An Innovative Approach for Advanced Packaging in 3D System Integration using Reconstructed Wafer-to-Wafer Bonding

Sanghoon Lee[1], Kyounglim Suk[1], Woojin Jang[1], Gwangjae Jeon[1], Jaegun Shin[1], Donghwi Kim[1], Kwangjin Moon[2], Jaewha Park[2], Seokhyun Lee[1] and Sujin Ahn[3]
Advanced Package Laboratory[1], Advanced Process Development Team[2], Advanced Technology Development[3]
Semiconductor R&D Center, Samsung Electronics Co., Ltd,
1, Samsungjeonja-ro, Hwaseong-si, Gyeonggi-do, 18448, South Korea
e-mail : svl.lee@samsung.com

Abstract— **Advanced package architecture and integration are among the most important technologies in high-performance HPC and AI systems of 3D system interconnection. This paper introduces an innovative approach for heterogeneous integration in a new 3D system called reconstructed wafer-to-wafer (W2W) for high-performance computing (HPC) and artificial intelligence (AI) systems. It discusses the package architecture, process flow, and 2um pitch bonding results of the reconstructed W2W technology. This approach utilizes advanced hybrid bonding technologies and offers several advantages, including high performance, efficient power delivery, high yield, and 3D system heterogeneous integration. It supports various architectures and can be configured flexibly according to system requirements, thereby improving the overall performance and efficiency of the system. Furthermore, it proposes a concept to recover alignment errors to enable the interconnection pitches of the integrated circuit package to be scaled down to sub-micron levels, significantly increasing the processing speed of the system.**

Keywords – Reconstructed wafer, Heterogeneous integration, 3DIC, Hybrid bonding, Wafer to wafer, Die to wafer, Copper to Copper, Multi-stack

I. INTRODUCTION

The increasing demands for high-performance computing (HPC) and artificial intelligence (AI) systems have emphasized the importance of advanced packaging technologies. Escalating requirements for ultra-high performance in HPC and AI products, including high performance AI chips and increased High Bandwidth Memory (HBM) capacity, has accelerated the development and advancement of packaging technologies like 2.5D or 3D system integration. Also, the increasing cost of advanced nodes and the limitation in reticle size have driven the widespread adoption of chiplet-based integration. In response to this, 3D system heterogeneous integration with hybrid bonding technologies has been proposed [1-2]. Despite the promise of 3D system technologies, current approaches to 3D system integration face many challenges especially in advanced package area. Package integration issues such as thermal dissipation, multi-stacking, thin chip handling, warpage deterioration, and fine bonding pitch are still critical and require further research and development to achieve optimal solutions.

To address these challenges, we introduce a novel scheme called "reconstructed wafer-to-wafer (W2W)", which integrates W2W FAB technology [3] with die-to-wafer (D2W) package technology [4-6]. Throughout the paper, advantages of this innovative scheme, process flows, and integrated package architecture will be discussed. First of all, it allows to improve the yield to the maximum by selectively processing only the known good dies (KGDs). While utilizing the conventional wafer bonding technology alone for chip-based integration can enhance yield, it is not possible to avoid the cumulative yield reduction of bad chips during wafer stacking process. However, integrating the D2W approach, selecting KGDs only, into the reconstructed W2W method can address this issue effectively.

The implementation of the reconstructed W2W scheme effectively eliminates the limitations of the D2W process on thin chips, addressing issues such as handling difficulties, chip cracks due to weakened chip strength, and critical warpage deterioration. Through the development and implementation of the reconstructed W2W scheme, we anticipate overcoming technological challenges not only in 3D system integration but also in expanding to high-density multi-stack HBM memory systems.

Also, as the demand for smaller semiconductor devices continues to rise, there is a need for higher interconnection circuit densities and narrower bonding pitches. To meet this challenge, reconstructed W2W technology has been developed. It involves the creation of a redistribution layer on the backside of the chip, which helps to correct misalignments that occur during the D2W bonding process. This innovative approach allows the reconstructed W2W scheme to achieve an accuracy specification margin that is more than twice as large as what is required in the D2W bonding process alone. Essentially, reconstructed W2W technology offers enhanced precision and alignment capabilities, making it suitable for meeting the demands of modern miniaturized semiconductor devices.

II. STRUCTURE AND PROCESS

A. An example of reconstructed 3D system package structure

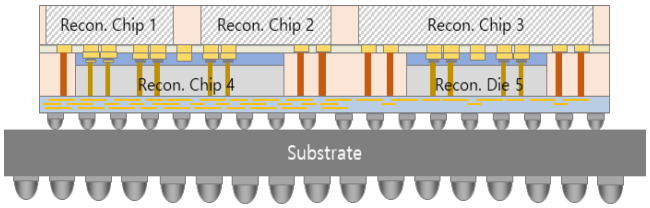

Figure 1. An example of reconstructed 3D system package structure consists of reconstructed chips and a top die.

Figure 1 illustrates an example of a W2W structure, which consists of reconstructed chips. Unlike conventional package technologies that utilize solder bump bonding, the top reconstructed die and bottom reconstructed dies are vertically connected through hybrid bonding interfaces, enabling sub-micron level bonding pitch for maximizing IC density. This hybrid bonding method involves Cu to Cu bonding and dielectric to dielectric oxide bonding. Power delivery is achieved by directly connecting the top reconstructed die to the substrate or package using through-dielectric vias (TDVs). Furthermore, the reconstructed dies are gap-filled with a high-thickness oxide dielectric layer equal to the chip thickness. The signal I/Os are directly connected through the through-silicon vias (TSVs) of the reconstructed dies, while the top reconstructed die and substrate are connected through through-dielectric vias (TDVs).

B. Reconstructed W2W Bonding Process

Simplified process flows of the typical D2W bonding and an innovative reconstructed W2W bonding are shown in figure 2. For the D2W bonding surface, dielectric oxide layers are deposited on the top of the chip wafer and the carrier wafer by chemical vapor deposition (CVD), and bonding align keys for D2W bonding are formed on the bonding surface. Then, a very flat surface with target Cu pad dishing height are obtained by a chemical mechanical polishing (CMP) process to prevent any voids formation. Then, the chip wafer is processed into chips through and sawing processes and bonded to the Si wafer. Since particle is one of the factors that directly cause voids in hybrid bonding, a surface protective coating is applied to minimize particle generation during the chip sawing, or a particle-free sawing is applied. Finally, both the singulated chips and the wafer undergo plasma activation and rinse processes to activate the bonding surface, and after D2W bonding, high-temperature annealing is performed to cause covalent bonding between oxide layers and Cu to Cu diffusion.

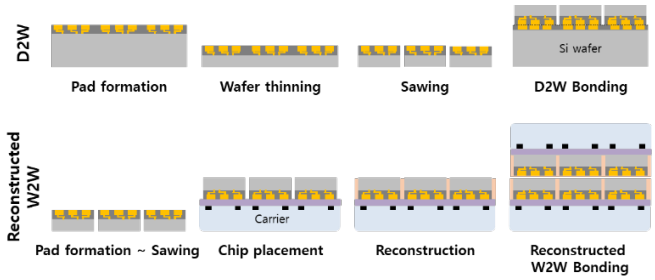

Figure 2. The typical D2W bonding and an innovative reconstructed W2W bonding process.

The reconstructed W2W bonding process is almost similar to the typical D2W process up to the D2W bonding, but it differs in that it bonds the chip to the carrier and undergoes complex subsequent processes such as reconstruction and W2W bonding. First of all, to proceed with the reconstructed W2W process, it is important to make a D2W bonded wafer into a wafer that can be bonded in W2W process. To do this, the inter-die gap between reconstructed dies is filled with dielectric material. Then, a planarization process is performed to match the total thickness variation (TTV) of each chip. The first advantage of the reconstructed W2W scheme is the standardization of chip thickness. The bonding quality varies depending on the chip thickness and tool shapes. For example, if the chip thickness is too thick, the bonding void quality is poor, and if the chip thickness is too thin, chip crack, chip handling, and warpage issues are aggravated [4].

However, reconstructed W2W has the advantage of not having any constraints on chip thickness selection because it can select a chip thickness that shows the best D2W quality and target a desired thickness. After the planarization process, the backside back end of line (BEOL) process is followed to create a hybrid bonding interface for W2W bonding interface. This backside BEOL step can be said to be the most important part of the reconstructed W2W process flow, as it can recover the misalignment that accumulates in the D2W process by rerouting the backside BEOL, and ultimately provide a sub-micron package bonding solution that can break through the precision limit of the D2W bonding equipment through the package scheme.

And W2W bonding can connect the top die and the reconstructed die through vertically hybrid bonding interfaces. Finally, the reconstructed carrier wafer, which is permanently bonded, is removed through the carrier removal process. At this time, another BEOL layer can be generated for heterogeneous D2W bonding, perform another W2W bonding to implement a multi-stack, or form a redistribution layer (RDL) to make a package product.

C. Misalignment recovery for reconstructed W2W

One of the most important factors that determines the performance of a package is the bonding pitch. The number of I/Os included in the same area depends on the bonding pitch, which directly determines the package's ability to transmit a large amount of data quickly. In conventional chip bonding technologies such as chip-on-wafer (CoW) and D2W, the bonding pitch was determined by the accuracy specification of the bonding equipment.

But, when using conventional chip bonding equipment, it is important to consider the production UPH (units per hour) as a large number of chips need to be bonded. Accuracy and production yield in bonding equipment have a trade-off relationship. While there are limitations to improving alignment accuracy through chip bonding equipment upgrades, achieving stable alignment is difficult due to the variability among hundreds of chips within a wafer. This issue makes it difficult to demonstrate 2um pitch or less sub-micron pitch with conventional package technology [7,8]. Figure 3 shows that a minimum misalignment of less than 400nm is required to achieve a 2um pitch bonding, based on a Cu to Cu pad contact area that overlaps by 50%. Although high-precision D2W bonders with a precision of 200nm or less have recently been developed, the recognition rate of real chips is low due to the complex stack-up structure of metal and dielectric, and there is a variation in recognition rate between chips. Moreover, it is difficult to implement pitches below 2um because mass quantities need to be processed.

To address this issue, the reconstructed W2W bonding technique has been introduced. This technique uses a backside BEOL process to reset the misalignment. In other words, through the backside BEOL process, the electrical connection paths of the chip can be redistributed, and the reset range can be defined as the relationship between the critical dimension of the metal pad, the photo resolution, and the photo overlay, etc. As a result, only the fine alignment achievable in W2W bonding is reflected in the final package. Although a high-precision facility with a precision of less than 400nm is required for the implementation of a 2um bonding pitch, there is a chance to make up for misalignments of 600nm or more in reconstructed bonding. This is one of the key advantages of reconstructed W2W bonding, as it further enables the implementation of sub-micron fine pitch bonding with the current D2W equipment.

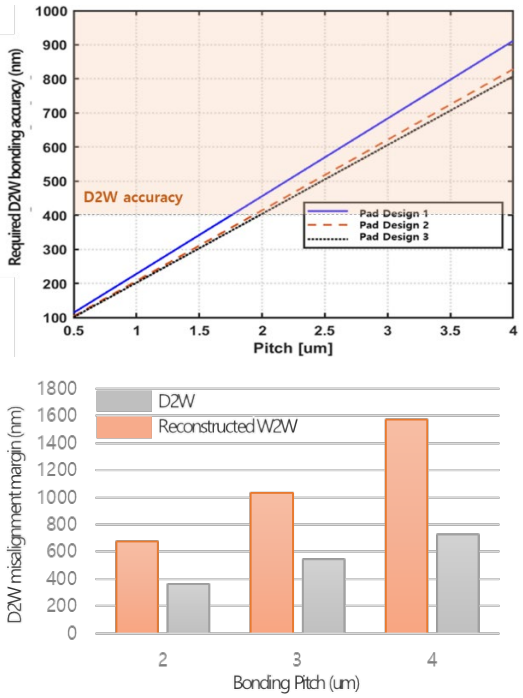

Figure 3. Required D2W accuracy and D2W misalignment margin for D2W bonding and reconstructed W2W bonding at different pitches.

III. RESULTS

A. Preparation of the reconstructed W2W bonding surface for a void-free bonding interface.

Excessive roughness or topology beyond a certain level in hybrid bonding can be the cause of void formation. The causes of roughness or topology degradation can vary including different materials, different critical dimensions of Cu bonding pads, and complex stack-up structures during the CMP process. For the reconstructed W2W structure, the top surface of the D2W chip and the inter-die gapfill area must be bonded simultaneously. Filling a thick inter-die gapfill region with dielectric material alone is already a challenging task. However, in the case of reconstructed W2W, where Si chips, Cu pads, and dielectric materials need to be planarized together, the ability to process the oxide bonding surfaces on-chip and inter-die gapfill to a flat state has become an extremely difficult technological challenges.

Figure 4 shows the results of oxide erosion measured by atomic force microscope (AFM) in both the on-chip region and inter-die gapfill region on the backside BEOL of reconstructed D2W. In the on-chip region, the Cu pads are uniformly distributed, similar to a typical W2W bonding interface. This allows for excellent oxide erosion and the desired target Cu dishing to be achieved. However, in the inter-die gapfill region, there are variations in the underlying structure of the bonding interface, differences in Si height, and

an absence of metal. These factors contribute to a significant disparity in metal density. As a result, improving oxide erosion in the inter-die gapfill region is challenging. Nevertheless, through process optimization, we were able to secure improved conditions, as depicted in the figure.

In addition, for achieving good W2W bonding quality, it is necessary to maintain the wafer edge profile at the same level as the wafer center and middle. However, in the case of reconstructed W2W, there is a tendency for the gapfill profile at the wafer edge to collapse steeply, making it extremely challenging to maintain a uniform wafer profile overall. Figure 5 illustrates the results showing the variation in wafer thickness under different wafer preparation conditions. In the pre-optimized condition 1, the thickness difference between the wafer edge and center was significantly high, at a level of 88%, and the deviation level was successfully reduced to 11%, after optimization. Through improvements in wafer thickness, oxide erosion, and dishing variations, stable W2W bonding without critical voids could be achieved after the reconstructed W2W bonding process.

Finally, figure 6 shows the level of 2um pitch bonding alignment based on W2W bonding conditions. Bonding alignment is an important factor in determining the bonding pitch, bonding pad design, and accuracy specification as it determines the un-bonding area and affects bonding yield and Cu to Cu bonding quality. Figures 6 (a) and (b) illustrate the transition patterns of W2W misalignment under conditions 1 and 2, respectively. Meanwhile, figure 6 (c) illustrates the enhanced misalignment under condition 3, with misalignment below 200nm, which could potentially enable us to achieve bonding pitches below 1um. Various factors such as wafer warpage, align key design, and equipment specification determine bonding alignment, and the tendencies of misalignment also vary depending on each factor, such as transition, rotation, run-out, etc. Through the optimization of W2W bonding, we were able to improve the level of misalignment and successfully implement 2um pitch bonding which was difficult to achieve with the D2W bonding technology.

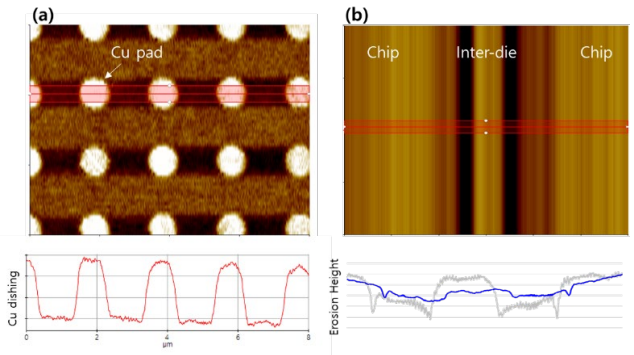

Figure 4. Cu dishing, oxide erosion, and inter-die gapfill erosion of reconstructed wafers analyzed by AFM.

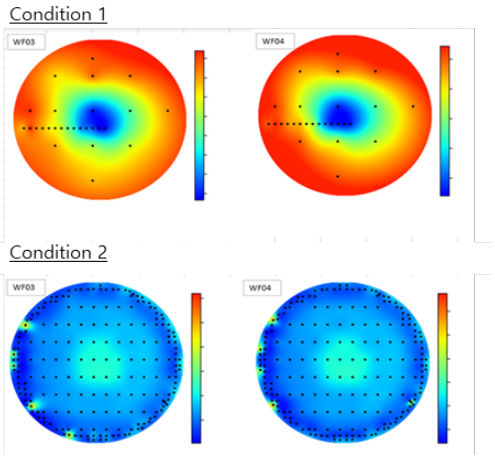

Figure 5. Variation in wafer thickness under different wafer preparation conditions at differences wafer locations - center, middle, and edge.

B. Demonstration of reconstructed W2W

The reconstructed wafer and the top wafer were W2W bonded, and a cross-sectional analysis was performed. The improvement in erosion of the inter-die gapfill region can be clearly observed by vertical cross-sectional SEM image in figure 7. In the on-chip area (a), a uniform arrangement of Cu pads can be observed, while in the inter-die gapfill region (b), it is evident that there are no Cu pads present. Instead, the area is filled solely with gapfill dielectric material. Controlling erosion becomes challenging when there are variations in Cu metal density or differences in the underlying bonding interface structure. However, through CMP optimization, it was possible to achieve a flat bonding surface resulting in successful oxide-oxide bonding without any instances of delamination or un-bonding areas.

Figure 8 displays magnified cross-sectional images of the Cu to Cu joint in the W2W bonding interface after reconstructed W2W bonding. Through high-temperature annealing, Cu expansion and metal diffusion occurred, confirming the successful Cu bonding without any observed micro voids or oxide delamination. Moreover, the development of the reconstructed W2W scheme allowed for the implementation of a 2µm pitch, which is still challenging to achieve with existing D2W equipment.

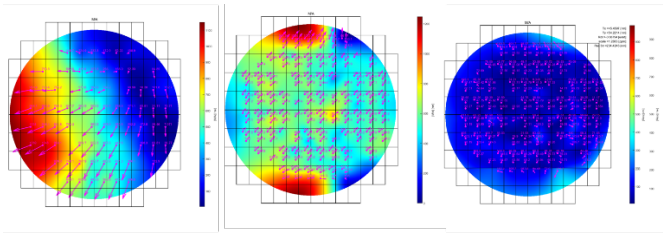

Figure 6. Bonding alignment of reconstructed W2W bonded wafers.

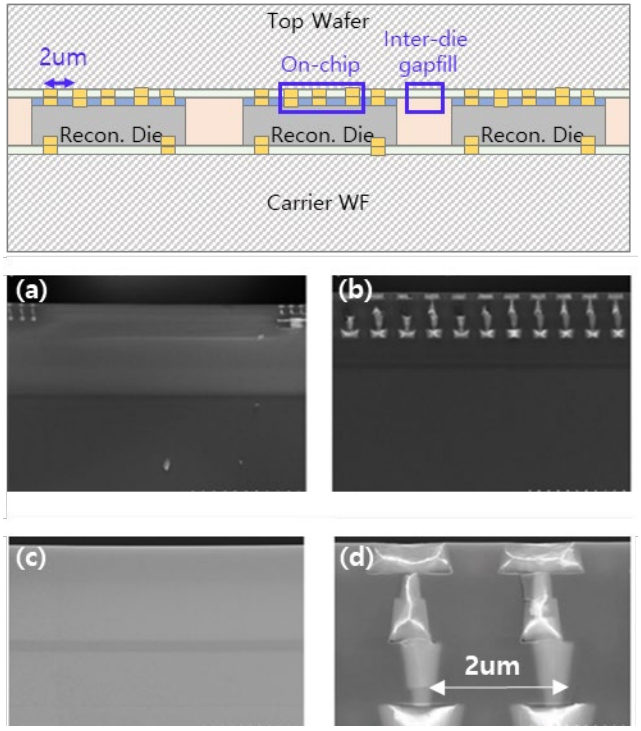

Figure 7. Cross-sectional SEM image of W2W bonding interface. Low-magnification SEM images of (a) the on-chip and (b) the inter-die gapfill. High-magnification SEM image of the (c) on-chip and (d) inter-die gapfill region.

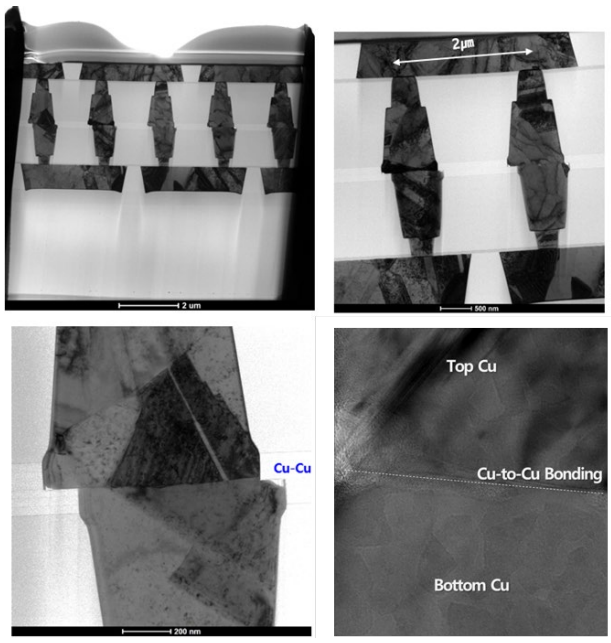

Figure 8. Cross-sectional images of the reconstructed W2W 2um pitch Cu to Cu bonding interface analyzed by TEM.

This achievement was made possible without the need for advanced bonding equipment, thereby enabling the extension of equipment usage by one generation. Additionally, daisy chains TEGs composed of 100 and 100,000 chains were tested and it was observed that the daisy chain was successfully connected through Cu to Cu bonding without any open pads.

IV. CONCLUSION

In conclusion, the reconstructed W2W technology emphasizes the importance of 3D interconnects, architectures, and advanced packaging techniques driven by the increasing demand for high-performance computing, high-performance HPC, and artificial intelligence systems. It enables the integration of 3D systems, which is a key technology to meet these demands. The technology supports various architectures and allows for flexible configurations based on system requirements, enhancing the overall system performance and efficiency. The concept of alignment error recovery at the package level and pitch reduction to sub-micron levels has been introduced. Through successful demonstration of reconstructed W2W through improvement activities such as gapfill CVD and CMP process development, wafer thickness profile improvement, on-chip and inter-die gapfill erosion improvement, and alignment enhancement, it has great potential to be applied to advanced 2.5D/3D packages that offer excellent thermal and electrical performance. By implementing a reconstructed wafer only on one side of the wafer or on both sides, it can be applied to various platforms and applications. Furthermore, we expect to further develop wafer stacking technology to enable wafer multi-stacking technology or even HBM applications. For further study, we plan to investigate factors such as Cu pad, via, and photo overlay for resetting misalignment in reconstructed W2W technology. We will also discuss the electrical yield level based on each of these factors.

REFERENCES

[1] A. Elsherbini et al., "Enabling next generation 3D heterogeneous integration architectures on intel process", 2022 IEEE International El ectron Devices Meeting

[2] Gordon Kuo et al., "A thermally friendly bonding scheme for 3D system integration", 2023 IEEE 73rd Electronic Components and Tec hnology Conference

[3] E Beyne et al., "Scalable, sub 2um pitch, Cu/SiCN to Cu/SiCN h ybrid wafer-to-wafer bonding technology", 2017 IEEE International El ectron Devices Meeting

[4] S. Lee et al., "A study on memory stack process by hybrid copp er bonding (HCB) technology", 2022 IEEE 72nd Electronic Compone nts and Technology Conference

[5] K. Kim et al., "C2W hybrid bonding interconnect technology for higher density and better thermal dissipation of high bandwidth mem

ory", 2023 IEEE 73rd Electronic Components and Technology Conference.

[6] J. Park et al., "Wafer to Wafer Hybrid Bonding for DRAM Applications", 2022 IEEE 72nd Electronic Components and Technology Conference.

[7] J. Mudrick et al., "Sub-10um Pitch Hybrid Direct Bond Interconnect Development for Die-to-Die Hybridization", 2019 IEEE 69th Electronic Components and Technology Conference.

[8] B. Bradstatter et al., "High-speed ultra-accurate direct C2W bonding ", 2020 IEEE 70th Electronic Components and Technology Conference.