Critical Challenges with Copper Hybrid Bonding for Chip-to-Wafer Memory Stacking

Wei Zhou
APTD
Micron Technology, Inc.
Boise, Idaho USA
zhouwei@micron.com

Michael Kwon
APTD
Micron Technology, Inc.
Taichung, Taiwan
mkwon@micron.com

Yingta Chiu
APTD
Micron Technology, Inc.
Taichung, Taiwan
yingtachiu@micron.com

Huimin Guo
APTD
Micron Technology, Inc.
Boise, Idaho USA
hguo@micron.com

Bharat Bhushan
APTD
Micron Technology, Inc.
Taichung, Taiwan
bharat@micron.com

Bret Street
APTD
Micron Technology, Inc.
Boise, Idaho USA
bstreet@micron.com

Kunal Parekh
APTD
Micron Technology, Inc.
Boise, Idaho USA
kparekh@micron.com

Akshay Singh
APTD
Micron Technology, Inc.
Boise, Idaho USA
akshaysingh@micron.com

***Abstract*—Due to nonmature wafer yield and customer demand for high-number die stacking, the chip-to-wafer stacking process with only known good die is a preferred solution to advanced memory products like high bandwidth memory (HBM). However, great challenges will arise if one wants to integrate it with the copper hybrid bonding technology. The memory wafer will be diced into individual chips where large amount particles will be generated and harm the hybrid bonding. In addition, the stacking process will take hours to complete rather than seconds as in a wafer-to-wafer bonding. Hence, the plasma lasting effect will be key to success. Finally, the bottom interface (IF) wafer is usually supported by a temporary carrier to sustain the wafer handling. The current wafer support system (WSS) for the IF wafer employs an organic glue, which substantially limits the thermal budget that the memory die stacking can go through. As a result, only a low-temperature annealing is allowed and low-temperature dielectric materials added. With those constraints, it was found that a porous bonding layer was generated along the interface. Failure analysis further pointed out that Cu creeping occurred along this porous interface, which might lead to leakage. An innovative solution was proposed in this work to replace the current organic-based WSS with a thin inorganic film, which can accommodate a much higher process temperature. The chemical mechanical planarization (CMP) process is found benefited too by displaying a much more consistent copper dishing as well as a uniform dielectric profile. With this new WSS, a satisfactory chip-to-wafer copper hybrid bonding process has been achieved.**

Keywords—die stacking, HBM, hybrid bonding, interconnection

I. Introduction

The copper hybrid bonding interconnection technology has been investigated in the industrial for decades as one of the most promising advanced packaging technologies. The research was intensified especially recently with the emerging of artificial intelligence (AI) and several other high computational applications. Advanced memory products, like high bandwidth memory (HBM) are found to play an important role to enable those advanced applications [1]. Compared with conventional solder-based interconnect technology, copper hybrid bonding offers great advantages in terms of electrical, mechanical, thermal, electrical-migration, and reliability performance [1], [2]. In particular, by replacing the solder inside the HBM chip-to-chip stacking with direct copper-to-copper connections, one can expect a much finer pitch, thinner bond line thickness (BLT), and more robust joints. Fig. 1 shows a typical HBM interconnection made by solder-capped pillars with a ~20μm pitch. By applying a direct copper-to-copper bonding, both the BLT and pitch can be easily reduced by half (Fig. 2). The joint itself is mechanically strong and free of intermetallic metals. Fig. 3 shows an image of electron backscatter diffraction (EBSD) across the interface where two copper pads are joined together. The original bonding interface is not observable anymore as the two copper pads become one with the copper crystal growth occurring across the bonding interface. Hence, the motivation is high to enable the copper hybrid bonding process for HBM memory products.

The current industrial mainstream of copper hybrid bonding mainly focuses on wafer-to-wafer bonding due to the facts of much easier wafer handling and better process defect control. Successful examples are advanced imager products [3] and 3D NAND [4], where wafer-to-wafer copper hybrid bonding has already been in a volume manufacturing. However, for relatively new memory products like HBM, where the wafer yield is not mature and customers are demanding a high number of die stacks, for example, 8 die or even 16 die inside one package, the accumulative yield loss will be so significant that the wafer-to-wafer stacking process is not economically feasible. Rather, chip-to-wafer stacking with known good die solutions is now the optimal path.

2377-5726/23/$31.00 ©2023 IEEE
DOI 10.1109/ECTC51909.2023.00063

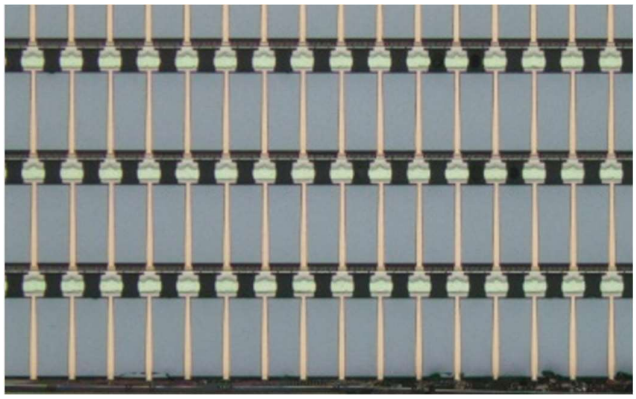


Fig. 1. Solder-based interconnection for HBM prodcuts with ~20μm pitch.

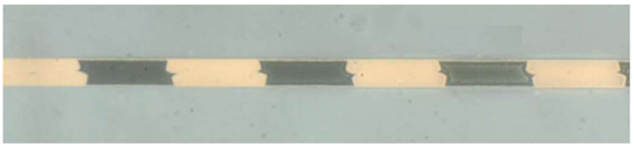


Fig. 2. Direct copper-to-copper connections for HBM.

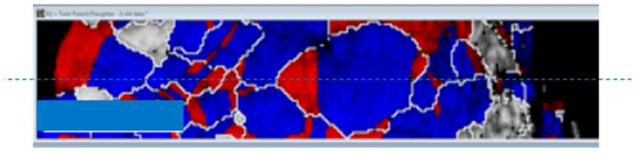


Fig. 3. Copper crystal growth across bonding inteface.

Unfortunately, a tremendous number of challenges will be encountered when conducting a chip-to-wafer stacking for HBM products by applying the copper hybrid bonding technologies. Fig. 4 illustrates a generic chip-to-wafer copper hybrid bonding process flow. The top wafer is singulated, clean, and plasma-activated before good die are picked up and stacked on top of a bottom wafer. The process will be repeated multiple times for multiple-die stacking. The bottom interface (IF) wafer is in wafer form. It is pre-attached with a temporary carrier and is similarly cleaned with plasma activation prior to bonding. After all the memory die are bonded on the bottom IF wafer, a batch annealing process takes place to formulate the permanent chip-to-wafer stacking process.

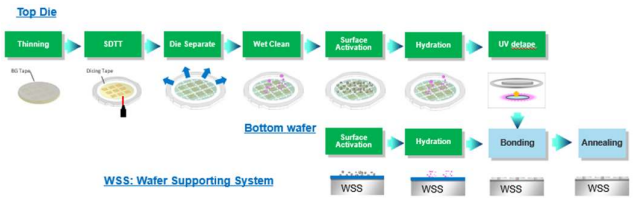


Fig. 4. HBM chip-to-wafer copper hybrid bonding process flow.

Several critical challenges unique to this process flow can be highlighted here. Hybrid bonding demands a nearly particle-free environment during stacking. However, the conventional wafer dicing process will no doubt generate many of them. The mitigation of dicing-induced particles is obviously significant to the success of this technology. The chip-to-wafer stacking process is run in a sequential mode, which means it will take hours to complete just one memory wafer stacking. Thus, the die and wafer surfaces must be kept clean and plasma-activated throughout the whole stacking process. To maintain such a large process window without degrading bonding quality, research must be undertaken to understand the staging time impacts versus the plasma activation, gases, and, most importantly, the choices of dielectric films used in the bonding. Lastly, the conventional chip-to-wafer stacking is accomplished on top of a wafer support system (WSS) with a layer of organic adhesive acted as a temporary bonding media. However, this organic adhesive will limit the choices of thermal budget of the bonding dielectric film and annealing temperature. As a result, those limitations will lead to a much weaker bonding interface and subsequently degrade the bonding integrity. This paper presents in detail the critical challenges related to HBM chip-to-wafer copper hybrid bonding and the innovative solutions available to overcome those challenges.

II. Impact of Wafer Dicing-Induced Particles

A. Particles generated by conventional wafer dicing

Conventional wafer dicing technologies include blade dicing, laser (stealth) dicing, and so on. The dicing methods are usually fine with solder-based die stacking, as the underfill is normally applied to fill the die gaps post die stacking. Alternatively, with the nonconductive film (NCF), the chip-to-chip gap filling and solder joint formation can occur at the same time. Both processes have a higher tolerance of particles as the underfill fulfills the bonding role to stick two chips together. Copper hybrid bonding, on the other hand, relies on surface hydrophilic force solely to adhere the chip dielectric surfaces together. A tiny particle, if present on the bonding surface, will be detrimental to the surface hydrophilic performance. Research has shown that particles lead to bonding voids with a size of 20x at the interface [4], [6]. Particle impact is affected by size, count, and source. Compared with blade dicing, laser stealth dicing is believed to provide a better dicing quality with less edge chipping or meandering issues. However, our studies have found that stealth dicing has generated far more particles than those allowed by a copper hybrid bonding process.

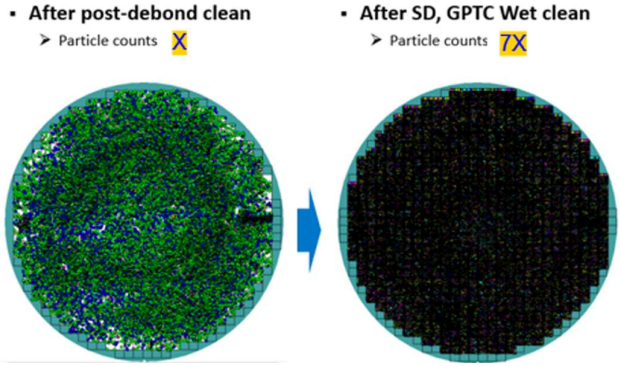
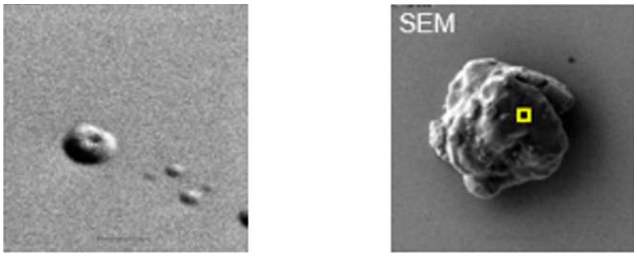


Fig. 5. Wafer particle counts increased by 7x with laser stealth dicing.

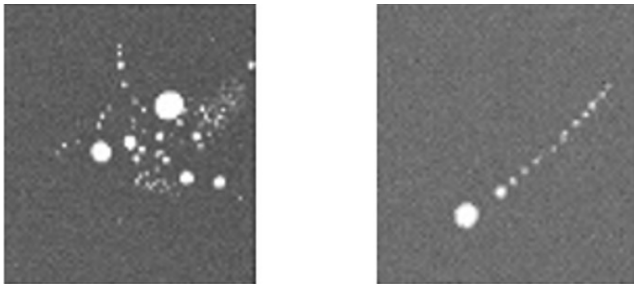
Fig. 5 shows particle counts before and after a laser stealth dicing per a sub-micron particle size criterion. The particle

counts was increased by 7x, and there were hardly any good die left for bonding. Classification has found that particle sources could be organics, like tape residues or inorganics, like silicon dust or recast—both having different impacts on the formulation of voids at bonding interface (Fig. 6). Fig. 7 shows confocal scanning acoustic microscopy (CSAM) post stacking where voids are present at the various bonding locations. Close observation of the voids easily identifies the signatures of particle impacts, as most voids are circular shaped and many are attached with a comet tail due to the bonding wave influence.



(a) (b)

Fig. 6. Classifcations of dicing paritices: a) inorganics and b) organics.



(a) (b)

Fig. 7. Sigatures of particle induced voids: a) clusters and b) comet tails.

B. Particle mitigation by combination of laser grooving and plasma dicing

Investigations have been carried out to explore the optimum dicing method to minimize dicing-induced particles. It is well believed that plasma dicing will give a very clean cut. However, multiple layers of dielectric films at the wafer scribe areas are usually present. Direct plasma cut through those dielectric films slows down the throughput significantly. Fig. 8 shows a typical memory scribe design where up to five dielectric films are common. In this work, a combination of laser grooving plus plasma etching is proposed to achieve an optimum between the dicing quality and a reasonable throughput. However, a conventional laser grooving process generates much debris or silicon dust at the kerf edge due to laser heat effects. Innovative cleaning methods must be explored to successfully remove this debris and dust before plasma etching can take place. Fig. 9 shows the effects of innovative cleaning. The burning debris are gone and the kerf chipping and burrs are significantly reduced. With the laser grooving removing those dielectric films and partial silicon, the plasma etching through the remaining bulk silicon is straightforward. Fig. 10 illustrates the results of a complete wafer cut. It shows that the saw streets are clean, and the particle counts are significantly reduced as well.

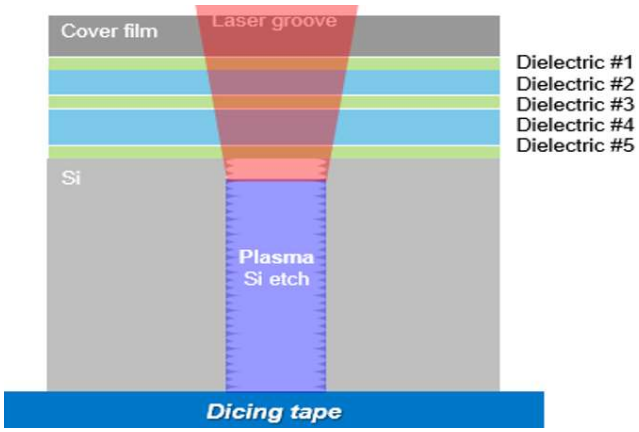
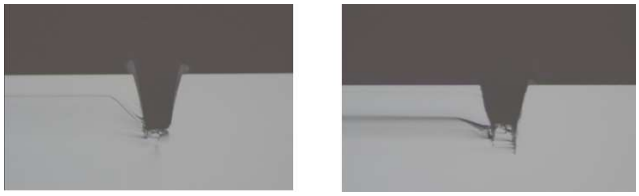
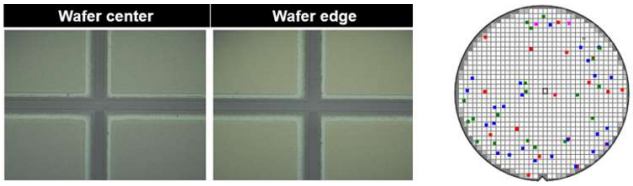


Fig. 8. Wafer dicing by laser grooving and plasma etching.



(a) (b)

Fig. 9. Innovative cleaning to remove the laser grooving induced edge burrs and Si dust: a) before and b) post cleaning.



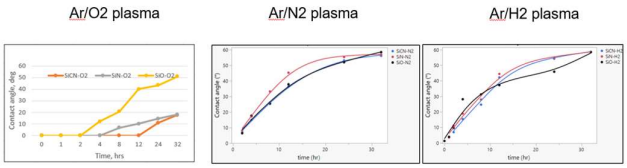
(a) (b) (c)

Fig. 10. Post wafer dicing images at saw streets at a) wafer center and b) wafer edge. c) Overal particle counts.

III. Impact of Plasma and Dielectric Materials

Plasma activations on die and wafer surfaces are critical to the success of copper hybrid bonding [5]. The outmost dielectric layer on die and wafer surfaces must be hydrophilic enough to ensure that the permanent bonding takes place at the room temperature without any external pressure assistance. In contrast to wafer-to-wafer bonding where the bonding time is only a few seconds, a complete memory die stacking on a bottom wafer will usually take hours to complete. Thus, the plasma effects over a period of wafer staging time must be kept without any degradation as long as possible. Investigations have been carried out to measure the contact angle changes with staging time for several dielectric films with different plasma gases. Fig. 11 shows the results of common dielectric films of SiCN, SiN, and SiO with the plasma of Ar/O_2, Ar/N_2, and Ar/H_2, respectively, with a staging time up to 32 hours. It was found that Ar/O_2

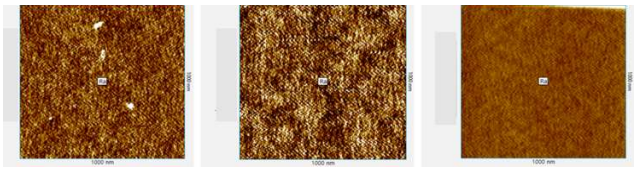
plasma has the longest endurance of surface effects by keeping almost 0 degrees of contact angles for all the dielectric materials up to at least 2h. Among them, SiCN behaves the best by keeping surface sufficiently hydrophilic up to 12h.



(a) (b) (c)

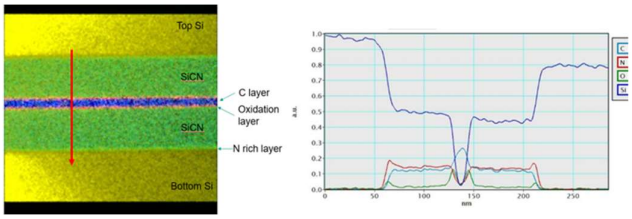
Fig. 11. Dielectric film contact angle variations with staging time after a) Ar/O_2, b) Ar/N_2, and c) Ar/H_2 plasma activation.

It is worth noting that the roughness impacts by Ar/O_2 plasma are also negligible. Measurement by atomic force microscopy (AFM) shows the Ra results of SiCN, SiN, and SiO films post plasma are 0.119nm, 0.121nm, and 0.213nm, respectively (Fig. 12). All of them are within the bonding specification requirements. Chip-to-wafer stacking was carried out to verify the bonding quality of the plasma impact versus the dielectric films. As shown in the X-section of Fig. 13, good bonding between SiCN-to-SiCN films was observed. TEM and EDS measurements were taken where there is an oxidation layer generated at each side immediately adjacent to the SiCN film, and there is an easily observable carbon-rich layer right in the center of the bonding interface. It is well believed that the presence of carbon inside the SiCN dielectric layer enhances the creations of dangling bonds upon plasma activation [2], [3], [6]. And the combined effects of Ar and O_2 plasmas further strengthens the staging endurance of dangling bonds on the wafer surface throughout the entire chip-to-wafer stacking.



(a) (b) (c)

Fig. 12. Surface roughness of a) SiCN, b) SiN, and c) SiO films after Ar/O_2 plasma.



(a) (b)

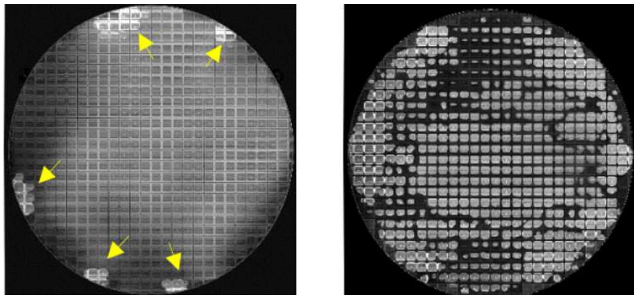
Fig. 13. a) TEM and b) EDS analysis across bonding interface of SiCN-SiCN films after Ar/O_2 plasma.

IV. Impact of Wafer Support System

A. Limitation of conventional WSS

In a conventional HBM chip-to-wafer stacking, the bottom interface (IF) wafer is usually very thin, 50ụm in thickness as an example, to minimize the final package height. However the IF wafer itself is too thin to stand alone during process handling. Instead, a carrier is generally attached in advance and removed at the end of assembly process [14], [16]. Fig. 14 demonstrates a typical process flow where a wafer support system (WSS) is employed. A thick organic glue is usually applied on a carrier before the IF wafer is temporarily bonded on the top. The temporary carrier will be removed later by either mechanical separation or laser agitation depending on the debonding mechanisms [17]–[19]. However, due to the limitation of the thermal budget of the organic glue, the whole process flow must run at a low temperature range with a thermal excursion usually below 200°C, including the critical post stacking annealing process and the bonding dielectric layer deposition. To mitigate the issues, industrial has been actively pursuing several kinds of high-temperature glues that claim abilities of a more than 300°C thermal budget. However, investigations here show that even with a high-temperature glue, large voids or even delamination are still observed with a typical annealing process of 350°C/2h. Fig. 15 demonstrates the severe CSAM voids encountered with two different organic glues both claiming process abilities of beyond 300°C. In addition, it was found that the organic glues contribute extra particles to the stacking, and cause excessive total thickness variation (TTV) during CMP process.

Fig. 14. Application of conventional WSS in the chip-to-wafer stacking process flow.

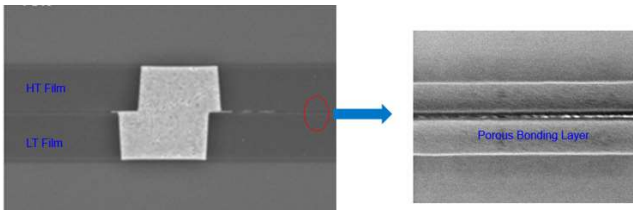


(a) (b)

Fig. 15. a) Voids and b) delamination observed in conventional WSS with two different types of organic glues.

The other way to overcome this issue that industrial may be thinking of is to go for a low-temperature annealing. Work has been accomplished here to investigate the feasibility of a 180°C annealing, which is the comfortable working temperature that most organic glues in the market can support. However, to enable such a low working temperature, the dielectric film that can be deposited at the backsides of DRAM and IF wafers must be constrained with low temperature too, once WSS is attached.

As a result, one of the dielectric layers employed at the bonding interface will be inferior to the other. This will ultimately degrade the bonding quality to a great extent. Fig. 16 shows the results of a copper hybrid bonding where the conventional WSS was present and a low-temperature-processed dielectric film was applied at the IF wafer backside. Although the Cu-to-Cu pad joint looks good, there is a porous layer formed inside the bonding interface between the memory die and bottom wafer dielectric films. Further reliability work pointed out that copper creeping has actually occurred along the porous bonding layers. Hence, working with conventional organic glues was found very challenging to achieve a robust chip-to-wafer hybrid bonding for the HBM memory products.



(a) (b)

Fig. 16. Issues of low-temperaure annnealing with a) Cu creeping and b) porous layer between the bonding interface.

B. Innovative new WSS

To generically solve the above dilemma, we propose an innovative WSS method. Fig. 17 explains the key differences of the new WSS method and the process flow. Instead of a thick organic coating, a thin inorganic film will be applied as a glue to temporarily bond the device wafer and carrier wafer together. The beauty of using an inorganic film is that it has a much higher thermal budget, enough to sustain post bonding anneal at a higher temperature, say 350°C. In addition, there is no limitation on the types of bonding dielectric films that can be used at the backside of DRAM die or IF wafer surface. It is theoretically feasible that the same type of a high-temperature film can be applied at both bonding layers across the interface. The removal of the carrier will require a new method, too.

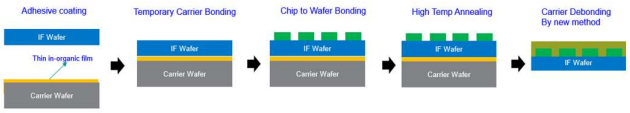


Fig. 17. Application of a new WSS in the chip-to-wafer stacking process flow.

The other added advantages of the new WSS are that it can significantly improve CMP performance. The thinner inorganic glue will minimize the CMP-induced wafer TTV greatly. And the employment of a high-temperature-based dielectric film as the bonding layer will make CMP control on the Cu pad dishing more stable. Together with optimizations in processes and slurry selections, Fig. 18 elaborates the improved CMP results for a fine-pitch (<10μm) memory product. AFM measurements were taken at the different wafer locations for the copper pad step height and dielectric profiles, respectively. It can be seen that both pad dishing and dielectric top surface profiles are very flat and consistent across the whole wafer. Fig. 19 demonstrates the improved CMP erosion at the critical areas transitioning from scribe to high Cu pad density regions. We determined that the erosion amount is minimal and within the process specs.

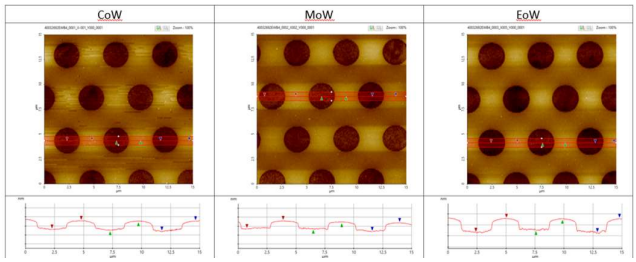


Fig. 18. Improved Cu pad dishing and dielectric surface profile at the different wafer locations.

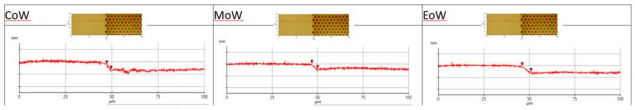


Fig. 19. Improved CMP erosion at the transition areas from scribe to a high copper density region.

With the adoption of the new WSS, DRAM die-to-IF wafer stacking was demonstrated. Fig. 20 shows the finished chip-to-wafer hybrid bonding. And Fig. 21 illustrates a X-section of the copper hybrid bonding around one signal pin. In addition, Fig. 22 further demonstrates the bonding quality of the interface between the top and bottom dielectric films and the copper grain structure. It can be seen that both the dielectric-to-dielectric bonding and the copper pad-to-copper pad connection between the DRAM die and the IF wafer are robust. More electrical measurement and reliability data collection will be reported separately.

(a) (b)

Fig. 20. Finished chip-to-wafer hybrid bonding for a fine-pitch HBM device with pics of a) full wafer and b) local image.

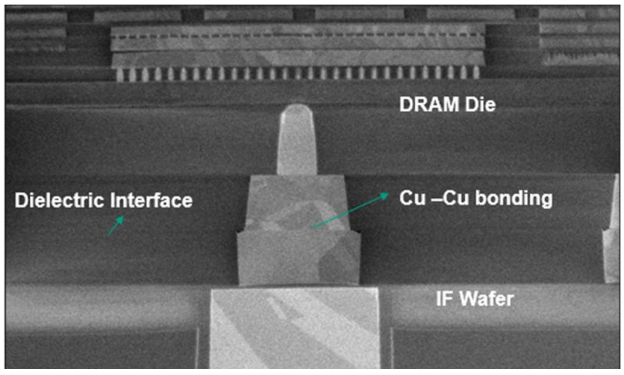
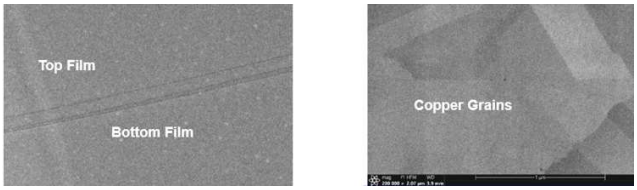


Fig. 21. X-section of copper hybrid bonding between a DRAM die and IF wafer.



(a) (b)

Fig. 22. Magnified images of the bonding interface between a) top and bottom dielectric films and b) Cu-Cu joints.

V. Conclusion

A chip-to-wafer hybrid bonding process for HBM memory products has been investigated in this work. Critical challenges were elaborated in detail, including dicing process-induced particles, the requirement of a long staging time, and the impact on plasma gases and the selection of dielectric materials. In particular, it was found the thick organic glue within the conventional wafer supporting system has a significant limitation on the thermal excursion of the annealing process as well as the choices of dielectric material depositions. It was found that those limitations lead to poor bonding quality. A porous bonding layer was generated along the interface where the copper creeping would be keen to occur. Innovative solutions including new cleaning methods and a thin inorganic film-based WSS were proposed, which have proved the efficiency in minimizing the particles and improving the CMP performance. As a result, a robust chip-to-wafer hybrid bonding for a fine-pitch HBM product has been achieved and demonstrated.

Acknowledgment

The authors greatly appreciate the support of the packaging failure analysis labs at Micron in Boise as well as Taiwan for their tedious and timely efforts in sample preparation and failure analysis. The team is also very grateful to TD management and other functional teams for their guidance and project execution throughout the entire technology development period.

References

[1] W. Kim, Y. Lee, W. Choi, K. Lim, BK Moon, and D. Rhee, "Plasma chamber environment control to enhance bonding strength for wafer-to-wafer bonding processing," *IEEE Electronic Components and Technology Conference (ECTC),* 2022, pp. 2008–2012.

[2] E. Beyne, S. W. Kim, L. Peng, N. Heylen, J. D. Messemaeker, O. O. Okudur, A. Phommahaxay, TG Kim, M. Stucchi, D. Velenis, A. Miller, and G. Beyer, "Scalable, sub 2µm pitch, Cu/SiCN to Cu/SiCN hybrid wafer-to-wafer bonding technology," *IEEE International Electron Devices Meeting (IEDM),* 2017, pp. 32.4.1–32.4.4.

[3] J. Jourdon et al., "Hybrid bonding for 3D stacked image sensors: impact of pitch shrinkage on interconnect robustness," *IEDM Dig. Tech. Papers,* pp. 7.3.1–7.3.4, Dec. 2018.

[4] L. Peng, S-W. Kim, S. Iacovo, F. Inoue, A. Phommahaxay, E. Sleeckx, J. De Vos, D. Zinner, T. Wagenleitner, T. Uhrmann, M. Wimplinger, B. Schoenaers, A. Stesmans, V. V. Afanas'ev, A. Miller, G. Beyer, and E. Beyne, "Advances in SiCN-SiCN bonding with high accuracy wafer-to-wafer (W2W) stacking technology," *2018 IEEE International Interconnect Technology Conference (IITC),* 2018, pp. 179–181.

[5] S. Son, J. Min, E. Jung, H. Kim, T. Kim, H. Jeon, J. Kim, S. Kim, K. Moon, H. Na, and K. Hwang, "Characteristics of plasma-activated dielectric film surfaces for direct wafer bonding," *IEEE 70th Electronic Components and Technology Conference (ECTC),* 2020, pp. 2025–2032.

[6] F. Inoue, L. Peng, S. Iacovo, A. Phommahaxay, P. Verdonck, J. Meersschaut, P. Dara, E. Sleeckx, A. Miller, G. Beyer, and E. Beyne, "Influence of composition of SiCN as interfacial layer on plasma activated direct bonding," *ECS J. Solid State Sci. Technol.,* vol 8, pp. 346–350, Jun. 2019.

[7] D. W. Fisher, S. Knickerbocker, D. Smith, R. Katz, J. Garant, J. Lubguban et al., "Face to face hybrid wafer bonding for fine pitch applications," *IEEE Electronic Components and Technology Conference (ECTC),* 2020, pp. 595–600.

[8] Y. Kagawa et al., "Novel stacked CMOS image sensor with advanced Cu2Cu hybrid bonding," *IEEE International Electron Devices Meeting (IEDM),* 2016, pp. 208–211.

[9] S.-W. Kim et al., "Novel Cu/SiCN surface topography control for 1 µm pitch hybrid wafer-to-wafer bonding," *Electronic Components and Technology Conference (ECTC),* 2020, pp. 216–222.

[10] G. Gao et al., "Development of hybrid bond interconnect technology for die-to-wafer and die-to-die applications," *IWLPC,* San Jose, California USA, Oct. 2017.

[11] G. Gao et al., "Scaling package interconnects below 20µm pitch with hybrid bonding," *IEEE 68th Electronic Components and Technology Conference,* 2018, p. 314.

[12] J. A. Thei et al., "Recent developments in fine pitch wafer-to-wafer hybrid bonding with copper interconnects," *IWLPC Dig. Tech. Papers,* pp. 1–6, Oct. 2019.

[13] C. H. Yang et al., "Self-annealing behavior of electroplated Cu with different brightener concentrations, *J. Electrochemical Society,* vol. 167, no. 8, May 2020.

[14] X. F. Brun, J. Burggraf, B. Ruxandra-Aida, and C. Mühlstätter, "Low temperature direct bonding of SiN and SiO interfaces for packaging applications," *2020 IEEE 70th Electronic Components and Technology Conference (ECTC),* Orlando, Florida USA, 2020, pp. 182–187.

[15] G. Abadias, E. Chason, J. Keckes, M. Sebastiani, G. B. Thompson, E. Barthel, G. L. Doll, C. E. Murray, C. H. Stoessel, and L. Martinu, "Review article: Stress in thin films and coatings: Current status, challenges, and prospects," *J. Vacuum Science and Technol. A,* vol. 36, (2) , pp. 1–48, Mar./Apr. 2018.

[16] P. Carazzetti, F. Balon, M. Hoffmann, J. Weichart, A. Erhart, E. Strolz, and K. Viehweger "Impact of process control on UBM/RDL contact resistance for next-generation fan-out devices," in *Proc. IEEE 67th Electronic Components and Technology Conf. (ECTC),* 2017, pp. 909–916.

[17] S. Guruvenket, S. Andrie, M. Simon, K. W. Johnson, and R. A. Sailer, "Atmospheric pressure plasma CVD of amorphous hydrogenated silicon carbonitride (a␣SiCN:H) films using triethylsilane and nitrogen," *Plasma Processes and Polymers,* vol. 8, no. 12, pp. 1126–1136, 2011.

[18] G. Gao, L. Mirkarimi, G. Fountain, T. Workman, J. Theil, G. Guevara, C. Uzoh, D. Suwito, and B. Lee, "Die to wafer stacking with low temperature hybrid bonding," *70th ECTC,* Florida USA, pp. 589–594. 2020.

[19] H. K. Seo, S. E. Kim, G. Kim, H. S. Park, and Y.-B. Park, "Effects of two-step plasma treatment on Cu and SiO_2 surfaces for 3D bonding applications," *70th ECTC,* Florida USA, pp. 167–1683. 2020.