Chip Scale Review®

Accelerating 3D and heterogeneous integration with high-volume D2W hybrid bonding – April1, 2023

Accelerating 3D and heterogeneous integration with high-volume D2W hybrid bonding

By Thomas Uhrmann [EV Group] and Nelson Fan [ASMPT]

The semiconductor industry is undergoing a revolutionary transformation with the adoption of heterogeneous integration and chiplet-based design, marking a fundamental turning point. Monolithic 2D scaling options often come with complex and costly issues and limited scaling benefits for a system. Chiplets, therefore, are an inevitable solution to meet the demands of the scaling roadmap and performance, power, area-cost and time-to-market (PPACt) requirements. High-performance applications, including artificial intelligence (AI), augmented/virtual reality, and autonomous driving, require specialized processors for each task, making chiplet integration necessary. This design approach is already being used in various forms, from hybrid bonding to 2.5D interposers, and is equally critical for consumer and mobile devices to keep up with performance and flexibility requirements.

The shift to chiplet integration, however, requires a complete overhaul of the semiconductor manufacturing process. While 2D transistor scaling remains relevant, the rising costs and complexity of scaling have prompted the industry to embrace 3D and heterogeneous integration. This approach involves assembling and packaging different components or dies with varying sizes and materials into a single device or package, thereby enhancing performance on new device generations that support these new applications and leading to more precise and customized mapping of customer and application requirements.

There are two different chiplet approaches: partition and add-on. Which approach is used depends on the application and purpose. The partition scheme involves breaking down the original monolithic die into two or more smaller chiplets and stacking them on top of each other in a 3D-integrated circuit (IC) configuration. In contrast, the add-on scheme uses a base die as one of the chiplets with little to no partitioning and adds another chiplet (or multiple chiplets) with additional features, such as extra memory. The add-on chiplets are stacked above or below the original monolithic chiplet in the 3D-IC configuration [1]. The partition scheme is focused more on cost and footprint savings, while the add-on method is geared towards performance and power improvement, such as for high-performance computing applications. Cost savings are greater as the monolithic 2D die area increases and wafer costs become more expensive. In both cases, savings are optimized when the partitioned two chiplets have the same size because it improves the yield for each die individually. In addition to cost savings, partitioning is also expected to lead to effective capacity improvement due to higher yields in smaller chiplets. Technology considerations of 3D-ICs and the various component flavors of 3D-ICs are depicted in Figure 1.

Heterogeneous integration relies heavily on wafer-to-wafer (W2W) hybrid bonding, which involves stacking and electrically connecting wafers from different production lines. This process has proven successful for complementary metal-oxide semiconductor (CMOS) image sensors and various memory and logic technologies. W2W hybrid bonding has been mature for over a decade, with equipment and process now well established. It enables contact pitch of less than 1µm in production, but die size and grid matching are required. Each bonding layer consists of only one node, and cumulative yield decreases the overall stack yield for high layer count. However, W2W bonding offers high-throughput capabilities.

Die-to-wafer (D2W) hybrid bonding is a relatively new technology, and its process and equipment maturity are still evolving, resulting in many challenges. The contact pitch for this bonding method is currently at 9µm in production, but this is expected to decrease rapidly to 2µm. One advantage of D2W bonding is that there are no limitations on die size or system segmentation. Additionally, chiplets of different nodes can be combined, providing a high level of flexibility. However, binning may be necessary because of the varying yields of individual dies. The throughput of D2W bonding is dependent on the size of the chiplets and the number of chiplets integrated into a system.

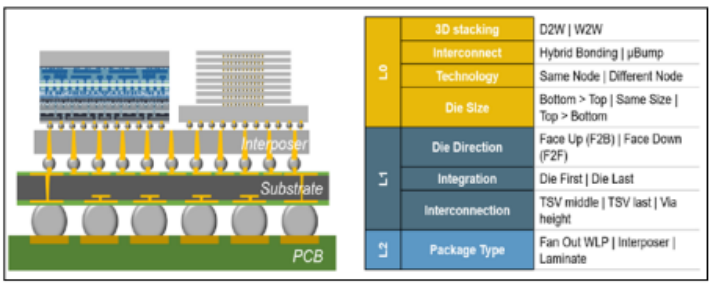

Figure 1: Heterogeneous integration and connection options along different packaging levels from chip- to board-level.

There are several D2W bonding methods available for heterogeneous integration, each with its own advantages and disadvantages, as shown in Table 1. Selecting the best approach for a given application depends on factors such as die size, thickness, total stack height, and interface considerations like contact design and density.

	Hybrid W2W Bonding	Hybrid D2W Bonding
Maturity	Wafer bonding equipment and process have been mature since 2016	Process and equipment maturity is starting to yield but still many difficulties
Contact Pitch	<1µm pitch is enabled in production	Currently 9µm pitch in production
Die Size	Die size and grid matching required	No limitations in die size and system segmentation
Segmentation	Each bonding layer consist of one node	Each chiplet can consist of a different node
Yield	Cumulative yield of each bonded layer	Cumulative yield can be avoided by binning
Throughput	>25 bonds per hour	Related to chiplet size and amount of chiplets per system

Table 1: Comparison between W2W and D2W hybrid bonding according to main decision criteria.

Collective D2W bonding process

Collective D2W bonding involves the bonding of multiple dies onto a wafer substrate in a highly accurate and reliable manner. The collective D2W bonding process typically consists of several key steps, shown in Figure 2, including carrier preparation with adhesive, die protection while handling, die population using a high accuracy D2W bonder, W2W die transfer of the carrier wafer to the product wafer, and finally, debonding of the die carrier and cleaning [2,3].

The first step in the process involves preparing the carrier wafer with a suitable adhesive material. The adhesive layer should be uniform and have a sufficient thickness to provide adequate bonding strength. The carrier wafer should also be compatible with the adhesive and should have a surface that can be easily cleaned and prepared for bonding.

In the next step, the dies are protected while being handled to prevent damage or contamination. This may involve the use of specialized handling equipment or techniques, such as vacuum or tweezers. The dies should be handled with care to avoid any potential damage, which can result in yield loss and reduced device performance.

Once the dies are protected, they are populated onto the carrier wafer using a high-accuracy D2W bonder. The bonder should be capable of achieving sub-micron alignment accuracy, and should also be able to handle a high volume of dies for efficient production. The next step involves transferring the dies from the carrier wafer to the product wafer. This is typically achieved through W2W bonding using high-precision alignment and bonding equipment. The bonding process should be performed under controlled conditions to ensure uniformity and reliability. After the dies have been transferred to the product wafer, the die carrier is debonded and removed. This involves separating the adhesive layer from the carrier wafer and cleaning any residual adhesive from the product wafer. The cleaning process should be carefully controlled to avoid damage to the dies or product wafer.

Direct placement D2W hybrid bonding

Direct die placement D2W hybrid bonding involves picking and placing a die onto a target wafer, followed by annealing to covalently bond the dies and electrically connect them [4]. The first step in this process is selecting a suitable die carrier, as shown in Figure 3. Depending on the requirements of the application, the carrier can be a film frame or a specially-designed and fabricated die carrier. The selection of the carrier should be based on factors such as die size, die thickness, and the number of dies to be bonded.

The next step is the carrier picking, which can be done from a completed singulated wafer or a reconstituted carrier consisting of different dies. Once the die has been picked, it is necessary to activate and clean the surface before bonding. Plasma activation is used to remove any contaminants and rehydrate the surface of the die, ensuring good bonding. The plasma activation process is critical to ensure strong bonding between the die and the wafer.

A high-accuracy pick-and-place process is essential for ensuring a high alignment accuracy with less than 200nm on opposite corners of the die. This high-accuracy pick-and-place process is achieved using advanced equipment and technologies. The controlled bond wave is initiated by contacting the die center, ensuring a stable and uniform bond between the die and the wafer.

	Co-D2W
Transfer Method	Collective die transfer by reconstituted carrier
Pros	• Proven technology • Die activation and cleaning equivalent to W2W hybrid bonding • Oxide management • Rework on carrier feasible
Cons	• Error propagation of D2W + W2W alignment • Cost of carrier prep, utilization and clean • Die thickness needs to be in narrow range
Maturity	Limited volume production proven for several years

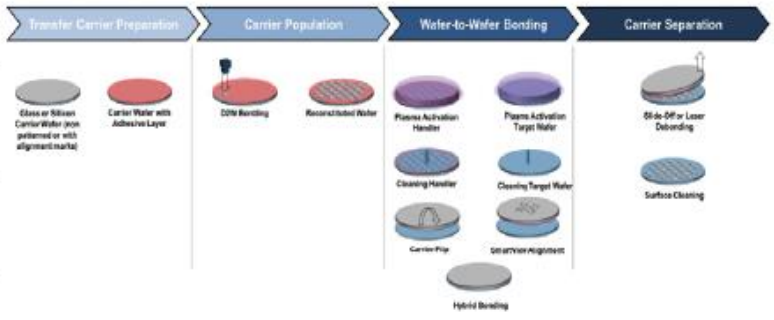

Figure 2: Collective D2W hybrid bonding process flow.

	DP-D2W
Transfer Method	Direct placement of activated dies using flip-chip bonder
Pros	• Versatile method • Die thickness invariant
Cons	• Bonding interface needs to be touched • Die handing especially for multi die stacks such as SRAM, DRAM • Particle management during die placement
Maturity	Scaling and in low-volume production

Figure 3: Direct placement D2W hybrid bonding process flow.

Reconstructed D2W hybrid bonding

A recent publication introduces a novel integration method called reconstructed D2W bonding (Figure 4) that combines direct and collective placement D2W bonding [5]. The direct placement approach is used to mechanically attach the dies to a carrier wafer, but it only provides a mechanical connection.

Once the chiplets are formed, they are permanently attached to a carrier wafer with the die either facing up or down, and the gap between them is filled with silicon oxide, which is a front-end compatible version of fan-out wafer-level packaging (FOWLP) that is inorganic. The challenge in this step is that the oxide thickness needs to be significantly higher to support the overall chiplet height, including the silicon substrate and metal interconnects. Afterward, the dies and oxide layer are planed thoroughly, and through-dielectric interconnects and hybrid bond pads are created at the wafer level. The actual hybrid bonding and electrical contact are done later in a W2W bonding process. Cleanliness must be strictly maintained throughout the process.

The major advantage of this process is its full front-end fab compatibility, and there are no materials that are incompatible with the fab present throughout the process flow. However, one of the main challenges is controlling and optimizing the oxide fill process. Analogies to FOWLP apply, where silicon content, die thickness, deposition temperatures, and oxide properties all impact the wafer shape and contact pitch scalability of the W2W hybrid bonding process. Despite this challenge, reconstructed D2W bonding has shown great promise in achieving full front-end fab compatibility while maintaining the highest level of cleanliness throughout the entire process flow. Further research is needed to optimize the oxide fill process and to address other potential challenges that may arise.

Self-aligned D2W integration

Currently, research institutes are focusing on self-aligned die bonding, which follows similar key steps to traditional die bonding such as cleaning and activation of the dielectric interface and copper bonding pad. Two guiding principles are being explored for self-aligned die bonding: 1) shaping the die to achieve ultra-precise dimensions; and 2) defining guiding pads on the die surface using hydrophilic and hydrophobic regions patterned with optical lithography at the wafer level. However, the singulation, cleaning, and activation processes still require the same level of precision as traditional die bonding. The placement of dies in self-aligned die bonding can be coarser, leading to potentially lower process costs, but the increased complexity in die preparation, guiding pad definition and combination with crucial chemical mechanical planarization (CMP) processes, must also be considered. Self-aligned die bonding has high potential, but further research and development is necessary to improve the integration process [6].

D2W bonding equipment status

The preparation and conditioning of surfaces for direct placement fusion and hybrid bonding of dies on wafers is a critical step. Challenges related to

	Reconstructed W2W
Transfer Method	W2W hybrid bonding using anorganic filled reconstructed wafer(s)
Pros	• Proven technology • Die activation and cleaning equivalent to W2W hybrid bonding • Oxide management • Rework on carrier feasible
Cons	• Error propagation of D2W + W2W alignment • Cost of carrier prep, utilization and clean • Die thickness needs to be in narrow range
Maturity	Scaling and in low-volume production

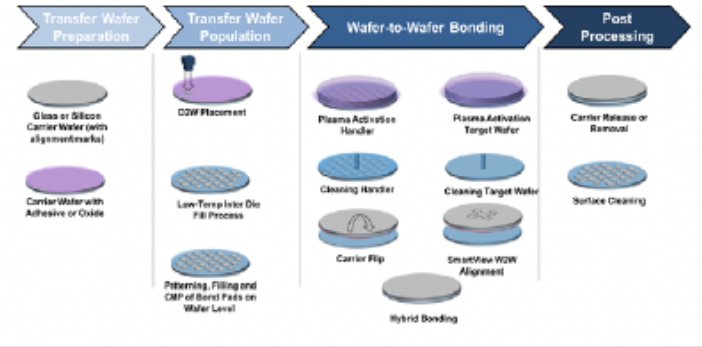

Figure 4: Reconstructed D2W hybrid bonding process flow.

cleanliness and activation are similar to those of other fusion bonding techniques. The dies require repopulation on a dedicated cleaning carrier wafer and may need optional cleaning during dicing before transport to the front-end clean hybrid bonding step. Surface coating on a readily CMP-treated wafer is crucial to preserve surface properties and cleanliness during dicing.

The EVG320 D2W is a flexible die preparation and activation system designed to seamlessly integrate with ASMPT's pick-and-place die bonding systems. It is equipped with a universal hardware/software interface and can be used as a stand-alone system. The system incorporates cleaning and plasma activation technology, and features an optional integrated metrology module that provides direct feedback to the die bonder on critical process parameters, such as die placement accuracy and die-height information, for improved process control.

After die preparation, the high-precision D2W die pre-bond step is carried out using ASMPT's LithoBolt—an automatic ultra-high-precision die bonding system. The system is designed to achieve high accuracy, throughput, and yield for volume production. Material handling for cleanliness control and alignment mechanism to achieve target alignment accuracy are crucial in this step. The cleaned and activated die and target wafer materials from the die preparation machine are transported to the load port of the equipment front-end module (EFEM) of the die bonding system through a cleaned front-opening unified pod (FOUP). The die is then bonded onto the target wafer under a compression force of 0.05 to 0.3MPa with a special curved collet. The force-controlled bond-arm of the bonder adjusts the compression force for bonding. Cleanliness must be well controlled throughout the entire material handling and bonding process inside the EFEM and bond chamber.

Die placement alignment accuracy is another crucial requirement in the die bonding step. Currently, the industry is calling for alignment accuracy below 200nm at 3σ for bond pad sizes below 10µm, and future bonding solutions should have accuracy down to 100nm or below. LithoBolt utilizes a new concept of alignment approach supported by a powerful optical system to assist in sub-micron-level pattern-recognition alignment. The bond head module incorporates intelligent design to enable true active alignment with real-time compensation. The alignment accuracy verification was performed using chip-on-glass (COG) for face-down mode bonding, and overall results achieved under 3σ were 106nm and 103nm (X-direction) and 131nm and 147nm (Y-direction) for the two corners accuracy as shown in Figure 5. Cleanliness assessment data showed that LithoBolt achieved ISO3 cleanroom standard both in idle and operation modes.

Accelerating process development

Manufacturers must undertake extensive development projects to determine the optimal bonding method for their devices. These projects must

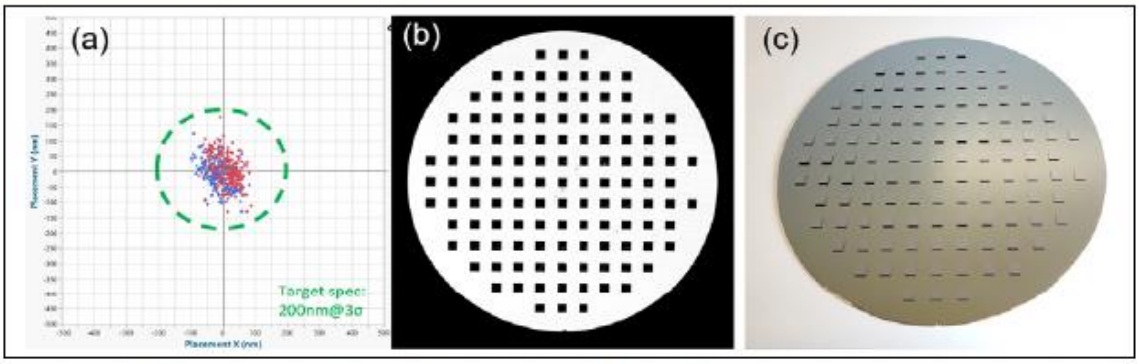

Figure 5: Recent DP-D2W bonding results using EVG320 D2W activation and cleaning followed by ASMPT LithoBolt die bonding: a) Alignment verification result using COG (9x9mm glass chip), face-down mode; achieving specification below 200nm at 3σ; b) post-bond scanning acoustic microscope; and c) photography of the bonded dies on the wafer.

consider not only the wafer bonding equipment, but also the materials involved in temporary and permanent bonding, as well as related processes such as die activation and cleaning, as well as subsequent die bonding. Process expertise and access to cutting-edge technologies are essential, but these systems are often already in use at customer sites and may be difficult to access for research and development purposes. To overcome these challenges, EVG established the Heterogeneous Integration Competence Center (HICC), which leverages EVG process solutions and expertise to enable new and improved products and applications driven by advancements in system integration and packaging, using ASMPT's latest-generation die bonding equipment. ASMPT has also established an advanced lab in Hong Kong, focused on overlay and including EVG's die cleaning and activation capabilities. These incubators were established to lower the barriers to development for customers.

Summary

The utilization of D2W hybrid bonding is crucial for the swift adoption of 3D/heterogeneous integration and the development of next-generation devices that offer superior performance, high bandwidth, and low power consumption. Even though the infrastructure for D2W hybrid bonding is still evolving, an increasing number of process solutions and collaborations across the supply chain are emerging and will be integral in establishing the best practices for D2W hybrid bonding. Close cooperation and seamless optimization between equipment design and process integration in appropriate testing labs are necessary for qualifying and refining D2W hybrid bonding.

Acknowledgements

The authors express their gratitude to Jürgen Burggraf and Mariana Pires from EV Group, as well as Hoi Ping Ng, Ming Li, and Siu Cheung So from ASM Pacific Technology, for their valuable contributions to this paper.

References

1. IEEE Heterogeneous Integration Roadmap, Ch. 2.
2. J. Burggraf, M. Pires, T. Uhrmann, "Collective die bonding - an enabling toolkit for heterogeneous integration," PRiME 2020 (ECS, ECSJ & KECS Joint Meeting), 2020.
3. T. Uhrmann, et al., "D2W hybrid bonding using high-accuracy carrier solutions for 3D system integration," 2023 IEEE ECTC.
4. C. H. Fan, et al., "Direct die-to-wafer Cu hybrid bonding for volume production," 2023 IEEE ECTC.
5. A. Elsherbini, et al., "Enabling next-generation 3D heterogeneous integration architectures on Intel process," 2022 International Electron Devices Meeting, pp. 27.3.1-27.3.4.
6. A. Bond, et al., "Collective die-to-wafer self-assembly for high alignment accuracy and high-throughput 3D integration," 2022 IEEE ECTC, pp. 168-176.

Biographies

Thomas Uhrmann is Director of Business Development at EV Group, Austria. He is responsible for overseeing all aspects of EVG's worldwide business development. Previously, he was Business Development Manager for 3D and advanced packaging, as well as compound semiconductors and Si-based power devices at EV Group. He holds an Engineering degree in Mechatronics from the U. of Applied Sciences in Regensburg, and a PhD in Semiconductor Physics from the Vienna U. of Technology (TU Wien). Email: T.Uhrmann@EVGroup.com

Nelson Fan is Vice President of Business Development, Advanced Packaging Technology, at ASM Pacific Technology (ASMPT), Hong Kong SAR, China, where he focuses on high-precision bonding and pick-and-place solutions for advanced packaging. Prior to joining ASMPT, he held multiple senior management roles in R&D and manufacturing in both OSAT and design houses. He has more than 40 US patents in semiconductor packaging technologies and holds a Bachelor's degree in Electrical Engineering from the U. of Colorado at Colorado Springs, USA.