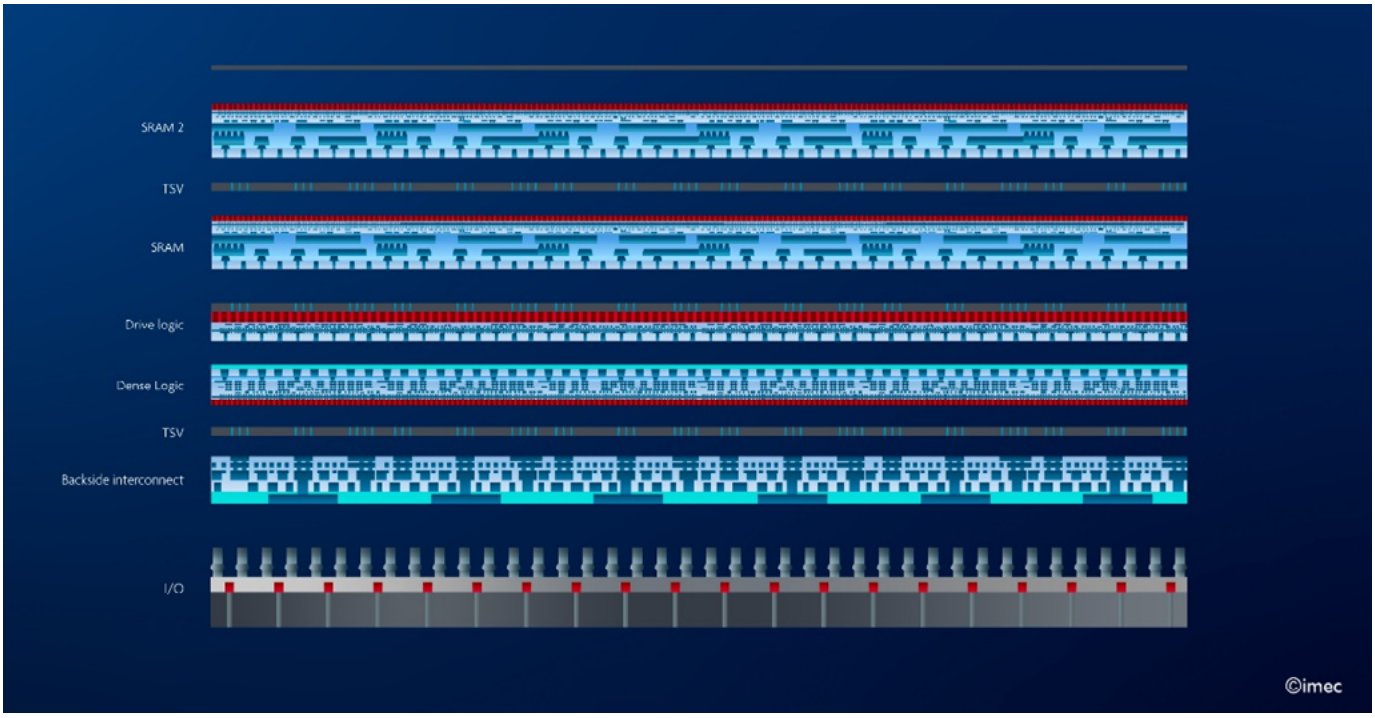


Figure 1: Example of a possible partitioning of a SoC in the CMOS 2.0 era.

served the semiconductor industry for decades, but struggles to adequately address the growing and diversifying compute demands. This approach helps solve compute system-scaling bottlenecks, bringing value to every player in the semiconductor ecosystem—including system and fabless companies.

3D interconnection and backside technologies: Foundational to CMOS 2.0

CMOS 2.0 relies on all semiconductor innovations of the past, including logic device scaling, memory density scaling, advanced lithography, 3D integration, and BSPDN technology. But the concept can only now become a reality thanks to recent breakthroughs in 3D interconnection and backside technologies. Wafer-to-wafer hybrid bonding, for example, starts to offer sub-µm interconnect pitch connectivity. It can, as such, provide an interconnect density matching the last metal layers of the BEOL—key to enabling logic-on-logic or memory-on-logic tier stacking with hybrid-bonding connections. Backside power delivery technology is expected to evolve to an even finer granularity level with the enablement of direct access to transistor terminals. This capability, although initially targeted for power connections, opens the possibility for fine-grain signal connectivity to also migrate to the backside. In this way, any device technology layer will become suspended between two independent interconnect stacks. The combination of fine-pitch bonding and fine-grain backside processing, as illustrated in Figure 2, is foundational to enable the vision of CMOS 2.0 illustrated in Figure 1. At the 2025 Symposium on VLSI Technology and Circuits, imec reported progress in wafer-to-wafer hybrid bonding and backside vias—two 3D integration technologies underpinning the realization of CMOS 2.0 [4]. These technologies lay the foundation for designing new system architectures around the CMOS 2.0 vision—guided by STCO—in which BSPDNs will play a central role. Also at the 2025 VLSI Symposium, imec researchers highlighted the power-performance-area-cost (PPAC) benefits that such BSPDNs can bring to advanced system architectures [5].

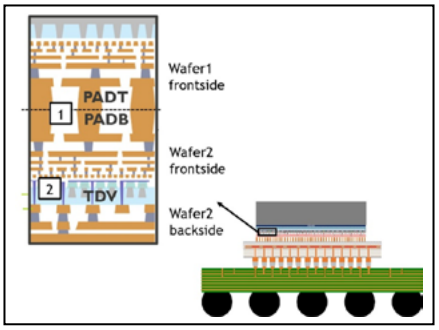


Figure 2: Schematic illustration of high-density face-to-face hybrid bonding connections and a backside high-density connectivity network (as presented at 2025 VLSI [4]). (PADT=top pad; PADB=bottom pad; TDV=through-dielectric via.)

Wafer-to-wafer hybrid bonding towards 250nm pitch: A roadmap view

A large variety of 3D interconnection technologies have been developed over the years, spanning a broad range of interconnect pitches and serving different applications needs (Figure 3). Of all these technologies, wafer-to-wafer hybrid bonding is best suited to provide the 3D interconnect pitches and densities required for memory/logic-on-logic tier stacking in a CMOS 2.0 context. Wafer-to-wafer bonded Cu pads offer short and direct low-resistive connections from one tier to the other. At scaled pitch, the wafer-level connections can deliver a high-bandwidth density, as well as reduced energy per bit during signal transmission.

The sections below discuss various aspects of wafer-to-wafer hybrid bonding.

The classical wafer-to-wafer hybrid bonding process flow. A classical

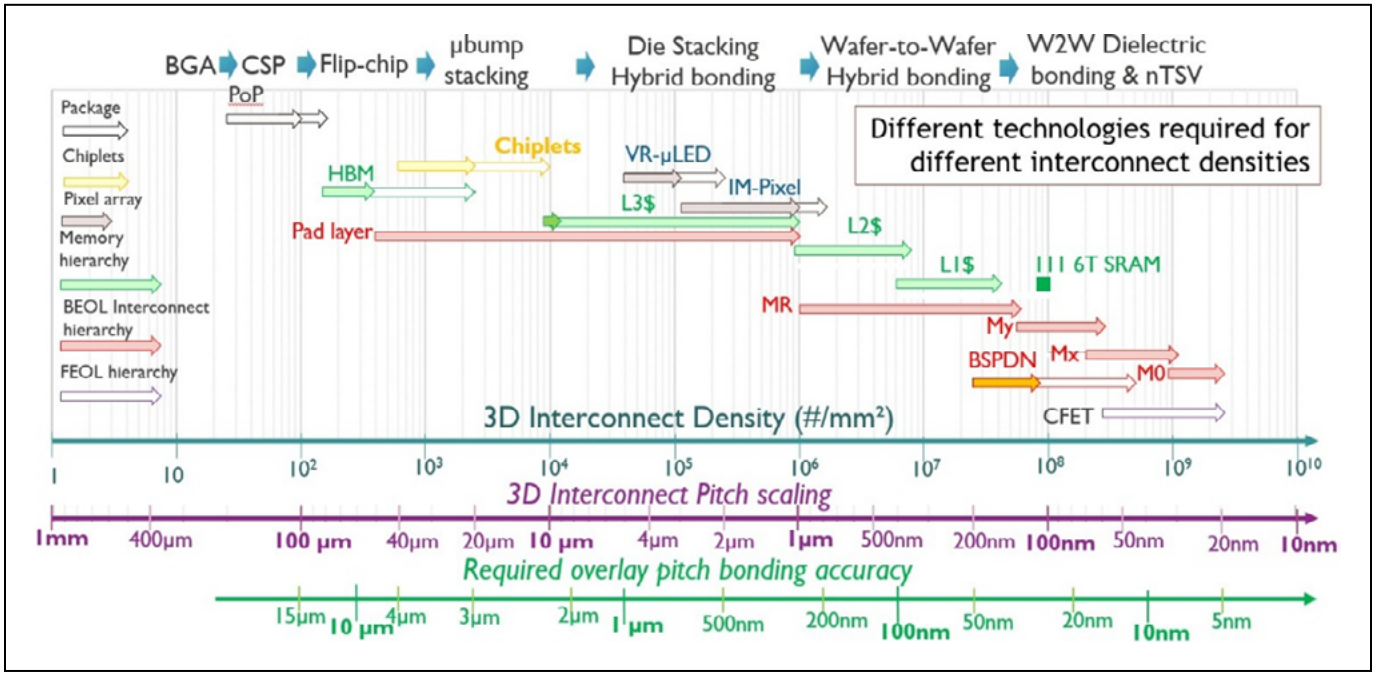


Figure 3: Imec's 3D interconnect technology scaling roadmap, showing the different technologies required for different interconnect densities. (BGA=ball grid array; CSP=chip scale package; W2W=wafer-to-wafer; Mx, My and MR represent the BEOL interconnect hierarchy.)

hybrid bonding process (Figure 4) starts from two fully-processed 300mm wafers, with completed FEOL and BEOL (see also Figure 2). The first part of the flow resembles an on-chip BEOL damascene process, where small cavities are etched into the bonding dielectric, for which SiO_2 is predominantly used. The cavities are filled with barrier metal, seed, and Cu. This is followed by a chemical mechanical polishing (CMP) step optimized for high across-wafer uniformity to produce extremely flat dielectric surfaces while achieving a controlled few nanometers of recess for the Cu pads. After accurate alignment, the actual bonding of the two wafers is performed at room temperature by bringing the wafers into contact at the center of the wafer. The polished wafer surfaces adherence results in a strong wafer-to-wafer attraction, resulting in a bonding wave, closing the wafer-to-wafer gap from the center to the edge. After this room temperature bonding step, the wafers are annealed at higher temperatures to obtain a permanent dielectric-to-dielectric and Cu-to-Cu bond.

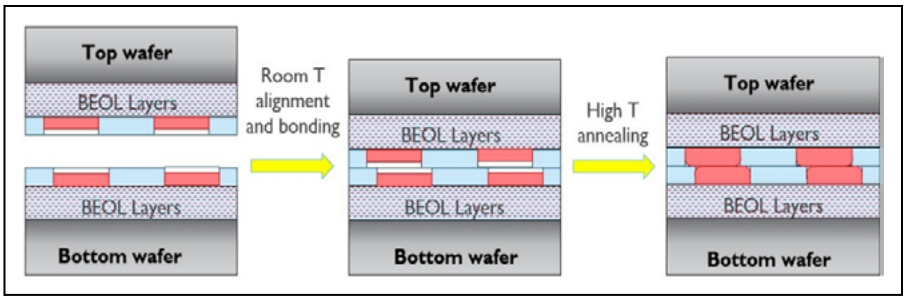


Figure 4: The classical wafer-to-wafer hybrid bonding process flow.

Reliable 400nm pitch wafer-to-wafer connections. At the 2023 International Electron Devices Meeting (IEDM), imec demonstrated reliable 400nm-pitch wafer-to-wafer connections at high yield, a significant improvement over the 1µm pitch connections used in industrial wafer-to-wafer bonding processes [6]. This leap in interconnect pitch was enabled by several process flow improvements, including enhanced control of the wafers' surface topology and the use of SiCN as the bonding dielectric. SiCN was found to offer better bonding strength and scalability than conventional SiO_2.

Pushing the hybrid wafer-to-wafer bonding roadmap towards 200nm pitch. Whenever we go deeper in the system hierarchy— ultimately splitting the logic part into specialized logic layers—bonding pitches below 400nm will be needed, driving the wafer-to-wafer hybrid bonding roadmap towards 200nm pitch. But as the pitch continues to scale, so do the requirements for the bonding overlay between two Cu pads. In general, the overlay accuracy of the bonding process corresponds to one

fourth of the pitch, translating into an overlay as small as 50nm for a 200nm pitch bonding process. Achieving this high level of accuracy at 300mm wafer scale is today's biggest challenge to achieve higher interconnect densities.

To continue the roadmap, imec researchers work towards a more fundamental understanding of the bonding process, and the factors that interfere with the high level of overlay accuracy. It is well known that during bonding, the two wafers get easily deformed and distorted, hindering precise overlay between Cu pads. The team found through simulations that the bonding wave that occurs when the two wafers adhere, does not propagate uniformly—a phenomenon believed to underlie wafer deformation. These insights can help build models that allow us to predict how much the wafers deform and to eventually fine tune the bonding recipes. This knowledge can also help to improve overlay accuracy in another way: Designers can shift the Cu pads in the pattern design before the actual wafer bonding. These pre-bond lithography corrections allowed imec to achieve wafer-to-wafer hybrid bonding at 300nm pitch with overlay error less than 25nm for 95% of the dies, using today's most advanced bonder tools.

At the 2025 Symposium on VLSI Technology and Circuits [4], imec researchers showed the feasibility to further extend the wafer-to-wafer hybrid bonding roadmap to an unprecedented 250nm pitch (Figures 5-6). However, to achieve the required overlay accuracy

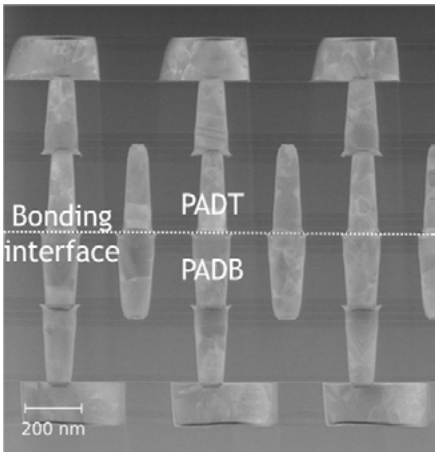


Figure 5: TEM of daisy chains (with bonded top (PADT) and bottom (PADB) pads of unequal size) on a 250nm hexagonal pad grid (as presented at VLSI 2025 [4]).

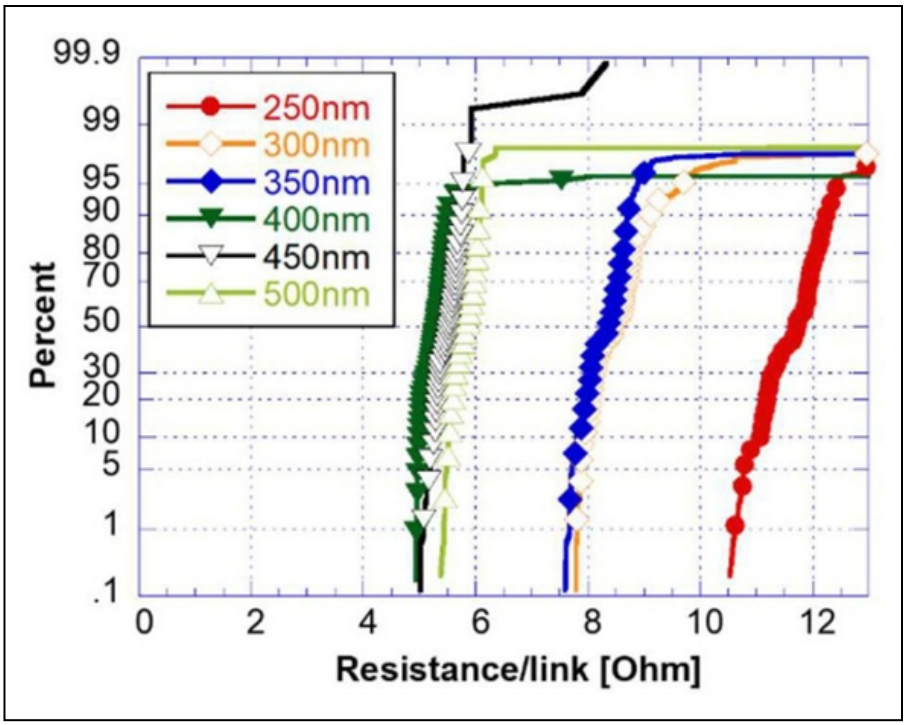


Figure 6: Electrical yield on hybrid-bonded daisy chains as a function of pad pitch for unequal pad sizes (as presented at 2025 VLSI [4]).

at industry-relevant yield over the full 300mm wafer, next-generation bonding equipment will be needed. Imec continues to work with its ecosystem of tool suppliers towards that ambitious goal.

Connecting the tier's front and backside metals with nano-through-silicon vias

In CMOS 2.0 implementations, the stacking of tiers will be much more complex than in today's industrial hybrid bonding cases. Not two, but multiple tiers will be stacked on top of each other. Most of the tiers will have metal lines on both sides—on their front as well as their backside—with an active layer (e.g., memory or logic) in between. Part of the backside metal lines may be used to power the active devices, as part of a broader BSPDN.

The sections below present recent evolutions in front-to-back connectivity technologies.

Enabling front-to-back connectivity with direct backside contacting and nano-through-silicon vias. Following this vision, tiers now have connections on both sides, with front- and backside metals connected to each other in a seamless way. This front-to-back connectivity can be realized with through-silicon vias (TSVs), at the granularity of logic or memory standard cell level. When going deeper in the system hierarchy, other front-to-back connections at finer interconnect pitch are needed, including direct backside contacting. This connectivity scheme can be used to directly connect the source/drain contact areas of advanced logic devices to the backside metal and is emerging in the logic roadmap of leading foundries.

Evolution in front-to-backside connectivity technologies must keep pace with the advancement of the wafer-to-wafer hybrid bonding roadmap, so as to offer tight pitch connections on both sides of the wafer in a balanced way (see also **Figure 2**). But combining all these technologies also brings challenges. Increasingly, more post-processing is required after the wafer-to-wafer bonding step, including wafer thinning (to support the fabrication of TSVs) and backside metal patterning. During the latter step, minimizing backside lithography distortion is crucial to ensure tight overlay between the backside metal lines and either the TSVs or source/drain contacts.

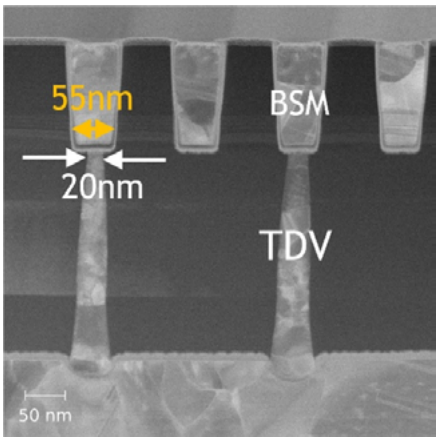


Figure 7: TEM of frontside-to-backside connections using barrier-less Mo-filled TDVs with 20nm bottom diameter (as presented at 2025 VLSI [4]).

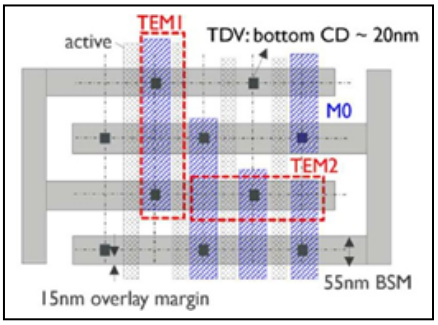


Figure 8: Layout showing 15nm overlay margin between the bottom of the TDV and the 55nm wide backside metal (as presented at 2025 VLSI [4]). (TEM1 represents the TEM cut used in **Figure 7**.)

Backside through-dielectric vias with 20nm bottom diameter. At the 2025 Symposium on VLSI Technology and Circuits Conference, imec presented progress in its nanoTSV (nTSV) roadmap, showing backside vias with a diameter as small as 20nm, at 120nm pitch (**Figure 7**) [4]. Vias with such a small diameter offer the benefit of consuming as little of the standard cell area as possible, but their fabrication requires extreme wafer thinning to ensure manageable aspect ratios.

Imec's roadmap offers several options for making nTSVs, including via-first, via-middle and via-last

integration. In addition, vias can be made with circular or slit-shaped bottoms [7], trading overlay tolerance for area consumption. In the 2025 VLSI demonstration, vias were made using a via-first approach, meaning that the vias are already patterned within shallow-trench isolation (STI) features on the wafer frontside prior to wafer thinning. The resulting through-dielectric vias (TDVs, so called because these vias run through the shallow trench isolation (STI) dielectric) were filled with molybdenum (Mo). Mo can be implemented without a barrier and offers smaller resistance than conventional Cu or W metals—benefiting both area and performance.

Connecting front- and backside with high overlay accuracy. The layout of a typical test structure showed 15nm overlay margin between the 55nm wide backside metal lines and the 20nm wide circular bottom of the Mo TDVs (Figure 8). This overlay specification can be achieved using higher order corrections per exposure in the backside metal lithography step, to compensate for the grid distortions from preceding wafer bonding and thinning steps.

In all previously discussed connecting schemes, achieving high total overlay accuracy in hybrid bonding as well as minimizing backside lithography distortion are crucial targets that both rely on the bonding process and on the capabilities of next-generation bonding equipment.

Performance and area benefits of BSPDNs in always-on and in switched-domain designs

BSPDNs are another key feature of future CMOS 2.0 architectures. With a BSPDN, the entire power distribution network is moved to the wafer's backside, where the power delivering interconnects can be made larger and less resistant. BSPDNs can therefore significantly reduce supply-voltage (or IR) drops. This facilitates designers to maintain the 10% margin allowed for the unwanted power loss that occurs between the voltage regulator and the active devices. By decoupling the power delivery network from the signal network, BSPDNs also allow decongesting the BEOL in the wafer's frontside, which can now be more efficiently designed for signal transport.

Imec pioneered the concept of BSPDN in 2019 and has meanwhile proposed several options for implementing BSPDNs [8]. Some major chip manufacturers have recently introduced the technology in their logic roadmaps, and plan to offer commercial products with advanced processors relying on BSPDNs. The technology also shows promise for 3D SoC implementations, and benefits are expected for CMOS 2.0 architectures as well. The section below discusses various aspects of BSPDNs.

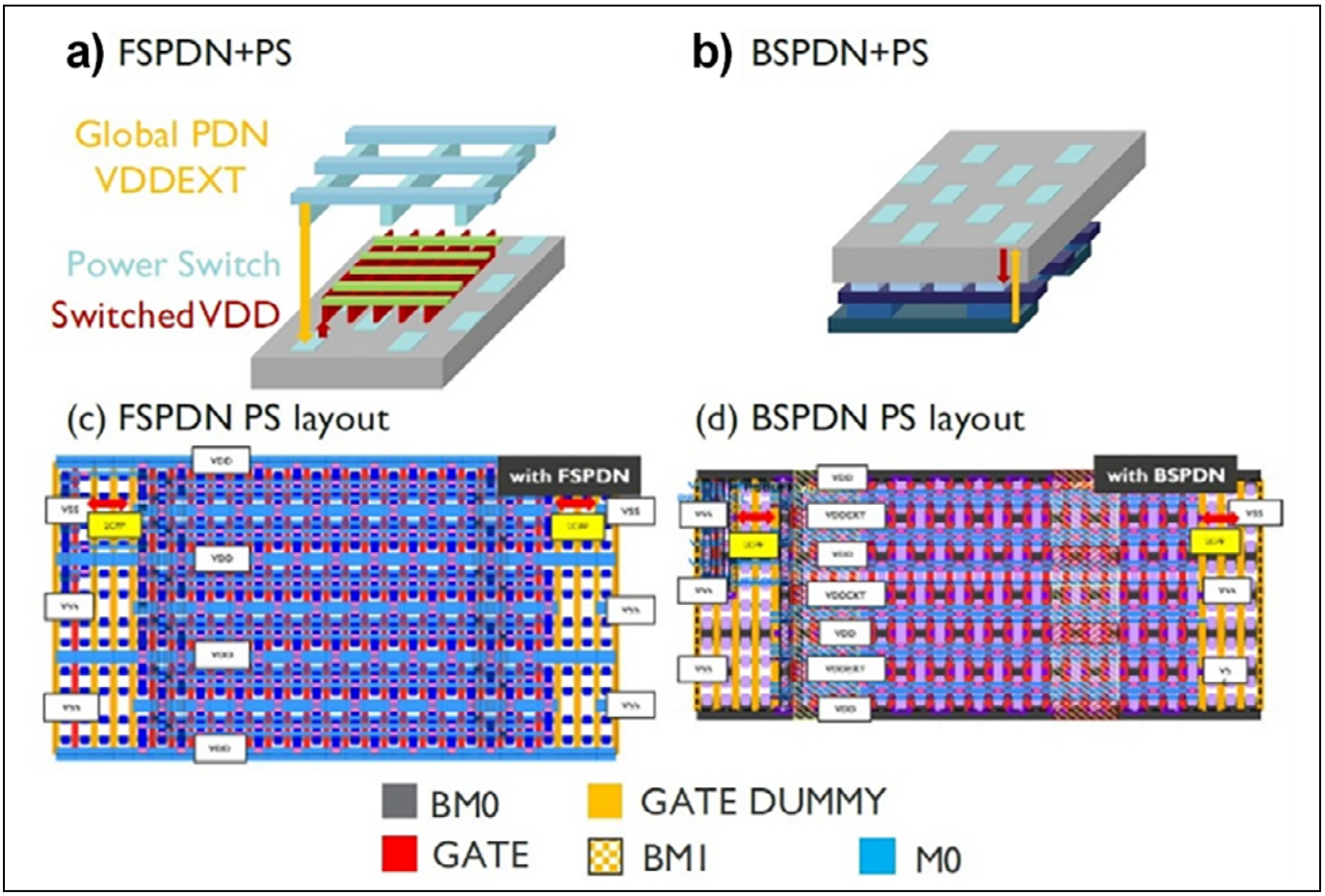


Figure 9: a-b) Power delivery of switched-domain designs with power switches placed in a checker board pattern; c-d) power switch layouts for frontside and backside PDNs (as presented at 2025 VLSI [5]). (VDDEXT=the always-on power; VDD=the switched power; PS=power switch.)

BSPDNs in always-on and switched-domain designs: Performance and area improvement over frontside implementations. In the past, imec has demonstrated the PPAC merits that BSPDNs can bring at block level, for high-density, as well as for high-drive logic use cases [8]. These benefits have been shown through design-technology co-optimization (DTCO) studies for always-on use cases, architectures where power (i.e., global V_{DD}) is continuously delivered to the active devices.

At the VLSI 2025 Symposium, imec also showed the benefits of implementing BSPDNs in switched-domain designs where blocks of standard cells are turned off for power management [5]. Switched-domain designs are realized by locally implementing power switches: Devices that distribute power (local VDD) to the transistors locally and can turn groups of standard cells on and off when needed. These designs are typically used in power-constrained applications such as mobile phones.

Imec researchers compared the impact of using BSPDNs in switched-domain designs with traditional frontside PDN implementations (**Figure 9**). The study was performed through physical implementation of a mobile compute processor design in 2nm technology. The BSPDN implementation resulted in improved performance as well as reduced area consumption compared to frontside PDN switched-domain designs. With a BSPDN, the IR drop could be significantly reduced (by 122mV). This allowed the BSPDN design to use fewer power switches and still manage an acceptable IR drop. The reduced amount of power switches takes up less core area space compared to frontside PDN implementations: A total 22% area reduction is achieved by using a BSPDN implementation.

Summary

With CMOS 2.0, a new scaling paradigm will unfold that can meet the growing diversification of compute applications. It relies on the stacking of functional tiers—each optimized using the most-suited technology (node). Fine-grain backside processing as well as fine-pitch hybrid bonding are key to enable this vision. Recent advances in wafer-to-wafer hybrid bonding motivated by SRAM partitioning and backside technologies driven by power delivery optimization bring the CMOS 2.0 concept closer to reality, offering tier-to-tier connectivity at the granularity of logic and memory standard cells. These foundational technologies will make it possible to bring heterogeneity—central to current chiplet approaches—within the SoC itself, creating more options for compute system scaling.

Acknowledgments

This work has been enabled in part by the NanoIC pilot line. The acquisition and operation are jointly funded by the Chips Joint Undertaking, through the European Union's Digital Europe (101183266) and Horizon Europe programs (101183277), as well as by the participating states: Belgium (Flanders), France, Germany, Finland, Ireland and Romania. For more information, visit nanoic-project.eu.

References

1. J. Ryckaert, "CMOS 2.0: bringing heterogeneity inside the system-on-chip," imec Reading Room.
2. J. Ryckaert, "What is CMOS 2.0?," IEEE Spectrum, Feb. 26, 2024.
3. J. Ryckaert, S. B. Samavedam, "The CMOS 2.0 revolution," Nature Reviews Electrical Engineering, March 2024.
4. L. Witters, et al., "High-density wafer level connectivity using frontside hybrid bonding at 250nm pitch and backside through-dielectric vias at 120nm pitch after extreme wafer thinning," Proc. 2025 Symposium on VLSI Technology and Circuits.
5. Y. Zhou et al., "Backside power delivery for power switched designs in 2nm CMOS: IR drop and block-level power-performance-area benefits," Proc. 2025 Symposium on VLSI Technology and Circuits.
6. S. A. Chew, "Wafer-to-wafer hybrid bonding: pushing the boundaries to 400nm interconnect pitch," imec Reading Room.
7. P. Zhao et al., "Backside Power Delivery with relaxed overlay for backside patterning using extreme wafer thinning and Molybdenum-filled slit nano Through Silicon Vias," Proc. 2024 Symposium on VLSI Technology and Circuits; pp. 1-2; doi: 10.1109/VLSITechnologyandCir46783.2024.10631441.
8. J. Ryckaert, "Backside power delivery options: a DTCO study," imec Reading Room.

Biographies

Zsolt Tokei is an imec Fellow, and incoming Program Director of 3D System Integration at imec, Leuven, Belgium. In 1998, he started working at the Max-Planck Institute of Düsseldorf, Germany, as a post-doctorate researcher. He joined imec in 1999 and, since then, has held various technical positions. First, as a process engineer and researcher in the field of copper low-k interconnects, then headed the metal section. Later, he became Principal Scientist, Program Director of nano-interconnects and more recently, transitioned to 3D interconnects. He earned an MS in Physics from the University Kossuth in Debrecen, Hungary. In the framework of a co-directed thesis between the Hungarian University Kossuth and the French University Aix Marseille-III, he obtained his PhD (1997) in Physics and Materials Science. Email Zsolt.Tokei@imec.be

Eric Beyne is a Senior Fellow, VP of R&D, and Program Director of 3D system integration at imec in Leuven, Belgium. He obtained a degree in Electrical Engineering in 1983 and a PhD in Applied Sciences in 1990, both from the KU Leuven, Belgium. He has been with imec since 1986, working on advanced packaging and interconnect technologies.

SPEA

NEXT-GENERATION SEMICONDUCTOR TESTING

DOT Test Platform

The capabilities of two traditional testers, combined in a single, test-head-only system. A true zero-footprint solution, with 100% test efficiency.

- Automotive
- PMIC
- MEMS & Sensors
- BMS
- MCU
- ADC/DAC
- Digital & Mixed Signal
- Power & Discretes
- SerDes
- Medical
- Linear
- 5G, IoT & Wireless

Discover how to save **up to 90%** on your Cost of Test
www.spea.com/savewithDOT

SPEA S.p.A.
Via Torino 16, 10088 Volpiano (TO) - Italy
Tel. +39 011 982 5400

Bridging performance and yield: The evolving role of interconnect technologies in HBM

By Damon Tsai, Woo Young Han, Tim Kryman [Onto Innovation]

As artificial intelligence (AI), high-performance computing (HPC), and advanced graphics processing continue to push the limits of memory throughput, the demand for high-bandwidth memory (HBM) has surged. These applications require faster access to massive datasets, driving the need for memory solutions that can deliver both speed and density without compromising power efficiency. HBM addresses this by stacking memory dies vertically and connecting them with ultra-fast interconnects. However, as data rates climb, traditional bump technologies—long relied upon as the primary interconnect method—are reaching their physical and performance limits.

Today, we stand at a pivotal moment in HBM interconnect technology. Hybrid bonding has emerged as one of the most talked-about technologies in advanced packaging. Promising finer pitches and superior electrical performance, hybrid bonding is generating excitement for its potential in high-performance applications. On the other hand, innovations in bump scaling are making bumps increasingly viable for a broad range of applications, beyond legacy and cost-sensitive applications. Despite these advancements, scaling bump pitch beyond 10µm and toward 2µm presents significant challenges. Maintaining uniformity and controlling bump height variation becomes more difficult, impacting yield and reliability. Meanwhile, taller HBM stacks, from 8-high to 24-high, require thinner dies, increasing the risk of die warpage and cracks during dicing.

Hybrid bonding is not without its challenges, too. At this scale, direct bonding demands plating uniformity and surface cleanliness to ensure reliable interconnects.

In this article, we will examine the challenges facing bump and hybrid bonding technologies, the solutions these two technologies enable, and how they stack up compared to each other. We also will highlight how manufacturability, reliability, and process control evolve as pitches shrink and stack heights increase. To start with, we focus our attention on bumps, and then move on to hybrid bonding.

Bump metrology evolution

Microbumps play a critical role in enabling vertically-stacked HBM structures by serving as interconnects between dies, and dies to interposers or substrates. These bumps need to be uniform in height, properly aligned, and defect free (**Figure 1**).

Inconsistent bump height in HBM can result from plating nonuniformity and process variability, and it negatively affects yield, reliability, and performance. Meanwhile, poor coplanarity can lead to mechanical

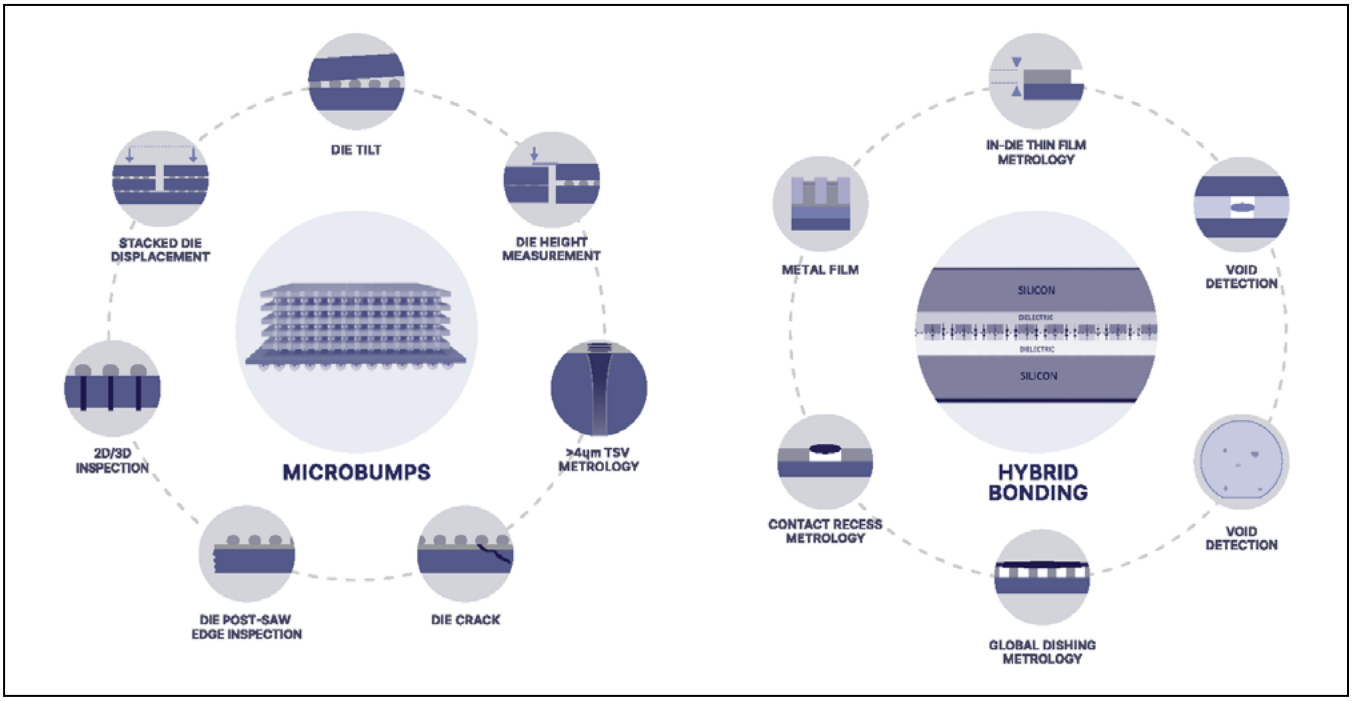


Figure 1: Interconnect challenges in microbump and hybrid bonding technologies.

Global No.1 Total Test Solution Provider

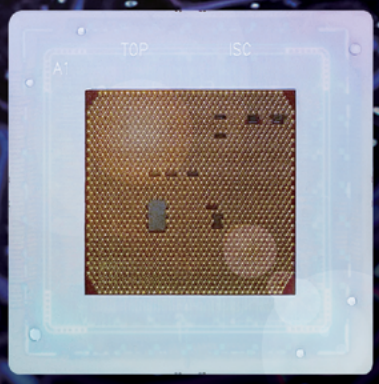
GLOBAL INNOVATION
ISC

HEAD OFFICE

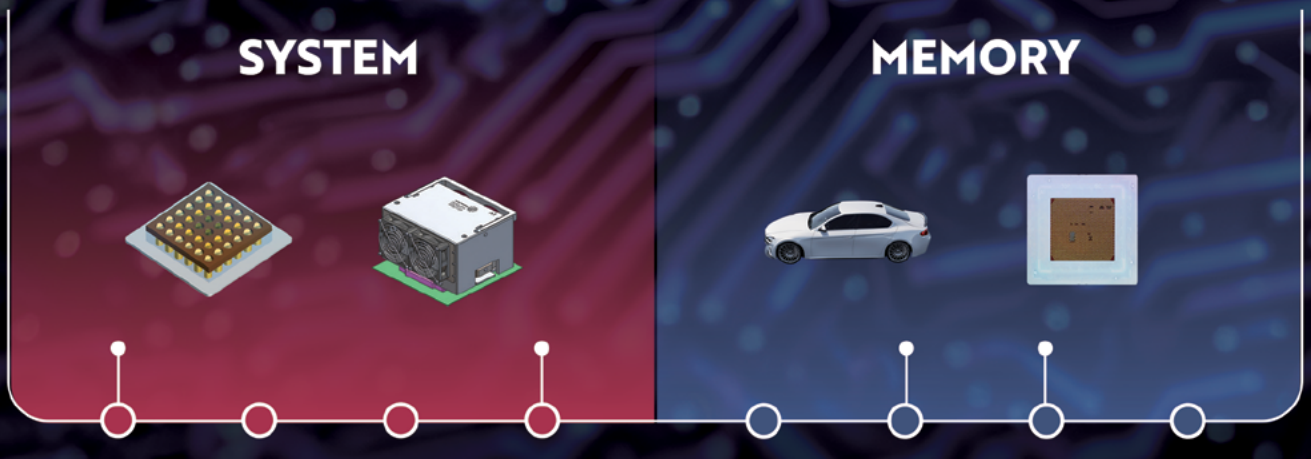
26, Geumto-ro40beon-gil,
Sujeong-gu, Seongnam-si,
Gyeonggi-do, Republic of Korea

CONTACT

sales@isc21.kr
+82 31 777 7675
www.isc21.kr



Capable of Mass Production : From Validation To SLT



stress, interconnect fatigue, or thermal cycling failures, while inconsistent contact can degrade signal integrity and power delivery. Misalignment during flip-chip bonding can result in open or short circuits. Given the breadth of these challenges, manufacturers should focus on identifying issues after the plating step and before the reflow step. After all, if you have a plating problem but move onto reflow, it will be too late to fix the problem.

With the number of layers in an HBM stack continuing to rise, addressing die warpage becomes even more essential (**Figure 2**). First of all, die warpage significantly compromises stack alignment and bonding quality. This is especially critical given the extremely tight tolerances of HBM. In addition, warpage can also result in voids, opens, and a host of other issues leading to electrical failures, mechanical stress and cracking, yield loss, and thermal performance degradation, resulting in overheating and reduced performance. Meanwhile, organic residue can result in surface contamination, voids and delamination, oxidation and corrosion, and diminished yield and reliability.

Properly identifying cracks and alignment errors in bumps poses another challenge (**Figure 3**). Cracks often occur during the dicing and backside grinding process and can break the electrical path, leading to open circuits. They often propagate due to thermal cycling, especially in materials with different coefficients of thermal expansion (CTE), weakening the bump structure. Even if initially functional, cracked bumps are prone to failure under thermal or mechanical stress during operation.

Misalignment of dies is another problem. It is often caused during the pick and place step due to the lack of backside patterns. This lack of patterning makes it difficult to tell if overlay is accurate—this is something with which automated optical inspection systems struggle.

Rise of hybrid bonding

The key reason hybrid bonding has emerged as a new technology for HBM is simple: improved interconnect density and smaller package sizes. To begin with, hybrid bonding enables finer interconnect pitches, less than 10µm, allowing for more I/O terminals in a smaller area. This increased density translates directly into higher bandwidth and improved overall performance.

Traditional bump-based stacking introduces gaps of about 30µm between dies. Hybrid bonding offers direct Cu-to-Cu connections, thereby achieving near-zero spacing between dies, significantly reducing overall package thickness and offering lower resistance and better thermal conductivity than bump-based methods. This improves signal integrity, reduces power consumption, and enhances heat dissipation, all of which are critical for HBM.

Finally, hybrid bonding supports several configurations: wafer-to-wafer, die-to-wafer, and die-to-die, offering flexibility in manufacturing and integration. This adaptability is vital for scaling HBM technologies across different applications and performance tiers.

With our introduction to bumps and hybrid bonding complete, we now turn to a comparative discussion of these two technologies, building on earlier points and exploring new ones.

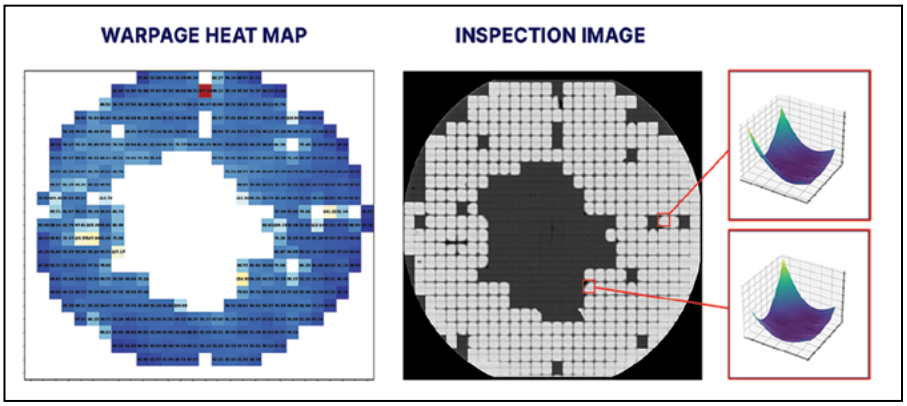


Figure 2: Die warpage measurements.

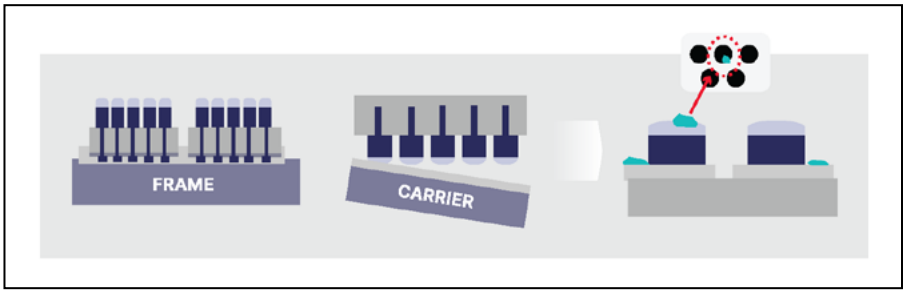


Figure 3: Organic residue post-die sawing and debonding impacts yield.

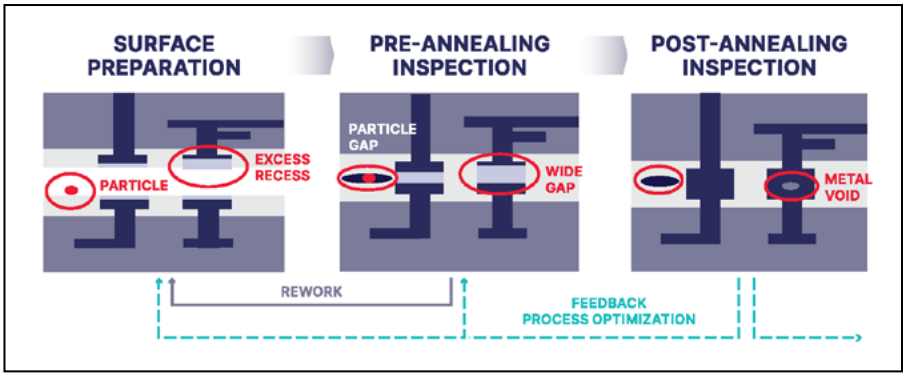


Figure 4: Hybrid bonding metrology and inspection challenges.

Bump vs. hybrid bonding

Bumps have several advantages over hybrid bonding. As a mature technology, they are widely used and well understood in the industry, are compatible with existing flip-chip and underfill processes, and are scalable for moderate density. The cost of bump technologies is lower than hybrid bonding technologies. Bumps, however, have a number of disadvantages compared to hybrid bonding. The most significant of which are the pitch limitations of bump technologies, which struggle at pitches below 10µm due to challenges in plating uniformity and solder reflow. Bumps also require underfill, which can introduce stress and complicate thermal management.

Despite the limitations noted above, bump technology continues to evolve. Leading suppliers of bump plating systems project a continued downscaling of bump dimensions, with diameters decreasing to the 4µm-5µm range and heights dropping as low as 1µm to 2µm. At a 10µm pitch, the lateral footprint remains sufficient to support the high-density I/O requirements of advanced memory architectures. Concurrently, the reduced bump height enables vertical integration of up to 16 stacked HBM dies within the 775µm maximum package height defined by Joint Electron Device Engineering Council (JEDEC) standards. Consequently, bump interconnects remain a viable and scalable solution for next-generation HBM, and major manufacturers are maintaining substantial R&D investments in bump technology.

Hybrid bonding, while promising, presents its own set of challenges. In addition to being more expensive than traditional bump technologies, hybrid bonding may require manufacturers switching from bump technology to purchase new equipment and adopt new process flows. In addition, hybrid bonding is especially sensitive to particles and organic residues; for example, even 1µm particles can cause defects. Such residues can prevent proper contact, trap gases or moisture, and lead to void formation during bonding. These voids can cause delamination or incomplete bonding, reducing mechanical integrity and electrical continuity.

Interconnect solutions

Process control for hybrid bonding is challenging, but these obstacles can be addressed by employing a suite of advanced metrology and inspection technologies, as well as analytic software solutions.

High-speed, sub-micron inspection can be used to detect surface anomalies such as particles, residues, and backside and edge defects (**Figure 4**). This ensures that bonding surfaces are clean and defect-free before the bonding process begins. As for overlay misalignment, the capability to measure wafer topography and alignment with sub-micron precision enables accurate die placement and reduces the risk of misalignment during bonding. As for voids, a non-contact, immersion-free acoustic metrology technology capable of detecting voids down to 1µm is effective at identifying bonding defects that could lead to electrical or thermal failures. Furthermore, a sub-micron inspection system can be used to detect defects like cracks and delamination caused by thermal or mechanical stress.

In the case of bump-based interconnects, the ability to measure Cu-to-Cu bump height down to 1.5µm will be key. However, several of today's technologies employ white light

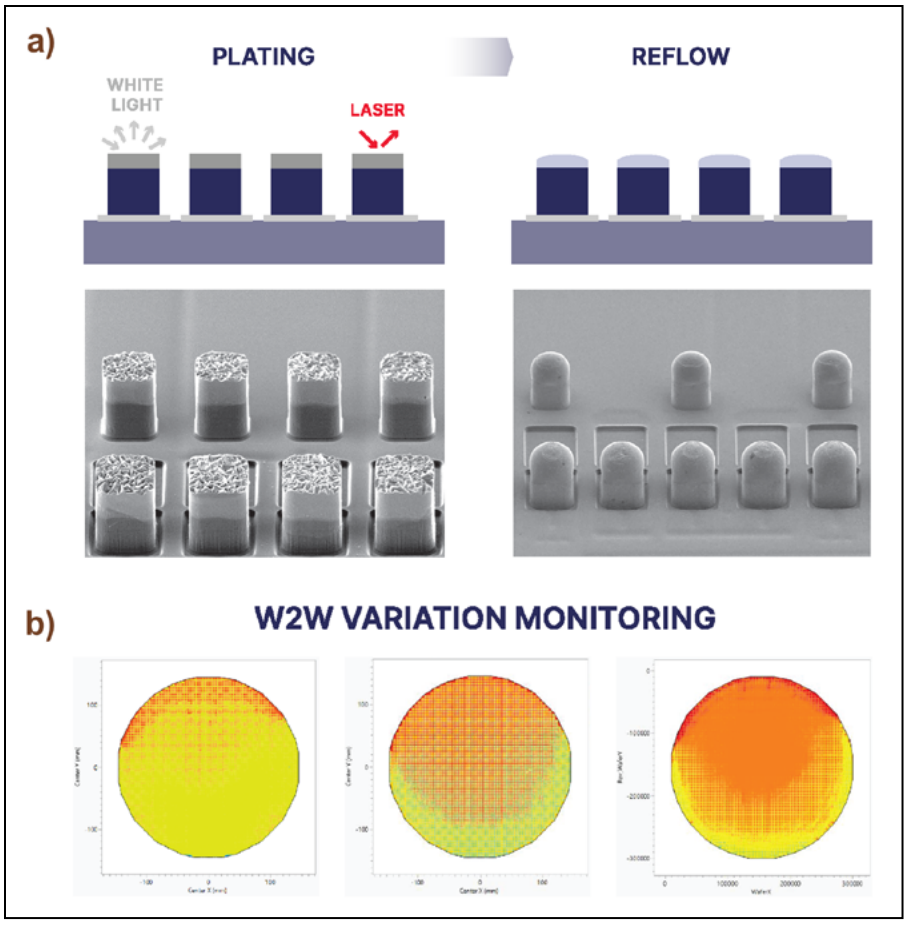


Figure 5: a) (top) Plating uniformity impacts coplanarity after reflow. Coherent laser technology overcomes roughness of the top surface, thereby overcoming the scattering noise that occurs when white light is used; and b) (bottom) Wafer-to-wafer variation monitoring.

illumination technologies that are unable to properly identify defects on these exceptionally small Cu pillars before reflow (**Figure 5**). After reflow, the bump shape is very clean and very smooth; but before reflow and after plating, the Cu is very rough. Typically, when white light hits a very rough surface, the light randomly scatters. However, an inspection system that uses a coherent wavelength laser technology can overcome this challenge.

In addition to the points noted above, an integrated system using analytical software can be used to detect errors and provide real-time defect analysis and statistical process control. Software can be used to provide in-line process control insights about defects and offer actionable corrections and is capable of handling millions of bumps per wafer, with each bump generating multiple data points. This integration helps manufacturers trace contamination sources and optimize cleaning steps, thereby improving bonding reliability.

Summary

New technologies are actively being adopted for interconnect control in HBM. In the case of bump process control, the sector is moving toward smaller dimensions, 12µm to 4µm in high-volume manufacturing, and below 2µm in R&D. Meanwhile, hybrid bonding is gaining traction due to its ability to support higher interconnect density and performance. This shift is driving the adoption of advanced inspection and metrology techniques, including an opto-acoustic void detection technology capable of detecting voids down to 1µm—a critical point for yield in HBM production.

Foundries are integrating bump inspection and metrology tools into their advanced packaging lines, especially for custom HBM and logic solutions. At the same time, hybrid bonding adoption is accelerating, particularly in the case of chiplet-based and heterogeneous integration strategies.

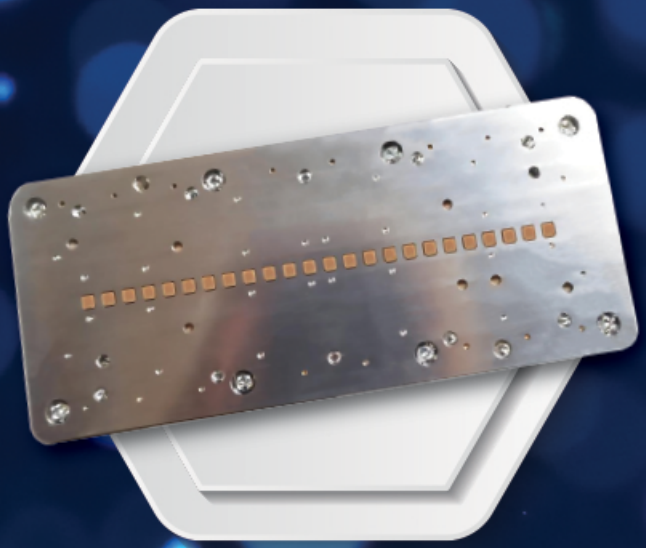
This year, R&D is expected to push microbump pitch to 6µm and hybrid bonding density to 60 interconnects/mm². In the coming years, microbump pitch may scale down to 3µm, while hybrid bonding density could reach 100 interconnects/mm². Long term, hybrid bonding is projected to become a mainstream interconnect technology, surpassing microbumps in many advanced applications, with densities reaching 150 interconnects/mm² and microbump pitch shrinking to 1.5µm.

As it stands today, there is no single solution for all HBM use cases. Instead, manufacturers may adopt an approach leveraging both bump and hybrid bonding, supported by a comprehensive process control strategy that integrates metrology, inspection, and advanced analytics.

Biographies

Damon Tsai is the Senior Director of Product Marketing, Inspection, at Onto Innovation, Milpitas, California. He joined the company in 2018 and has extensive experience in inspection and metrology, with a specialized focus on semiconductor FEOL, advanced packaging, OSAT, and specialty markets like RF, Power, and CIS. Email damon.tsai@ontoinnovation.com

Woo Young Han is Product Marketing Director, Inspection, at Onto Innovation, Richardson, Texas. He joined Onto Innovation in 2000 and holds an Electrical Engineering degree from the University of Toronto.



WLCSP SOLUTION

Min. Pitch
≥0.125mm

Pin Count
≤6000

Longevity
>1000K

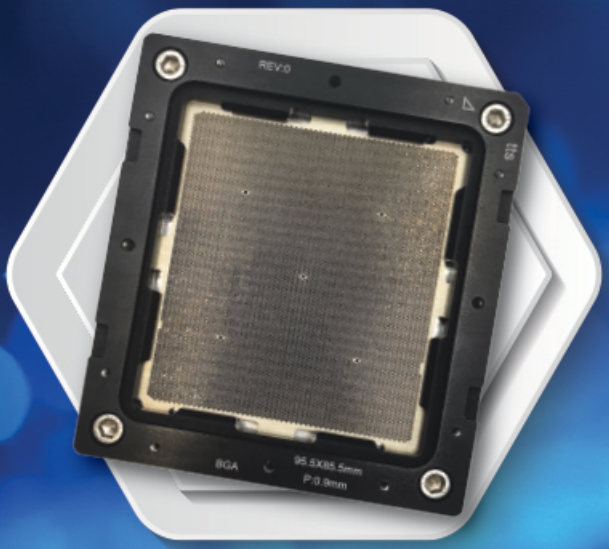
COAXIAL SOLUTION

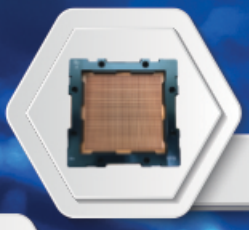
Min. Pitch
≥0.80mm

Pin Count
>10,000

HIGH SPEED
224Gbps

CROSSTALK
-75dB@56GHz





HIGH PIN COUNT SOLUTION

RF & mmWAVE SOLUTION

HYBRID ELASTOMER SOLUTION

PROBE PINS

tts

Test Tooling Solutions Group

www.tts-grp.com sales@tts-grp.com

WLP, PLP, FOWLP, FOPLP, CoWoS® (TSV-interposer), CoPoS (TGV-interposer), and advanced packaging

By John H. Lau [Unimicron Technology Corporation]

Fifteen years ago, the boundaries and tasks among the semiconductor, packaging, and carrier/PCB (printed circuit board) functions were very clear, and each properly executed the requirements of their respective functions (**Figure 1**). During TSMC's 2011 third quarter investor conference, Dr. Morris Chang (founder of TSMC), without any advance warning, shocked everybody by announcing his company would move into the packaging and testing field. The first product would be chip-on-wafer-on-substrate (CoWoS®), which integrates logic computing and memory chips by mounting them on a silicon interposer and then placing them directly on a package substrate. Today, the industry calls CoWoS® 2.5D integrated circuit (IC) integration.

Since that investor conference, the boundaries and tasks among the semiconductor, packaging, and carrier/PCB entities have become blurred and there is overlap between the semiconductor and packaging functions, and between the packaging and carrier/PCB functions, as shown in **Figure 2**. For example, it can be seen that both the semiconductor and packaging functions include work on wafer bumping, wafer-level packaging (WLP), panel-level packaging (PLP), fan-out wafer-level packaging (FOWLP), FOPLP (fan-out panel-level packaging), CoWoS®, CoPoS (chip-on-panel-on substrate), chiplets and heterogeneous integration, 2.D, 2.3D, 3D, 3.3D, 3.5D, RDLs (redistribution layers) and glass substrate/interposer, thermal management, etc. Also, both the packaging and carrier/PCB functions include work on FOPLP, organic substrate/interposer, glass substrate/interposer, chiplets and heterogeneous integration, 2D, 2.3D, RDLs, thermal management, etc.

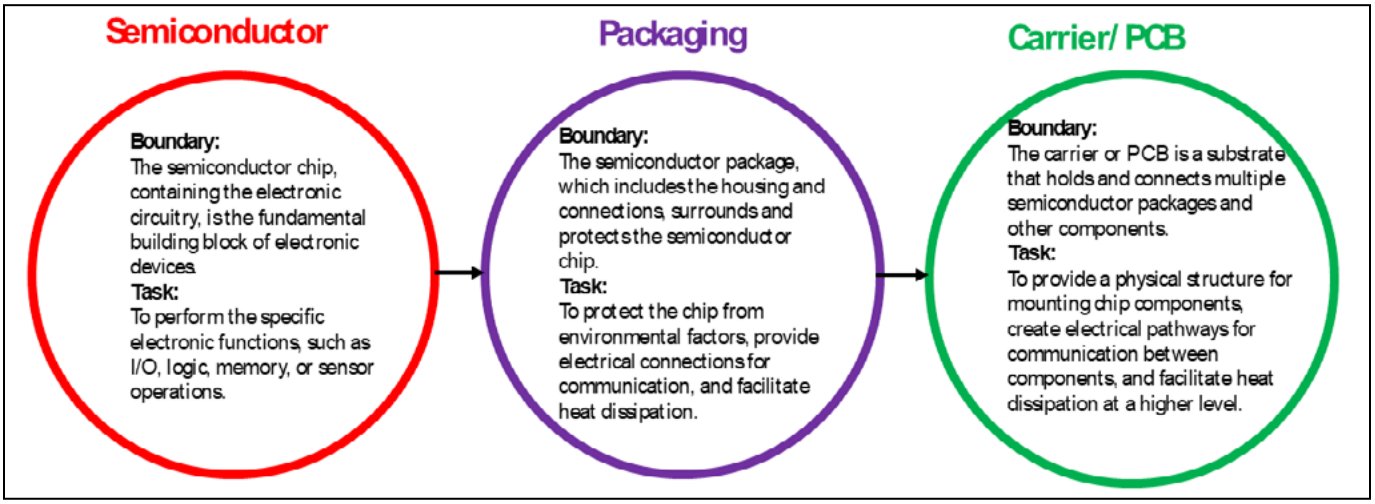


Figure 1: Boundaries and tasks among semiconductor, packaging, and carrier/PCB.

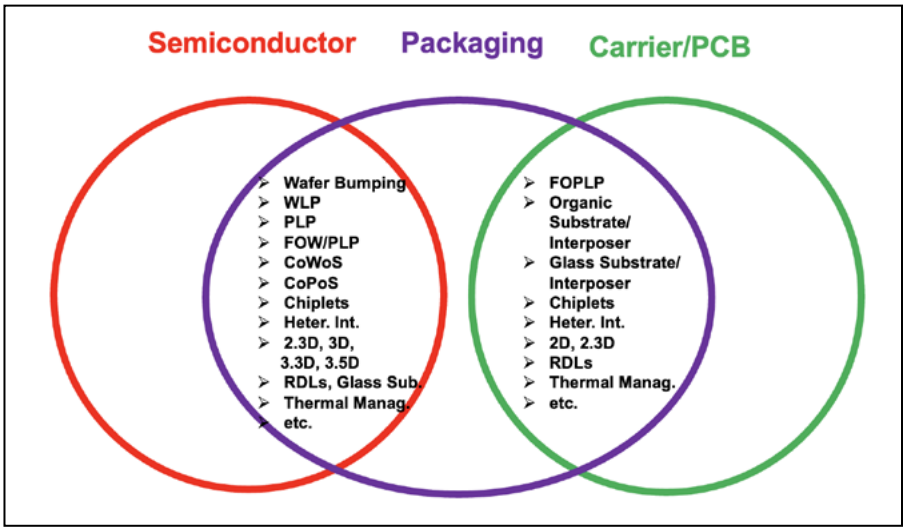


Figure 2: Overlapping boundaries and tasks among semiconductor, packaging, and carrier/PCB functions.

In this article, WLP, PLP, FOWLP, FOPLP, CoWoS®, CoPoS, and advanced packaging, will be briefly mentioned. Some recommendations will be provided.

WLP vs. PLP

WLP has been in high-volume production for more than 25 years, e.g., [1,2]. In the past few years, PLP has been gaining a lot of traction, e.g., [3-5]. The one and only reason is PLP has better area efficiency (which leads to lower costs) compared to WLP as shown in **Figure 3**. For both WLP and PLP, the materials of the substrate or interposer could be ceramic, organic, silicon, glass, fan-out RDLs, etc.

In the past 40 years, there have been (on and off) discussions on using a rectangular silicon panel to fabricate semiconductor devices (rectangular chips). However, due to a number of reasons, it did not materialize, and the robust silicon-wafer infrastructure is the key reason.

FOWLP vs. FOPLP

One of the applications of WLP is FOWLP, and one of the applications of PLP is FOPLP. **Figure 4** shows an example of FOWLP by using a temporary wafer carrier for the fan-out heterogeneous integration of four chips (one is 5mm x 5mm, and three are 3mm x 3mm) and four capacitors [6]. **Figure 5** shows an example of FOPLP by using a temporary panel carrier for the fan-out heterogeneous integration of the four chips and four capacitors [7]. After fan-out packaging, the temporary carrier (either the round wafer or rectangular panel) is removed. Today, FOWLP is in high-volume manufacturing, while FOPLP is in small-volume manufacturing.

CoWoS® vs. CoPoS

One of the applications of WLP is CoWoS®, while one of the applications of PLP is CoPoS. The area efficiency of CoPoS is better than that of CoWoS®, therefore, CoPoS technology is potentially lower cost than CoWoS®. Today, CoWoS® is the packaging technology of choice for high-performance computing (HPC) products driven by artificial intelligence (AI) [4]. TSMC has been producing CoWoS® since 2013 [4], and

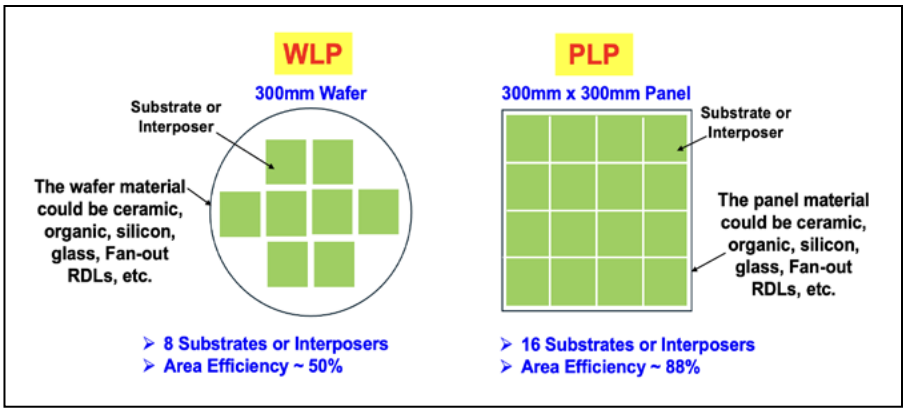


Figure 3: Substrates or interposers (not chips) fabricated on: a) A round wafer (WLP), or b) A rectangular panel (PLP).

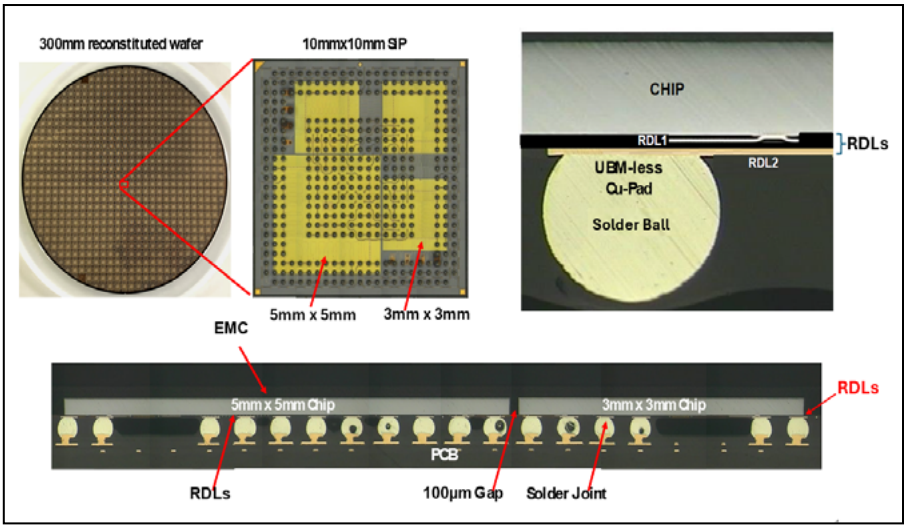


Figure 4: Fan-out wafer-level packaging for heterogeneous integration.

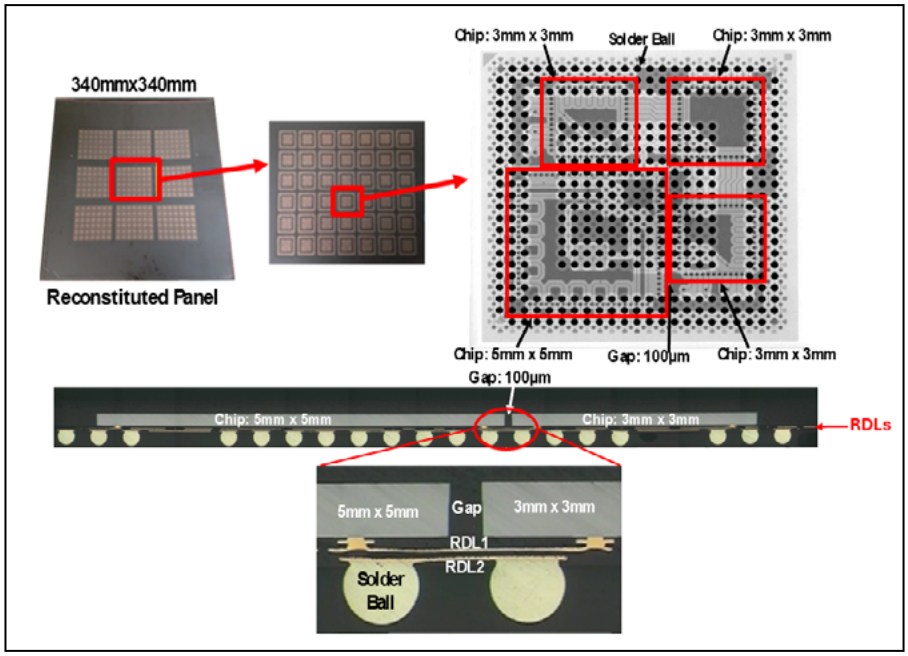


Figure 5: Fan-out panel-level packaging for heterogeneous integration.

now is the sole provider of CoWoS®. During the TSMC 2025 North America Technology Symposium (April 23), the company announced its CoPoS technology, which is competing with its CoWoS® technology.

Figure 6 schematically shows a CoWoS® package. It can be seen that the system-on-a-chip (SoC) and the high-bandwidth memories (HBMs) are supported by a through-silicon via (TSV) interposer with RDLs that are fabricated on a silicon wafer. There are at least two methods in fabricating the TSV: laser and deep reactive iron etching (DRIE). The process flow of making the TSV interposer with DRIE is shown in **Figure 7a**, and the image of the TSV cross section is shown in **Figure 7b**, [8,9].

Depending on the linewidth (L) and spacing (S) of the RDLs, there are at least two methods in fabricating the RDLs for CoWoS®, as shown in **Figures 8b-c**. **Figure 8b** shows the process flow for the conventional RDLs (L/S≥2µm) with polymers such as photo-imageable dielectric (PID) or Ajinomoto build-up film (ABF) for the dielectric layer and electrochemical deposition (ECD) Cu for the conductor layer. The polymer is spin-coated on the silicon wafer.

Figure 8c shows the other process flow for the RDLs (L/S<2µm). In this case, the RDLs are fabricated with the 64nm process technology. The dielectric layers (SiO_2) are fabricated

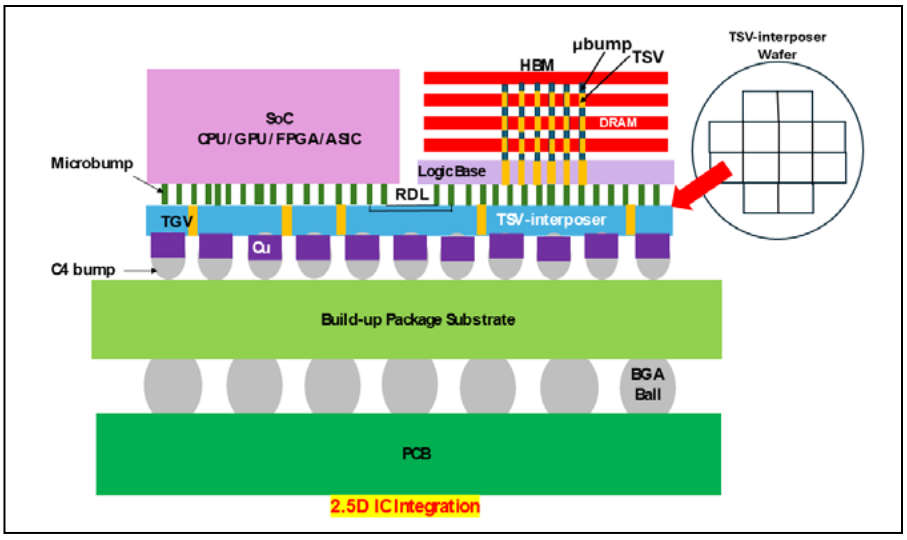


Figure 6: Packaging technology for HPC products driven by AI (CoWoS®).

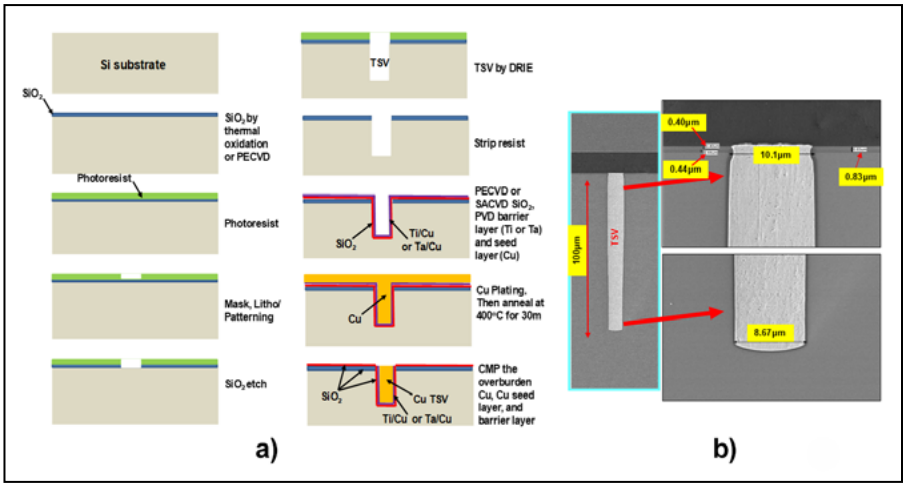


Figure 7: a) TSV process flow; b) SEM image of the cross section of a TSV.

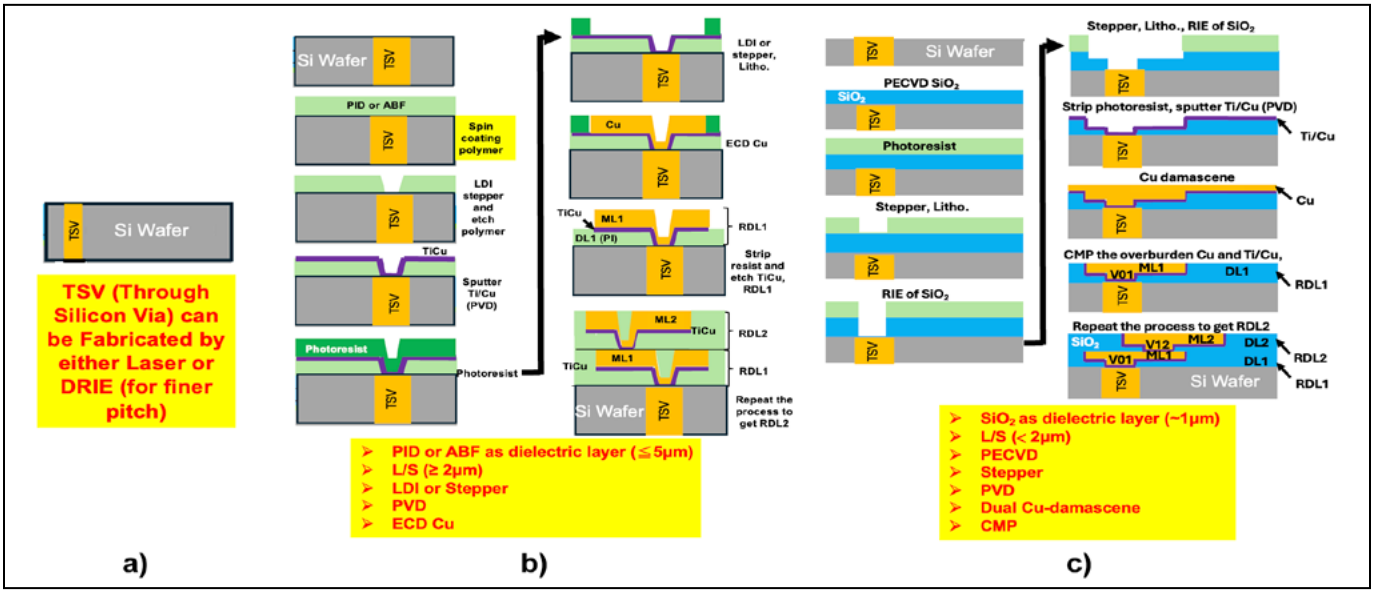


Figure 8: Fabrication of: a) TSVs; b) Ordinary L/S (≥2µm) RDLs; c) L/S (<2µm) RDLs.

by plasma enhanced chemical vapor deposition (PECVD). The metal layers are fabricated by dual Cu-damascene and chemical-mechanical polishing (CMP). For example, the scanning electron microscope (SEM) image of the TSV with RDLs is shown in **Figure 9** [8,9]. Today, in all the CoWoS®, the RDLs are fabricated by this method and the minimum pitch of the RDLs is 0.4µm.

Figure 10 schematically shows a CoPoS package. It can be seen that the organic or glass interposers (made from a panel) are supporting the SoC and HBMs. Because the objectives of CoPoS are to compete with and replace the CoWoS® package, therefore, a glass panel is used (the organic panel is not flat enough for fabricating the 0.4µm pitch RDLs). The process flow for fabricating the through-glass via (TGV) is shown in **Figure 11a** and the flow for the RDLs is shown in **Figures 11b-c**. **Figure 11b** shows the process flow for the conventional RDLs (L/S≥2µm) with polymers such as PID or ABF for the dielectric layer, and ECD Cu for the conductor layer. The polymer is slit-coat on the glass panel. **Figure 11c** shows the process flow for the RDLs (L/S<2µm), which is the same as that for the CoWoS® package (**Figure 8c**), except for the panel materials and structures. CoWoS® is built on a silicon wafer, while CoPoS is built on a glass panel. Again, in order to compete with the CoWoS®, PECVD + dual Cu-damascene + CMP for the RDLs of the CoPoS package are a must. According to TSMC's announcement (April 23, 2025), its glass panel size is 310mm x 310mm for the CoPoS® package. In the future, it could be changed to 510mm x 515mm, or 600mm x 600mm.

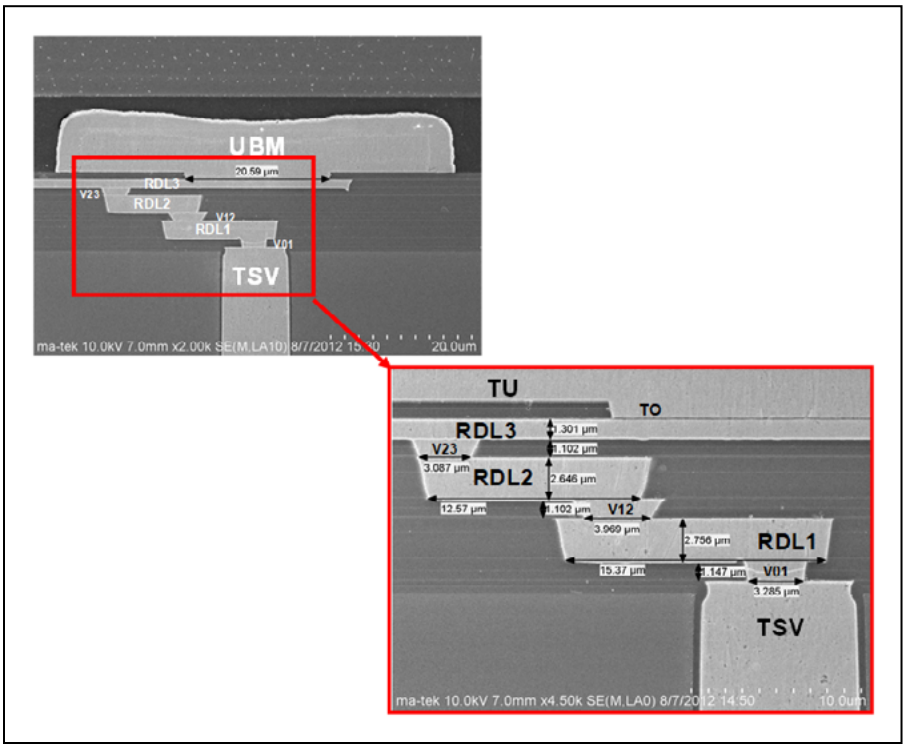


Figure 9: SEM image of a TSV fabricated by DRIE and RDLs by dual Cu-damascene + CMP.

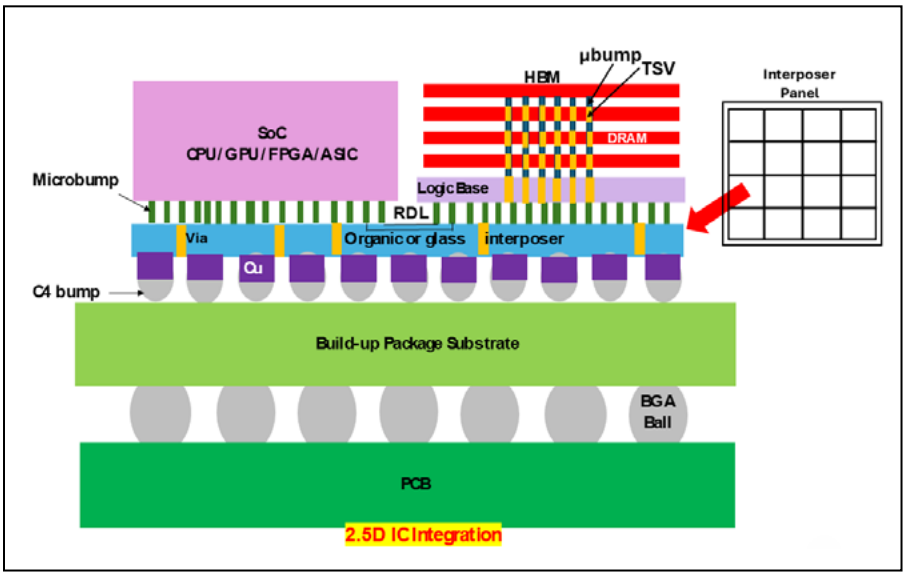


Figure 10: Packaging technology for HPC products driven by AI (CoPoS - organic or glass interposer panel).

Advanced semiconductor packaging

In [4], advanced semiconductor packaging, such as 2D, 2.1D, 2.3D, 2.5D, and 3D IC integration, has been discussed. In this study, we would like to update [4] to include 3.3D and 3.5D IC integration (**Figure 12** [10]).

3.3D IC integration

Because of the ever-increasing size of the TSV interposer of the CoWoS® package and the TGV interposer of the CoPoS package, the semiconductor manufacturing yield loss of the large-size TSV interposer (or TGV interposer) is becoming unsustainable [10]. Most recently, TSMC and Hynix have been working on HBM4 with hybrid bonding and placing the HBM4 on top of the SoC so the size of the TSV-interposer (or TGV interposer) can be smaller. On the other hand, NVIDIA and Hynix are working on HBM4 with hybrid bonding and placing the HBM4 on top of the SoC, and then placing the SoC directly on top of the build-up packaging substrate (the TSV-interposer or the TGV interposer are eliminated) [10].

During the Samsung Foundry Forum & SAFE™ in Santa Clara (June 12-13, 2024), Samsung announced, basically, the same thing as TSMC, NVIDIA, and Hynix (**Figure 13**). Samsung

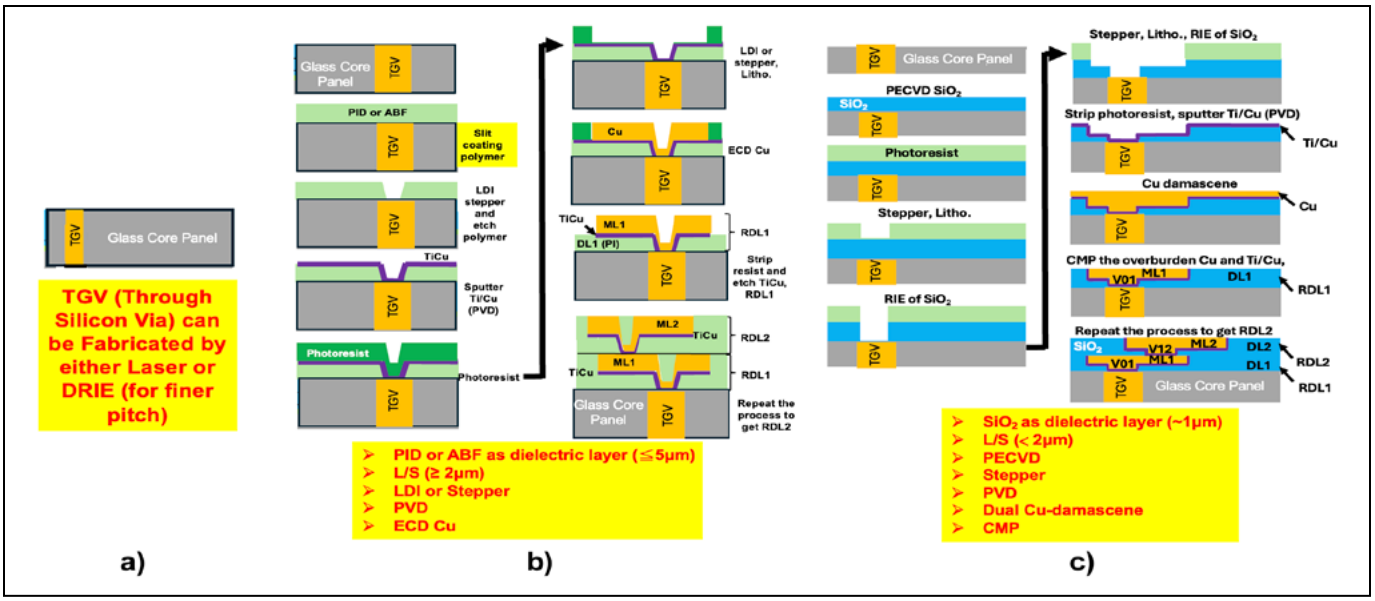


Figure 11: Fabrication of: a) TGVs; b) Ordinary L/S (≥2μm) RDLs; c) Fabrication of the L/S (<2μm) RDLs.

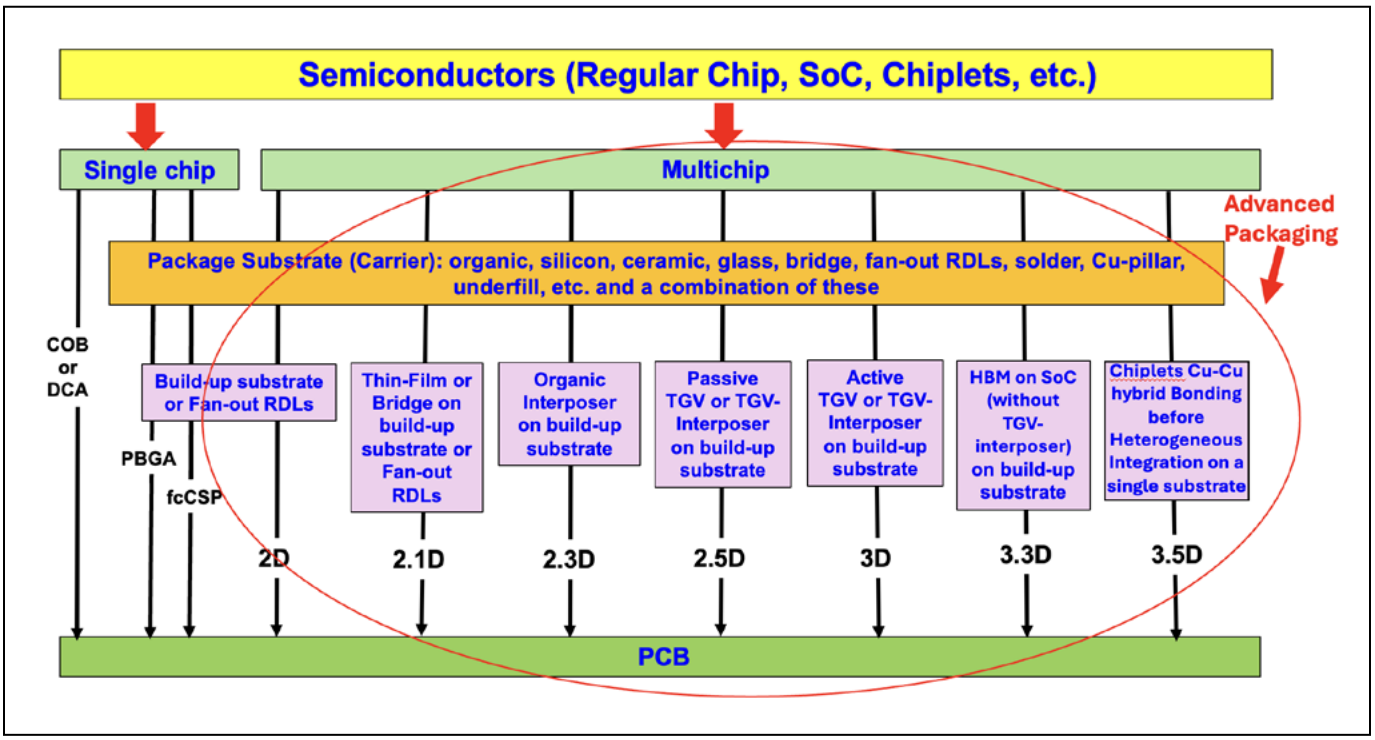


Figure 12: Advanced semiconductor packaging.

has a platform called Samsung advanced interconnect technology (SAINT) that includes the SAINT-D for dynamic random access memory (DRAM) stacking on top of logic chips like central processing units (CPUs) or graphics processing units (GPUs). Samsung's new 3D packaging method involves stacking HBM chips vertically on top of processors, which differs from the existing 2.5D IC integration technology that connects HBM chips and GPUs horizontally via a TSV interposer. This vertical stacking approach eliminates the need for the TSV interposer, but requires a new base die for HBM memory that is made using a sophisticated process technology. For the future, the vertical stacking of GPUs (artificial intelligence [AI] computing chips) and last-level cache (LCC) form a single unit interconnected with HBM memory. A silicon bridge chip directly connects the die, and a transparent medium replaces the TSV interposer in the copper RDL redistribution layer—what Samsung calls the all-in-one heterogeneous integration, or 3.3D IC integration [10].

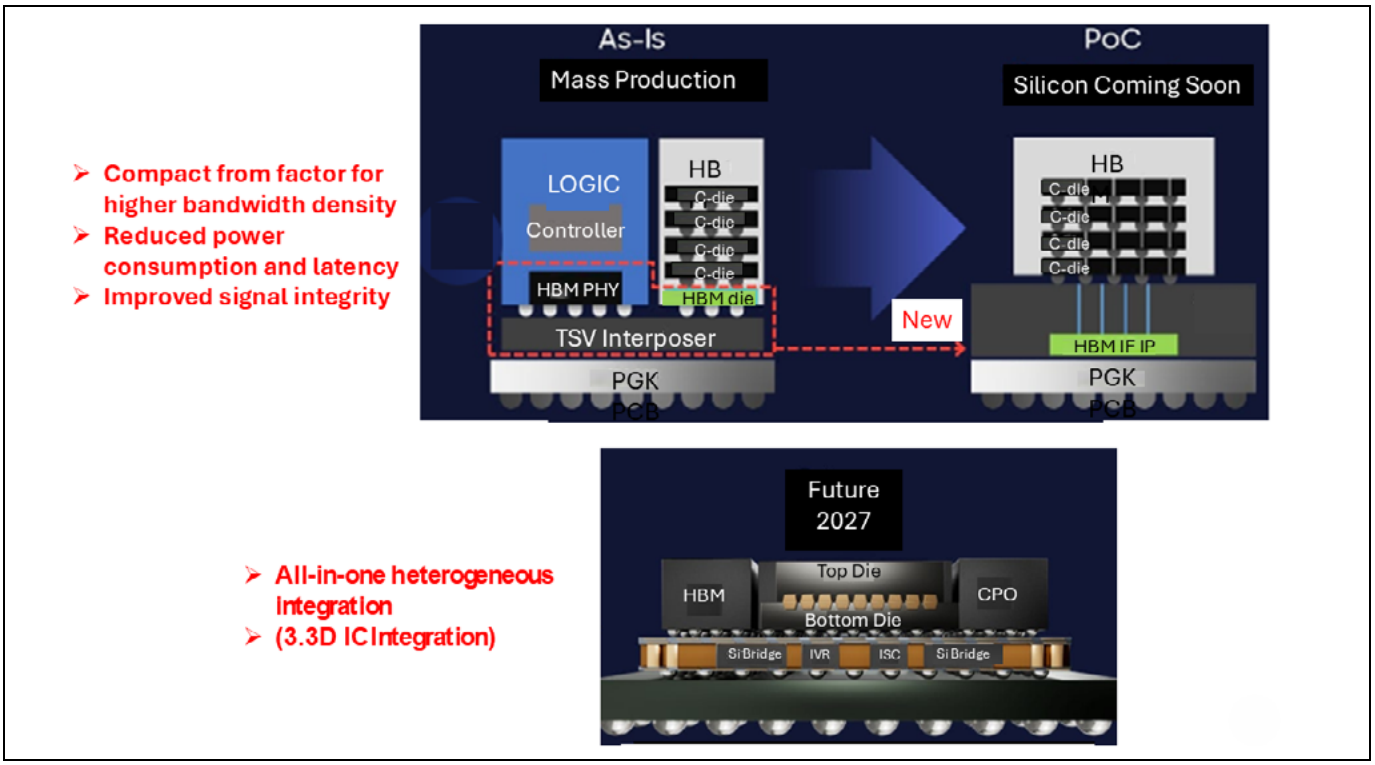


Figure 13: 3.3D IC heterogeneous integration.

Figure 14 shows a new 3.3 IC integration structure [11]. It can be seen that the interconnects between the DRAMs and logic base of the HBMs are Cu-Cu hybrid bonding [4,12-14]. The HBMs are then (Cu-Cu) directly attached to the top surface of the SoC. Finally, the SoC with HBMs are Cu-Cu directly attached to the top surface of the glass substrate.

3.5D IC integration

The sections below discuss various 3.5D IC integration platforms.

Unimicron's 3.5D IC integration. As mentioned in [15], frontend integration of some of the chiplets (before package heterogeneous integration) with Cu-Cu hybrid bonding [4,12-14] can yield a smaller package size and a better performance. **Figure 15** shows an example of Cu–Cu hybrid bonding between some chiplets before they are attached to the TSV interposer [15].

Samsung's 3.5D IC integration. **Figure 16** shows Samsung's extremely large 3.5D heterogeneous integration for the next-generation packaging technology [10,16]. It can be seen that the large application-specific IC (ASIC) is Cu-Cu hybrid bonded to

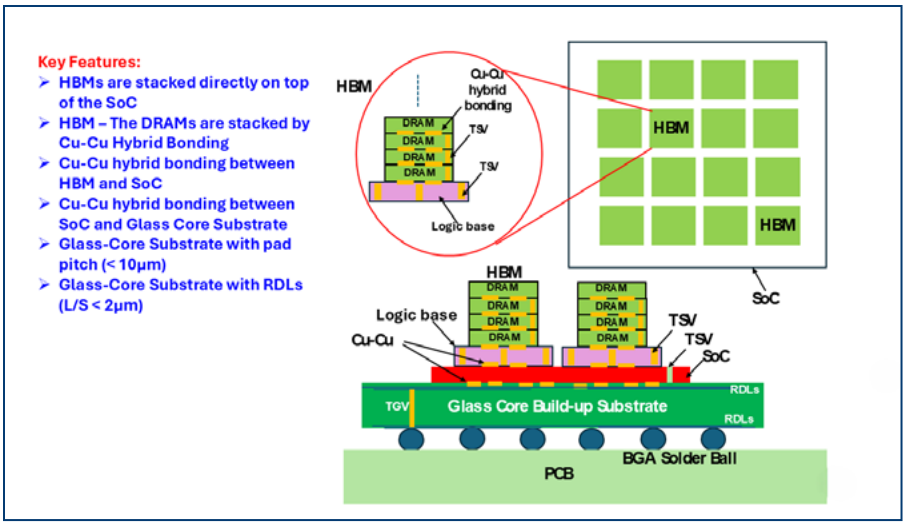


Figure 14: 3.3D IC heterogeneous integration on a glass substrate.

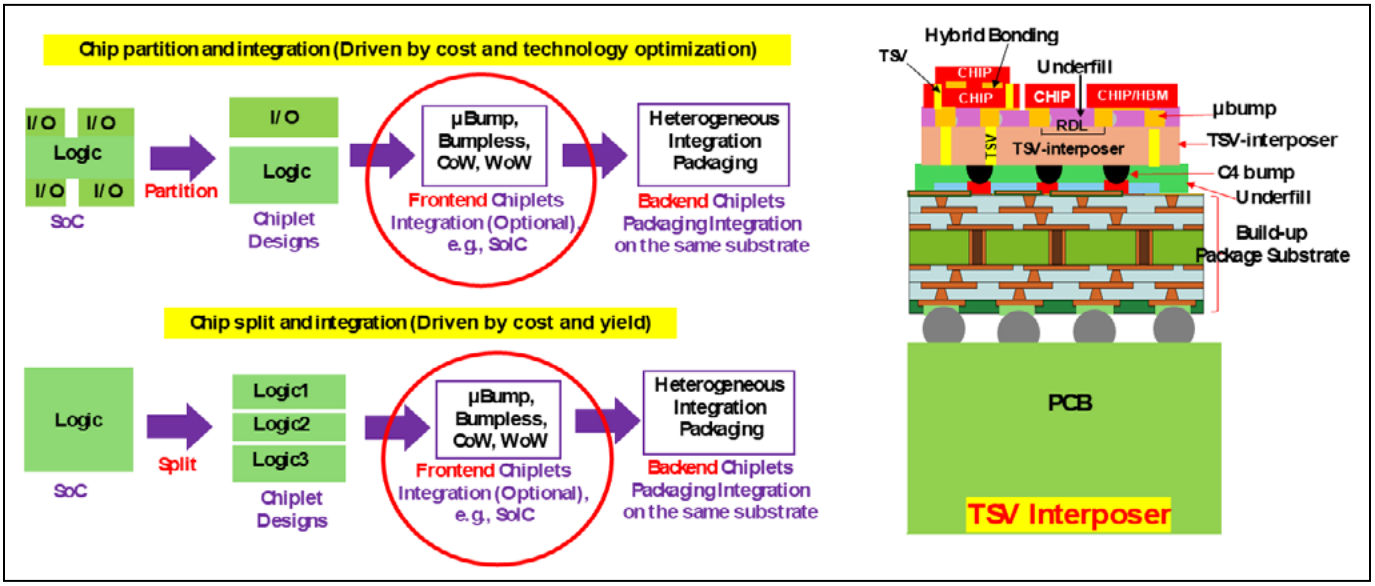


Figure 15: Hybrid bonding of chiplets (frontend integration) before heterogeneous integration packaging (3.5D IC heterogeneous integration).

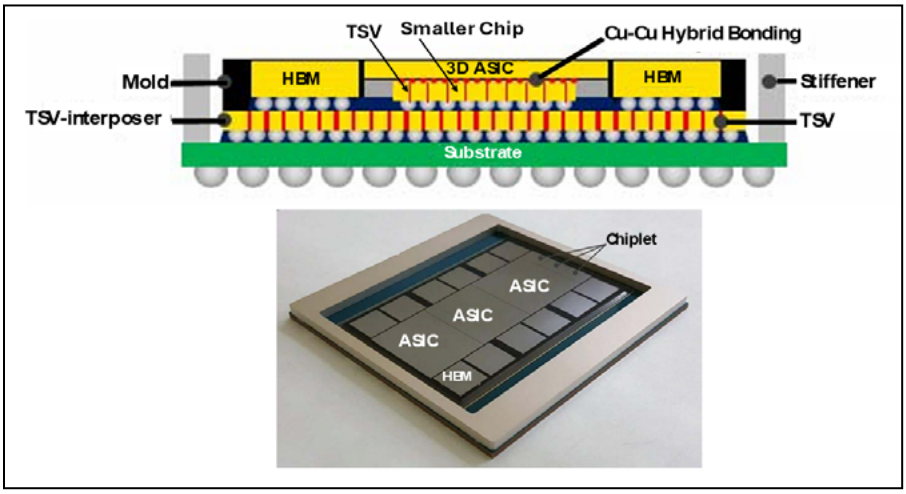


Figure 16: 3.5D IC heterogeneous integration.

another chip before they are micro-bumped on the TSV-interposer (this is the front-end integration shown in **Figure 15**).

AMD's 3.5D IC integration. **Figure 17** shows AMD's compute and AI chiplets, which are stacked on top of the I/O and cache chiplets in the company's MI300 chip. All the compute chips are built using TSMC's N5 process. Neither the I/O functions, nor the system's cache memory, benefit from N5, so AMD chose a less-expensive technology (N6) for those. Therefore, those two functions could then be built together on the same chiplet. MI300 stacks (with hybrid bonding by TSMC's system on integrated chips (TSMC-SoIC®)

platform) the front-end optional part of (**Figure 16**) three CPU chiplets (called compute complex dies, or CCDs, using AMD's "lexicon") and six accelerator chiplets (XCDs) on top of four input-output dies (IODs), on top of a piece of TSV-interposer that links them together to eight stacks of high-bandwidth DRAM HBM that ring the "super chip." AMD calls this 3.5D IC integration [10,17].

Broadcom's 3.5D XDSiP™. Broadcom has introduced its 3.5D eXtreme Dimension System in Package (3.5D XDSiP™) platform for ultra-high-performance processors driven by AI workloads. This platform

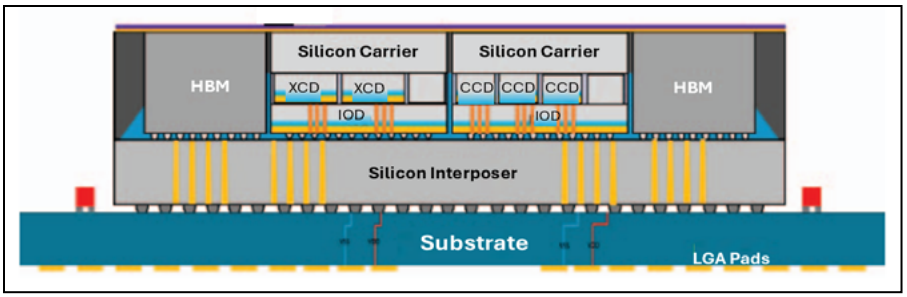


Figure 17: 3.5D IC integration.

uses TSMC's CoWoS®-L packaging technology that offers a maximum interposer size of approximately 5.5 times that of a reticle (about 858mm^2), or 4719mm^2 for compute chiplets, I/O chiplets, and up to 12 HBM3/HBM4 packages. To maximize performance, Broadcom suggests disaggregating the design of compute chiplets and stacking one logic chiplet on top of another in a face-to-face (F2F) manner using Cu-Cu hybrid bonding, **Figure 18** [10,18].

Summary

Some important results and recommendations are summarized as follows:

- In the past 15 years, the boundaries and tasks among the semiconductor and the packaging, and the packaging and carrier/PCB, have been blurred and overlapping. Both semiconductor and packaging sectors are working on such efforts as wafer bumping, WLP, PLP, FOWLP, FOPLP, CoWoS®, CoPoS, chiplets and heterogeneous integration, 2.D, 2.3D, 3D, 3.3D, 3.5D, RDLs, glass substrate/interposer, thermal management, etc. The packaging and carrier/PCB sectors are working on such efforts as FOPLP, organic substrate/interposer, glass substrate/interposer, chiplets, and heterogeneous integration, 2D, 2.3D, RDLs, thermal management, etc.
- PLP has a much better area efficiency (which leads to lower costs) than WLP.
- FOWLP is one of the applications of WLP, while FOPLP is one of the applications of PLP. The uniqueness of FOWLP and FOPLP is that they both need a temporary carrier (a wafer carrier for FOWLP, and a panel carrier for FOPLP).
- Today, FOWLP is in high-volume production, while FOPLP is in low-volume production.
- The minimum pitch of the CoWoS® RDLs is 0.4µm, which is fabricated using 64nm process technology (PECVD + dual Cu-damascene + CMP).
- The CoPoS platform is meant to compete with, and replace, the CoWoS® platform. Therefore, the CoPoS RDLs must be fabricated

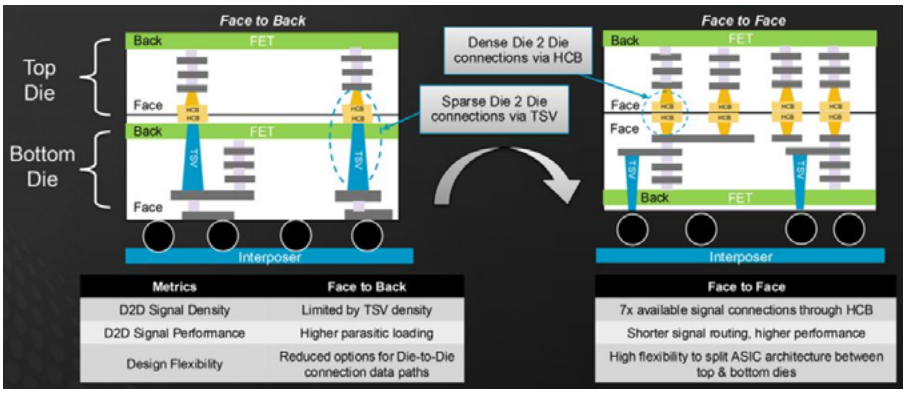


Figure 18: F2F and face-to-back (F2B) 3.5D IC integration.

using 64nm process technology with PECVD + dual Cu-damascene + CMP. Because the organic interposer is not flat enough for the fabrication of the 0.4µm-pitch RDLs, a glass interposer should be used. Today, TSMC's glass panel size for making the RDLs of the CoPoS is 310mm x 310mm. In the future, it could be changed to 510mm x 515mm, or 600mm x 600mm.

- The trend in HPC products driven by AI is to redesign the HBM (DRAMs and logic base) and SoC platforms so the whole module (HBM on top of the SoC) can be attached directly to a package substrate (without the TSV interposer or the TGV interposer), i.e., 3.3D IC integration. The package substrate could be made from a glass panel.
- To have a higher performance and smaller-size package substrate, stacking some of the chiplets by Cu-Cu hybrid bonding before heterogeneous integration of all the device components on a common substrate (3.5D IC integration) is recommended.

References

1. J. H. Lau, R. Lee, *Chip Scale Package: Design, Materials, Process, Reliability, and Applications*, McGraw-Hill, New York, 1999.
2. J. H. Lau, *Flip Chip, Hybrid Bonding, Fan-In, and Fan-Out Technology*, Springer, New York, 2024.
3. J. H. Lau, *Fan-Out Wafer-Level Packaging*, Springer, New York, 2018.
4. J. H. Lau, *Semiconductor Advanced Packaging*, Springer, New York, 2021.
5. J. H, Lau, *Chiplet Design and Heterogeneous Integration Packaging*, Springer, New York, 2023.
6. J. H. Lau, M. Li, M. Li, T. Chen, I. Xu, X. Qing, et al., "Fan-out wafer-level packaging for heterogeneous integration," IEEE Trans. on CPMT, Vol. 8, Issue 9, Sept. 2018, pp. 1544-1560.
7. C. Ko, H. Yang, J. H. Lau, M. Li, M. Li, C. Lin, et al., "Chip-first fan-out panel-level packaging for heterogeneous integration," IEEE Trans. on CPMT, 2018, Vol. 8, Issue 9, Sept. 2018, pp. 1561-1572.
8. J. H, Lau, *Through-Silicon Via (TSV) for 3D Integration*, McGraw-Hill, New York, 2013.
9. J. H. Lau, P. Tzeng, C. Lee, C. Zhan, M. Li, J. Cline, et al., "Redistribution layers (RDLs) for 2.5D/3D IC integration," IMAPS 2013 Proc. Also, IMAPS Trans., Jour. of Microelectronic Packaging, First Quarter 2014, pp. 16-24.
10. J. H. Lau, X. Fan, *Hybrid Bonding, Advanced Substrates, Failure Mechanisms, and Thermal Management for Chiplets and Heterogeneous Integration*, Springer, New York, 2025.
11. J. H. Lau, M. Ma, T. Lan, T. Tseng, "3.3D IC integration for HPC Applications Driven by AI," U.S. patent application Date: May 2025.
12. J. H. Lau, "Current advances and outlooks in hybrid bonding," IEEE Trans. on CPMT, Vol. 15, No. 3, April 2025, pp. 651-681.

13. J. H. Lau, "State of the art of Cu-Cu hybrid bonding," IEEE Trans. on CPMT, Vol. 14, No. 3, March 2024, pp. 376-396.
14. J. H. Lau, "Recent advances and trends in Cu-Cu hybrid bonding," IEEE Trans. on CPMT, Vol. 13, No. 3, March 2023, pp. 399-425.
15. J. H. Lau, "Recent advances and trends in chiplet design and heterogeneous integration packaging," ASME Trans., Journal of Electronic Packaging, published online: June 2023, and hard-copy: March 2024, pp. 010801-1 – 31.
16. I. Lee, S. Nam, S. Kim, S. Shin, Y. Kim, S. Seo, et al., "Extremely large 3.5D heterogeneous integration for the next-generation packaging technology," IEEE/ECTC Proc., June 2023, pp. 893-898.
17. C. Mandalapu, C. Buch, P. Shah, R. Topacio, P. Cheng, L. Wang, et al., "3.5D advanced packaging enabling heterogeneous integration of HPC and AI accelerators," IEEE/ECTC Proc., May 2024, pp. 798-802.
18. https://www.tomshardware.com/tech-industry/artificial-intelligence/broadcom-unveils-gigantic-3-5d-xdsip-platform-for-ai-xpus-6000mm2-of-stacked-silicon-with-12-hbm-modules (Dec 7, 2024).

Biography

John H. Lau is a Senior Special Project Assistant at Unimicron Technology Corporation, Taoyuan City, Taiwan (ROC). He has more than 40 years of R&D and manufacturing experience in semiconductor packaging, 532 peer-reviewed papers (a principal investigator on 380), 53 issued and pending U.S. patents (a principal inventor on 34), and 24 textbooks on semiconductor packaging. He is an ASME Fellow, IEEE Fellow, and IMAPS Fellow. He earned a PhD from the U. of Illinois at Urbana-Champaign. Email John_Lau@unimicron.com

iMAPS 2025 SAN DIEGO

58th INTERNATIONAL SYMPOSIUM on MICROELECTRONICS

SEPT 29 - OCT 2, 2025 ▪ SAN DIEGO, CA ▪ IMAPS2025.org

The **58th International Symposium on Microelectronics** is organized by the International Microelectronics Assembly and Packaging Society (IMAPS) and held in San Diego, California. The event offers one of the most robust programs for microelectronics and advanced packaging technical content and will feature:

- **3 Days of Packaging**
- **5 Tracks with 20 Sessions**
- **125+ Speakers**
- **100+ Exhibitors**
- **19 Professional Development Courses**
- **3 Panels and Interactive Poster Session**

Keynote Speakers

Glen Daves
NXP

Subramanian Iyer
UCLA

Tarek Ibrahim
Intel Foundry

Hemath Dhavaleswarapu
AMD

Symposium 2025 Welcome Reception on the USS Midway!

This truly unique event welcomes in Symposium 2025 on Monday, September 29.

The USS Midway is a historical naval aircraft carrier museum located at Navy Pier in San Diego. Experience a ship tour, dinner and optional flight simulator, as well as some awesome networking!

SPONSOR & EXHIBIT SALES UNDERWAY!

For sponsorship information, contact
Brian Schieman, Executive Director, bschieman@imaps.org

Conference:
Sept 29-Oct 2, 2025
Exhibition:
Sept 30-Oct 1, 2025
Professional Development Courses:
Sept 29, 2025

Co-located with
Semiconductor Thermal Management Conference 2025

ThermCon
SEMICONDUCTOR THERMAL MANAGEMENT
October 2-3, 2025 | San Diego, CA

Conference Hotel

Town & Country Resort
500 Hotel Cir. N
San Diego, CA 92108

Register Today! **IMAPS2025.org**

34th SWTest Conference and EXPO: Sold Out with a Record Attendance!

By Jerry Broz, PhD *[General Chair, SWTest US and SWTest Asia Conferences, and VP at Delphon Industries]*

Semiconductor Wafer Test (SWTest) Conference and EXPO in Carlsbad, California (June 2-4, 2025), is the premier annual technical event exclusively focused on the complex challenges and infrastructure behind wafer-level-test and probe technologies. The Conference and EXPO was a resounding sold-out success with one of the highest attendances in its long history.

The two-and-a-half-day conference welcomed 698 international attendees from 18 countries with an excellent technical program that included a Visionary Keynote, 8 themed podium sessions with 28 outstanding peer-reviewed presentations, two technical poster sessions, and a sold-out EXPO featuring 70 global suppliers. A truly international gathering, with more than 45% of the attendees coming from the Asia and European wafer test communities, the 34th SWTest brought together probe professionals and technologists to share knowledge, learn from colleagues, and informally network with other experts in a relaxed, friendly setting.

At Monday's plenary session, the Visionary Keynote was delivered by Raffaele Vallauri, who serves as the Vice President Director at Technoprobe S.p.A. – Italy. Mr. Vallauri made a thought-provoking presentation entitled, "Mastering Emerging Probing Challenges: The Power of Agile and Continuous Innovation." He explored how agile and lean development methodologies can transform probe card development, enabling vendors to iterate faster, collaborate earlier with customers, and deliver high-quality solutions in record time for driving the next wave of semiconductor innovation.

Over the next two-and-a-half days, each session of the program focused on the many key challenges faced by probe technologists in low- and high-volume testing environments. From the podium and poster presentations delivered during the conference, the Technical Program Committee was pleased to award the Best Overall Presentation to Emmett Ricks (Micron Technology - USA) for his insightful work discussing "How High-Speed Probe Cards Accelerate Time-to-Market."

The Best Data Presentation was awarded to the collaborative effort from Hsu Hao Chang, Amit Agnihotri, Andrew Yick, Dongwon Lee (Marvell - USA) and Giulia Rottoli (Technoprobe - Italy) on "First Silicon Photonics High Speed (up to 67GHz) Wafer Probe Card Demonstration for S-Parameter Testing on the Production Wafer."

During his Visionary Keynote Session on Monday, Raffaele Vallauri, Vice President Director at Technoprobe S.p.A. – Italy, challenged the attendees to follow lean development methodologies that would enable technologists to collaborate earlier with customers and iterate faster to deliver the high-quality solutions driving the next wave of semiconductor innovation.

SWTest 2025 Poster Sessions, which featured 12 outstanding posters from international professional attendees and universities, were a constant buzz of networking, socializing, and connecting between technologists, colleagues, and key suppliers.

Emmett Ricks (Micron Technology - USA) received the Best Overall Presentation Award from Jerry Broz (SWTest General Chair) for his team's innovative work entitled "How High-Speed Probe Cards Accelerate Time-to-Market."

Out of the session focused on the Pursuit of High Parallelism, the Best Presentation-Tutorial in Nature was awarded to Oscar Lee and Harvey Lin (TSMC - Taiwan) who discussed "Effective Solutions for Large Format Testing Challenges in High Parallelism HPC Device Testing." Dr. Edgar Hepp coauthored with Wabe Koelmans, Francesco Colangelo, Patrik Schürch (Exaddon - Switzerland) and Francois Gix, Sebastien Schoene (Synergie CAD - France) to receive the Most Inspirational Presentation award for their joint work entitled, "Ultra-low Leakage Probe Card for Wafer Parametric Testing Enabled by µ3D Printing." The Best Poster Presentation was awarded to Takaharu Ohyama, Shigeki Oka, Tomohisa Hoshino, Yasushi Watanabe (Yokowo, Co. - Japan) for their work on "Probing Test System with Optical Fiber Array for Photonic Integrated Circuits." The "People's Choice" Award (which was voted on by attendees in real-time via the mobile-app) went to "Wafer-Level Testing of Photonic Devices," as presented by Philipp Dietrich, Andrés Machado, Florian Rupp, Roman Zvahelsky (Keystone Photonics - Germany). The SWTest Program Committee is also proud to announce that *Chip Scale Review*, the SWTest Media Partner, will select one of the presentations from the 2025 Program for publication as a full article in an upcoming issue.

698 registered attendees enjoyed the SWTest EXPO 2025, which featured a total of 70 booths from the platinum, gold, and silver sponsors, as well as international and domestic companies–many of which were participating for the first time.

On Tuesday morning, Karen Armendariz and the SWTest Executive Team reviewed the SWT-Crew mentoring initiative, which connects women technologists

with experienced professionals for career development purposes. Additionally, student participation at SWTest is facilitated by the William Mann Student Travel Grant Program, which receives partial funding from the annual charity golf tournament.

The SWTest EXPO, which was open on Monday and Tuesday evenings, was a constant buzz of networking, socializing and connecting with colleagues, suppliers, and customers. At the SWTest Conference, the EXPO and the technical program do not compete, so the attendees can easily attend both events during the conference. Our 70 exhibit booths brought together 30 sponsors and 40 exhibitors that play integral roles in the wafer probe industry and support infrastructure. This year's EXPO included a record number of new multinational companies that were participating for the first time.

Throughout the event schedule, attendees are provided with substantial opportunities to engage with key exhibitors and to network during extended breaks, daily meals, and evening social or hospitality functions. Feedback from SWTest 2025 has been highly positive, and as General Chair, I am pleased that our conference and EXPO offer a diverse range of meaningful technical and professional interactions. I would also like to take this opportunity to thank the sponsors, exhibitors, authors, speakers, session chairs, committee, and the SWTest Team members, whose deep commitment made this year a fantastic and record-setting conference and EXPO.

Looking ahead in 2025, mark your calendars for the upcoming 6th SWTest Asia Conference and EXPO that will be held November 20 to 22, 2025 at the Hilton Sea Hawk in Fukuoka, Japan. We invite you to participate in this Asia-centric probe technology forum, where you will have the opportunity to learn about the latest industry advancements and network with peers. Additionally, the event will feature up to 70 prominent international suppliers from the probe and wafer test community. Registration is already open at https://www.swtestasia.org/ and discounted registration fees are available until October 10, 2025. The SWTest Conference and EXPO in San Diego and in Asia provide valuable opportunities for all wafer-level-test industry professionals, from end-users to suppliers, engineers, and sales and marketing staff.

ADVERTISER INDEX

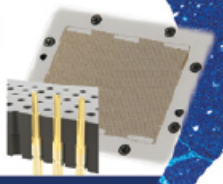
LEENO

www.leeno.com

92㎛ Pitch~
200K Pins in 7 Days!

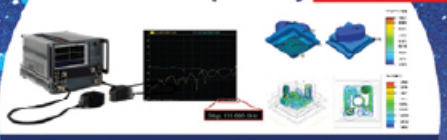
Spring Contact Probe

10K Pins in 7 Days!
Probe Head Pin Counts 30K
2 Probe Heads in 20 Days!
92㎛ Pitch~



Probe Head

110GHz Actual Measurement Capability
Report in 3 Days!

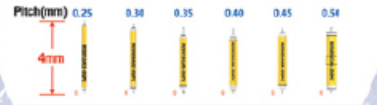


RF/Thermal Simulation
Actual Measurement
2-Port: 110GHz
4-Port: 43GHz

10 Sockets in 7 Days!
Logic & SLT Socket

AI

Specification
Frequency : 80GHz(BGA), 100GHz(QFN, LGA)
Pitch : 0.25mm~
Crosstalk : -60dB
Impedance : 50Ω±10%
1K Pins in 7 Days!



Coaxial Probe 100GHz

5 Sockets in 10 Days!
100GHz Coaxial Socket
Excellent RF & Electrical Performance!

MP Socket

Pin Counts 5K~
Device Power ~1,200W

10 Sockets in 10 Days!
Large Device Socket/Lid

LEENO LEENO INDUSTRIAL INC.

GLOBAL LEADER LEENO

HEAD OFFICE
10 105beon-gil MieumSandan-ro Gangseo-gu, Busan Metropolitan City, Korea

CONTACT
hskang@leeno.co.kr / +82 10 8737 6561
kimmj@leeno.co.kr / +82 10 5314 6560

Amkor Technology®

Unleash the Future with S-Connect™ Technology

Amkor's S-Connect technology transforms semiconductor packaging with high-density integration, modular design, and compact form factors.

Ideal for high-performance computing, edge devices, AI, and data centers, S-Connect significantly reduces power consumption, supporting energy-efficient computing and driving sustainable technological advancements.

Watch our latest video.