

Inverse Hybrid Bonding with Metal Organic Framework as Infill for Heterogeneous Integration

A Master Thesis Topic Summary
Presented to
The Academic Faculty

by

Rohan Sahay

In Partial Fulfillment
of the Requirements for the Degree
Master of Science in the
School of Electrical and Computer Engineering

Georgia Institute of Technology
August 2024

COPYRIGHT © 2024 BY Rohan Sahay

Inverse Hybrid Bonding with Metal Organic Framework as Infill for Heterogeneous Integration

Approved by:

Dr. Muhannad S. Bakir, Advisor
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Azadeh Ansari
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Albert B Frazier
School of Electrical and Computer Engineering
Georgia Institute of Technology

Date Approved: [July 26, 2024]

ACKNOWLEDGEMENTS

I want to thank my advisor, Dr. Bakir, for giving me this opportunity and for his constant support. He has been nothing but encouraging, and moreover, he showed trust in me even during the passage of failures, which helped me grow as a student and a researcher. In addition to his role as a research advisor, Dr. Bakir has been more of a teacher and a guide to me. His insightful feedback on research progress, active involvement, and comprehensive support have fostered my professional development, and personally, I could not be more grateful to him.

I want to thank Dr. Frazier and Dr. Ansari for serving as my thesis committee members and providing me with their strong guidance. I also want to thank all my lab mates and cleanroom staff members, especially Ashita Victor, Madison Manley, Durga GaJula, Christopher White, and Ankit Kaul, for their support and guidance. I want to thank Dr. Kummel and his students at UCSD - Dipayan Pal, Victor Wang, Chenghsuan Kuo, and Jit Dutta for bringing different ideas and their support for our collaborative work on Inverse Hybrid Bonding.

Lastly, I want to thank my family and friends, especially my father, Rajiv Sahay, mother, Sonali Prasad, and brother, Rohit Sahay, who have been with me through thick and thin. Knowing they have my back has helped me overcome uphill challenges.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	iii
LIST OF TABLES	vi
LIST OF FIGURES	vii
SUMMARY	x
CHAPTER 1: INTRODUCTION	1
1.1 Slowdown of Moore's Law	1
1.2 The Rise of Heterogeneous Integration	3
1.3 Recent Advancements in Heterogeneous Integration	9
1.3.1 Enhanced 2D Heterogeneous Integration (2.5D Heterogeneous integration) .	9
(a) Silicon Bridge – Intel’s EMIB & TSMC’s CoWoS-L	10
(b) Fan-out Wafer-level Packaging – TSMC’s InFO-WLP.....	11
1.3.2 3D Heterogeneous Integration	13
1.4. Challenges with Contemporary Interconnect Technologies in Heterogeneous	
Integration	13
1.5 Scope of Research.....	17
CHAPTER 2: LITERATURE SURVEY OF CURRENT INTERCONNECT	
TECHNOLOGIES	19
2.1 Need for High-density I/Os.....	19
2.2 Direct Copper Bonding	20
2.2.1 Surface Activated Bonding (SAB)	21
2.2.2 Surface Modification	22
2.2.3 Surface Passivation	22
2.3 Challenges with Underfill for Direct Copper Bonding.....	23
2.4 Hybrid Bonding	25
2.4.1 Wafer-to-wafer Hybrid Bonding	27
2.4.2 D2D/D2W Bonding	28
(a) AMD 3D V- Cache Technology	28
(b) Intel’s Foveros Direct.....	29

(c) Hybrid bonding as an emerging technology in HBM	31
2.4.3 Challenges in D2D/D2W Hybrid Bonding	32
(a) Alignment accuracy	33
(b) Low-thermal budget	35
(c) Surface cleanliness	36
2.5 Conclusion	38
CHAPTER 3: DIRECT COPPER BONDING	39
3.1 Testbed for Direct Copper Bonding.....	39
3.1.1 Flip-chip bonding process.....	41
3.2 Ruthenium for Low-Temperature Cu-Cu Bonding	43
CHAPTER 4: STUDY OF ALD DEPOSITION PARAMETERS FOR ENHANCED INFILL COVERAGE.....	47
4.1 ALD-based Aluminum Oxide as Infill for IHB	47
4.2 Metal Organic Framework for Infill using ALD-CVD Deposition	54
4.2.1 MOF Deposition Process	54
4.2.2. Coverage vs ALD deposition parameters	56
CHAPTER 5: Inverse Hybrid Bonding	62
5.1 First Demonstration of Inverse Hybrid Bonding	63
5.2 Inverse Hybrid Bonding with Fine Pitch Dimensions – 5 μ m Pitch	68
CHAPTER 6: FUTURE WORK	73
6.1 Direct Copper Bonding.....	73
6.2 Infill for Inverse Hybrid Bonding for Fine Pitch Heterogeneous Integration	74
6.3 Modeling and Simulation for the IHB Integrated Structures.....	76
References.....	76

LIST OF TABLES

Table 1 Ideal parameters for an underfill	24
Table 2 List of bonding parameters used for flip chip assembly	42
Table 3 Properties of ZIF-8 Metal Organic Framework.....	55

LIST OF FIGURES

Figure 1 (a) Gate cost trends vs scaling; (b) Rising bandwidth and corresponding I/O density increase over time [3] [4]	1
Figure 2 (a) Monolithic vs. chiplet integration architecture; (b) yield, performance, and risk analysis for monolithic vs. chiplet approach; (c) More silicon area and hence more transistor density with HI over monolithic architecture [10] [11][12]	4
Figure 3 The difference in the scaling of the bandwidth for compute (measured in peak compute capability (i.e., FLOPS)), memory, and interconnects [15]	6
Figure 4 (a) I/O density vs pitch scaling; (b) Bandwidth/stack for different memory technologies [18].....	7
Figure 5 (a) simplified schematic depicting the route for die-to-die signal transmission with bumping technology; (b) Reduction in parasitics through use of hybrid bonding [20]	8
Figure 6 (a) Hybrid bonding (TSMC’s SoIC) vs. flip-chip using bump interconnect technology comparison for (a) Insertion loss; (b) RLC parasitics [21]	9
Figure 7 Conventional 2.5D architecture using an interposer [24].....	10
Figure 8 (a) Intel’s Embedded multi-die interconnect bridge (EMIB) using silicon bridge embedded in package itself; (b) TSMC’s Chip-on-wafer-on-substrate (CoWoS-L) using a silicon bridge embedded in an interposer [25] [26]	11
Figure 9 (a) Traditional FO-WLP with the dia embedded in mold; (b) Commercial application of FO-WLP in TSMC’s InFO-WLP technology [30] [31]	13
Figure 10 Bandwidth density and I/O pitch for different interconnect technologies [36] 15	15
Figure 11 Challenges with contemporary fine pitch bonding schemes and research objective for the current work.....	18
Figure 12 Projected I/O density and pitch scaling trends for different integration schemes such as Die-to-Die, Die-to-wafer etc. [51].....	21
Figure 13 Capillary flow-based underfill dispensing process [61]	24
Figure 14 No-Flow underfill process [64]	25
Figure 15 Process sequence for hybrid bonding	27
Figure 16 Wafer-to-wafer bonding [73]	28
Figure 17 AMD 3D V-Cache using die-to-wafer hybrid bonding [74].....	29
Figure 18 Foveros Direct technology by Intel using hybrid bonding [78]	30
Figure 19 Evolution of Intel’s interconnect technologies [81]	31
Figure 20 Low form factor advantage of using hybrid bonding over conventional TCB bonding using microbumps [83]	32
Figure 21 The IMEC 3D interconnect technology landscape [90]	34
Figure 22 D2D/D2W assembly using (a) pick and place tool; (b) Carrier wafer [91]	35
Figure 23 Self-alignment process using water droplets [93]	36
Figure 24 Xperi’s DBI Ultra D2D/D2W technology [95]	37
Figure 25 Number of particles generated by various sources during C2W hybrid bonding [76].....	38
Figure 26 IHB in comparison to state-of-the-art D2D/D2W bonding technologies.....	39
Figure 27 (a) schematic of the testbed depicting the cross-section; (b) microscope images of top die depicting copper pillars on top of pads; (c) microscope image of bottom die depicting copper pillars and metal traces; (d) schematic of the integrated testbed used ..	41
Figure 28 Fabrication steps for Cu-Cu bonding testbed	42

Figure 29 (a) Kelvin structure for per I/O resistance measurement; (b) SEM Cross-section of the bonded copper pillar	44
Figure 30 Testbed for Cu-Cu bonding with Ru capping	45
Figure 31 Deposition process of ruthenium on copper	45
Figure 32 Percentage of Cu and Ru on the surface after 30 min FGA (Forming Gas Annealing) anneal [98]	46
Figure 33 Diffused Cu on Ru surface with pinholes after 450°C FGA anneal	47
Figure 34 Fabrication flow for testbeds used for study of infill	49
Figure 35 (a) Schematic of cross-section of testbed used for infill analysis; (b) Top view of the integrated structure using Keyence 7000 Microscope depicting the through hole density on the top die	50
Figure 36 (a) SEM image of the area scanned for EDX mapping post 200 cycle aluminum oxide ALD; (b) Top view of the integrated structure highlighting the area under study for infill coverage; (c) EDX mapping results depicting the areas where aluminum and oxygen pe.....	51
Figure 37 FIB cross-section of the de-bonded bottom die after 200 cycles of aluminum oxide ALD depicting	52
Figure 38 (a) Different regions of EDX study after 200 cycles ALD with vacuum purge – areas of bottom die right beneath through holes (Region A), areas 300 μm from the nearest through hole (Region B), Areas near the edge (Region C); (b) Al/Si ratio in the three r	52
Figure 39 (a) Diagonal Dicing of the integrated structure; (b) Close view of the cross-section of the integrated structure after depicting smooth infilled area and damaged area with no infill; (c) non-magnified image of cross-section depicting infill up to 300 μm ..	54
Figure 40 (a) ZIF-8 MOF chemistry and structure; (b) Schematic represents ALD ZnO and conversion to vapor-deposited MOF with around 10-15x volume expansion and gap fill.....	56
Figure 41 EDX mapping of the cross-section of integrated structure 300 μm from nearest through hole after full infill of die-to-substrate gap.....	57
Figure 42 SEM image of cross-section depicting non-penetration of polishing mold into the vias and die-to-substrate gap rendering the results discussed above credible	58
Figure 43 (a) Schematic of the testbed with a top die having through holes Cu-Cu bonded to a bottom substrate; (b) SEM cross-section of the Cu-Cu bonded test structure (Face A); (c) copper pad/pillar on the surface of the top die before bonding (Face B); (d) copper pillar with metal traces on the bottom die’s surface before bonding (Face C)	59
Figure 44 De-bonded bottom die overview after 20 nm ZnO ALD; (b) Regions of undeposited infill beneath the through holes	60
Figure 45 Area on the top die near through hole depicting spread radius of (a) 572 μm for 240 μm diameter via; (b) 364 μm for 75 μm diameter via	61
Figure 46 De-bonded bottom die overview after 20 nm ZnO ALD with a) pulse time 250 ms & temperature 150°C; (b) pulse time 1 sec & temperature 150°C	62
Figure 47 Process sequence for Inverse Hybrid Bonding.....	63
Figure 48 (a) Overview of the de-bonded bottom die after 1 cycle of MOF; (b) MOF grains observed right in the middle of bottom die depicting complete penetration.....	64
Figure 49 FIB cross-section near the center of the bottom substrate depicting 500 nm MOF as expected	65
Figure 50 (a) Overview of the de-bonded bottom die after 5 cycles of MOF infill; (b)	

Analysis area at the center of the bottom substrate after de-bonding; (c) Spectrum and EDX mapping depicting Zn and O peaks and hence the presence of MOF in the substrate's center	66
Figure 51 (a) Bond shear using Xyztec's shear tester; (b) Comparison of Peak shear force for the bonded dice before and after MOF infill.....	67
Figure 52 (a) Schematic of the cross-section of testbed used for infill analysis with 5 μm pitch I/Os; (b) SEM image of top die depicting 5 μm pitch copper I/Os	68
Figure 53 Overview of the top die after 20 nm ZnO deposition with a temperature of 160°C and pulse time of (a) 250 ms (b) 1 sec (c) 8 sec	69
Figure 54 Atomic percentage of Zn,O and Si at different regions of the de-bonded top die with 8 sec pulse time to analyze for coverage uniformity	70
Figure 55 Atomic percentage of Zn at different regions of the de-bonded top die after 3 cycles of ALD/CVD MOF deposition	71
Figure 56 Tilted SEM images of the top die at edges (A and C) and at the center after 3 cycles of MOF infill.....	72
Figure 57 (a) The region where FIB cross section was made (b) FIB cross-section of a 100 μm x 100 μm copper pad on de-bonded bottom die after 3 cycles of MOF.....	73
Figure 58 Cyclic clean process by using HOOH as oxidant to remove carbides over copper and followed by N ₂ H ₄ as reductant to remove oxides.	74
Figure 59 Unconverted ZnO at the center for 5 μm I/O pitch sample with 500 nm die-to-substrate gap	75
Figure 60 Top silicon die with (a) 1 μm ; (b) 2 μm Cu pillars; (c) Silicon-to-glass assembled testbed for fine pitch IHB.....	76

SUMMARY

This thesis presents Inverse Hybrid Bonding (IHB) as a novel fine-pitch die-to-die (D2D)/die-to-wafer (D2W) bonding scheme for heterogeneous integration. IHB is a two-step process wherein (1) Direct copper bonding is used for I/O connections, and (2) ALD-CVD deposition of Metal Oxide Framework (MOF) is used for the infill. The post-bond infill can potentially mitigate extreme particle control requirements for fine-pitch D2D/D2W bonding frameworks. Furthermore, compared to the viscous epoxy-based conventional underfills, a conformal, void-free infill deposition can be achieved with minimal die-to-substrate gap using gaseous precursors/reactants in the ALD-CVD infill deposition process.

In the present work, direct Cu-Cu thermocompression bonding with formic acid preclean was first demonstrated at 300°C. The bond quality was characterized by I/O resistance and shear strength measurements. Furthermore, ruthenium was studied as passivation capping over copper to achieve low-temperature Cu-Cu bonding.

For the post-bond infill in IHB, two different materials – aluminum oxide and ZrF₈ MOF were analyzed. A sequential set of experiments were conducted to study the effect of ALD process parameters on area coverage for infill, based on which the ALD/CVD infill process was optimized to successfully demonstrate IHB with a 5 mm x 5 mm effective bonded area and a 1 μm die-to-substrate gap. The infilled material was characterized for coverage and the effect of infill on the mechanical strength of the package. Further, current progress and future directions on IHB for fine-pitch have been discussed.

CHAPTER 1: INTRODUCTION

1.1 Slowdown of Moore's Law

From the earliest transistors to the System-on-chips (SoCs) in modern-day smartphones, the impact of the semiconductor industry on human life is unparalleled. While these life-changing innovations have been a major driving force, the production cost that goes behind them plays an equally important role in determining the dynamics of the semiconductor industry. Over the years, continuous scaling enabled more computing components or transistors on a single monolithic chip. Through this, the industry was able to meet the minimum component cost with enhanced performance capabilities [1]. The same was suggested by Moore's law, which stated that the number of transistors on a microchip would roughly double every two years [2].

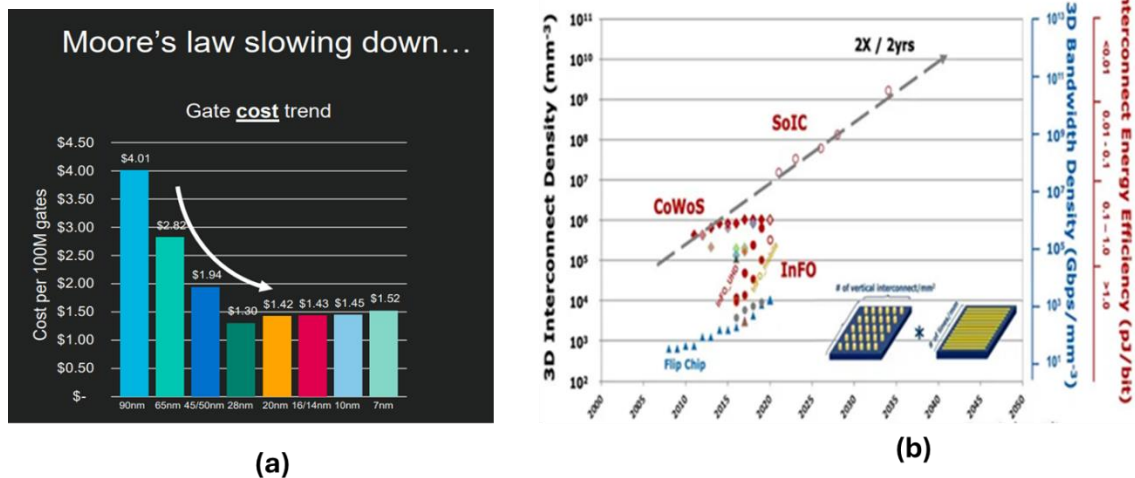


Figure 1 (a) Gate cost trends vs scaling; (b) Rising bandwidth and corresponding I/O density increase over time [3] [4]

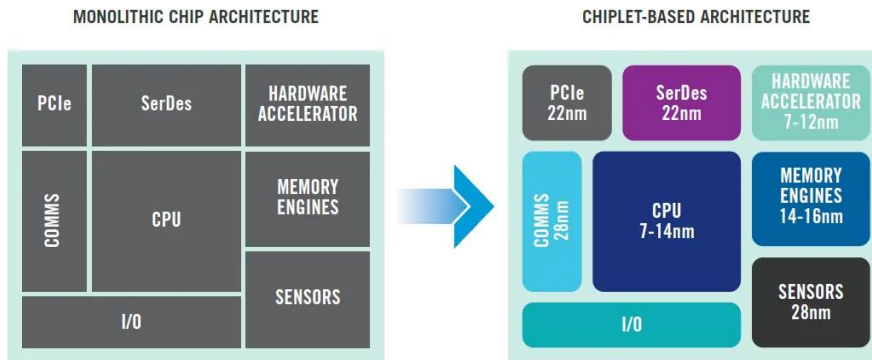
Moore's law has been able to extrapolate the rate of scaling for decades. However, there has been a recent slowdown in Moore's law, as evident in Figure 1, where an increase in cost per gate has been observed with scaling beyond the 28 nm technology node. The smaller technological nodes (below 28 nm) demand additional process steps, which add to the complexity of fabrication and hence increase the cost [5]. In addition, accommodating more components on a single chip, until recently, has been instrumental in lowering cost, but with continuous scaling, the yield has become more difficult [6].

Other major challenges with the monolithic integration are:

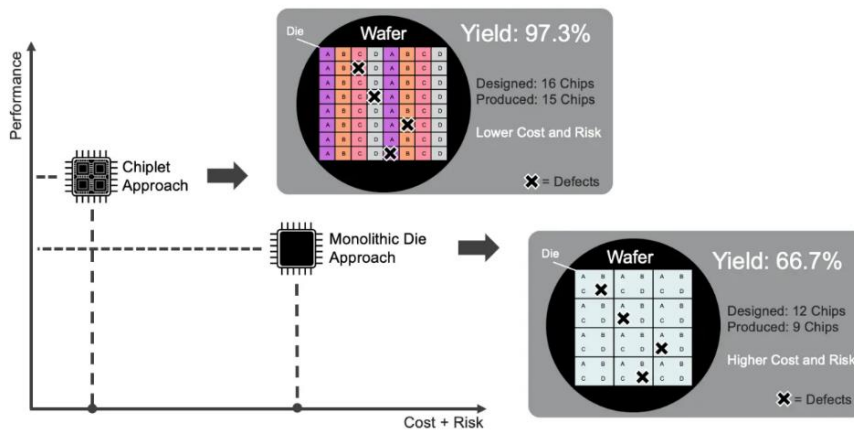
1. **Reticle limit:** The lithography tools in fabrication facilities have a reticle limit, which is essentially the largest area that could be accurately exposed [7]. This limits the silicon area, which needs to be increased to accommodate more computational components, achieving more computational capabilities and lower costs. Different methods, such as stitch chips and using multiple reticles, are being explored to go beyond the current reticle limit of ~ 26 mm x 33 mm [8].
2. **Power consumption:** With the monolithic integration approach, which has logic, cache, memory, passives, etc., on the same die, most of the power is consumed during data transmission through the interconnects in high-performance compute applications. Lengthy off-chip interconnects consume power and induce losses and delays [9].

In the backdrop of rising costs and challenging yield, new, high-performance computing technologies such as artificial intelligence, machine learning, the internet of things, etc., have emerged. These bandwidth-hungry applications pose multifaceted challenges with power delivery, signal integrity, thermal management, reliability, etc., which are difficult to achieve with conventional monolithic architecture.

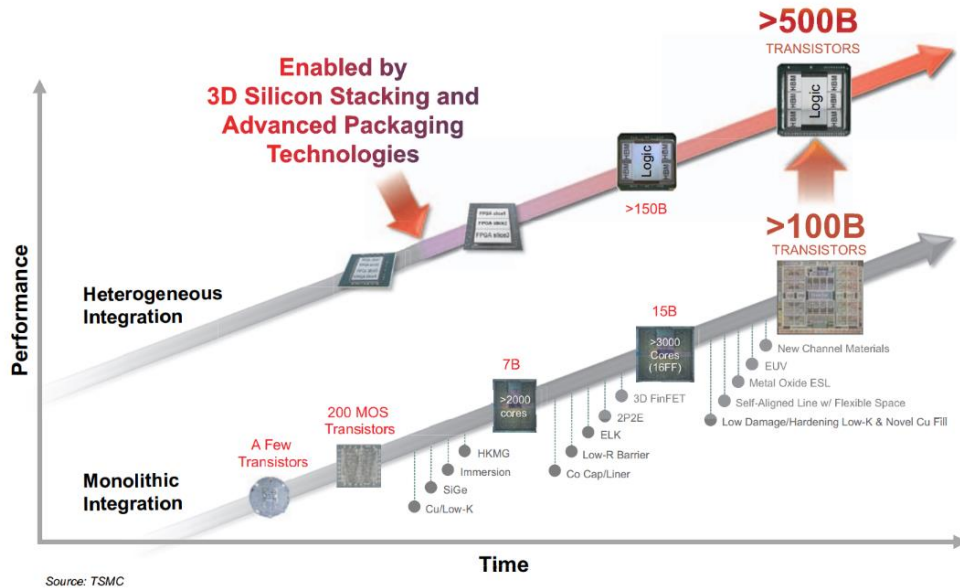
1.2 The Rise of Heterogeneous Integration



(a)



(b)



(c)

Figure 2 (a) Monolithic vs. chiplet integration architecture; (b) yield, performance, and risk analysis for monolithic vs. chiplet approach; (c) More silicon area and hence more transistor density with HI over monolithic architecture [10] [11][12]

Over the last decade or so, there has been a tectonic shift in the semiconductor industry, which is now gradually shifting away from scaling and process-centric innovations towards heterogeneous integration to cater to the high-performance requirements of modern-day devices. Heterogeneous integration allows disaggregation of a large monolithic chip into small functional chiplets, which are then integrated using

off-chip interconnect, as shown in Figure 2.a. Integrating good known dice via the chiplets approach helps increase yield, as shown in Figure 2.b. Other advantages include more silicon area for chiplets, which are now smaller and hence not limited by the reticle limit of the lithography tools, modularity, and increased flexibility as chiplets can be fabricated differently at application-specific fabs and integrated at last [13], etc. The increased silicon area allows more transistors, which leads to enhanced performance, as depicted in Figure 2.c.

In addition to the cost and yield benefits, enhanced performance and power efficiency are among the most important factors driving the shift toward heterogeneous integration. In conventional monolithic architecture, transistor scaling has paid dividends in terms of computing performance. However, there exists a gap between computing and memory access speed, which is exacerbated in the case of modern-day multi-core CPUs with shared memory [14].

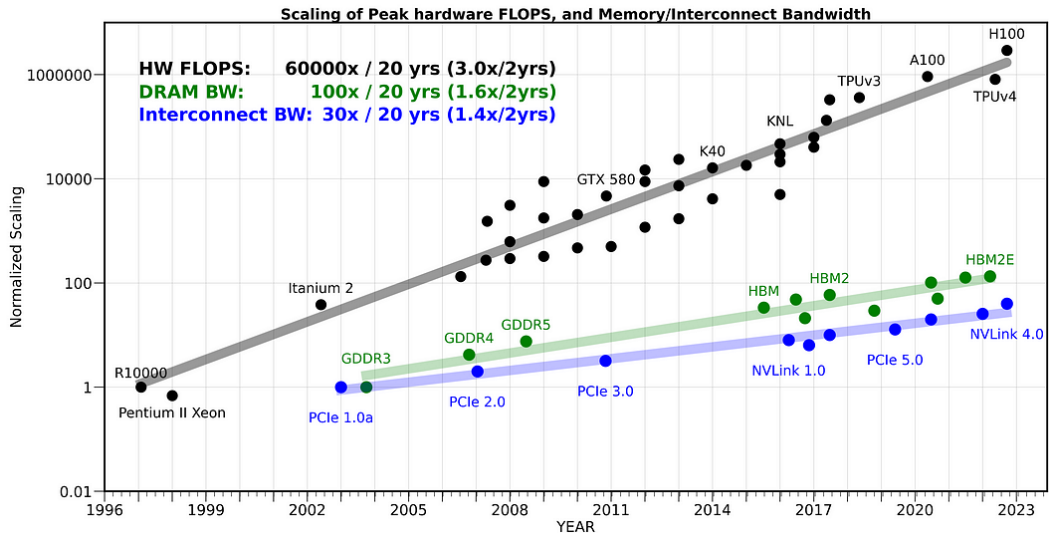


Figure 3 The difference in the scaling of the bandwidth for compute (measured in peak compute capability (i.e., FLOPS)), memory, and interconnects [15]

Heterogeneous integration has proved to be efficient in mitigating this gap by bringing the memory close to logic using shorter, high-density interconnects [16]. HI offers solutions such as High bandwidth memory (HBM) where DRAMs are vertically stacked or integration of GDDR adjacently to the processor using 2.5D HI [17].

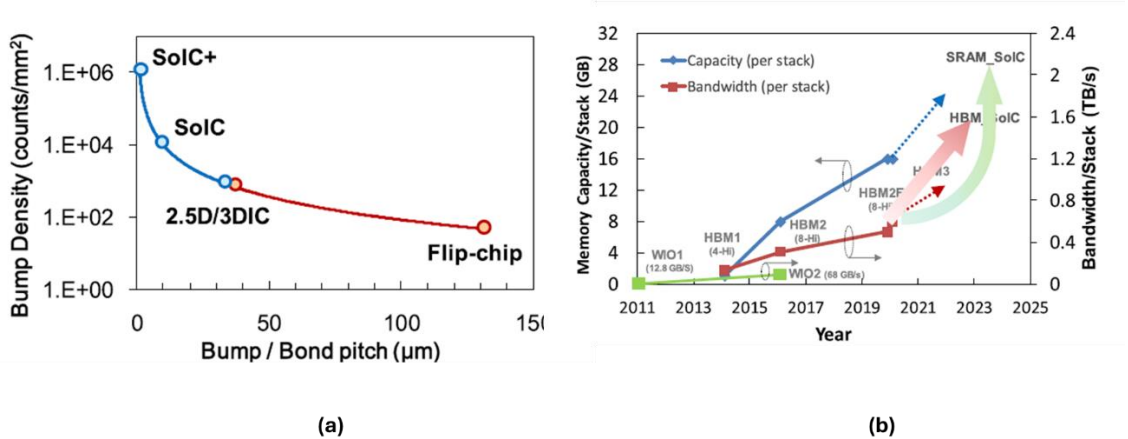


Figure 4 (a) I/O density vs pitch scaling; (b) Bandwidth/stack for different memory technologies [18]

Utilizing I/O pitch scaling, HI allows up to 10^6 interconnects/mm² rendering 1.6 Tbsp/stack bandwidth for technologies such as HBM, as shown in Figure 4 [18]. Along with bandwidth, latency is another key parameter determining compute-memory efficiency. High-performance applications require an extremely low latency, which is difficult to achieve with the monolithic package [19]. Next-gen interconnects in heterogeneous integration, such as hybrid bonds, enable sub-micron pitched, shorter I/Os, reducing latency and having large-scale implications for signal integrity. The fine-pitch hybrid bonds eliminate signal fan-out at the die-interconnect boundary, reducing I/O capacitive parasitics by 5x [20].

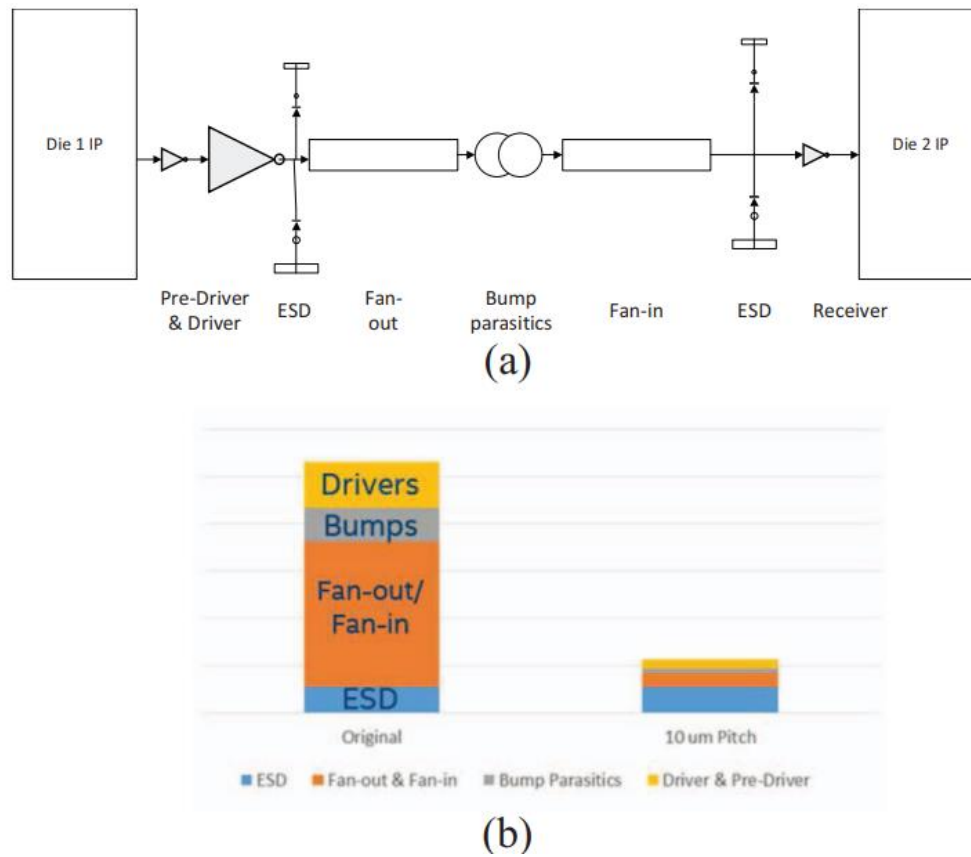


Figure 5 (a) simplified schematic depicting the route for die-to-die signal transmission with bumping technology; (b) Reduction in parasitics through use of hybrid bonding [20]

Further, using hybrid bonding in TSMC’s SoIC die-to-wafer technology, negligible insertion loss up to 30 GHz frequency has been reported, along with a considerable decrease in RLC parasitics compared to typical micro-bumps in conventional packages [21]. The low-loss transmission with scaled interconnects allows less than 0.05 pJ/bit energy consumption [22]. The effect of pitch scaling on performance can be better understood from Energy Efficiency Performance (EEP), which is essentially the product

of bandwidth density and energy efficiency [23]. A 16x increase in EEP was reported when the I/O pitch was scaled down from 9 μm to 2 μm [23].

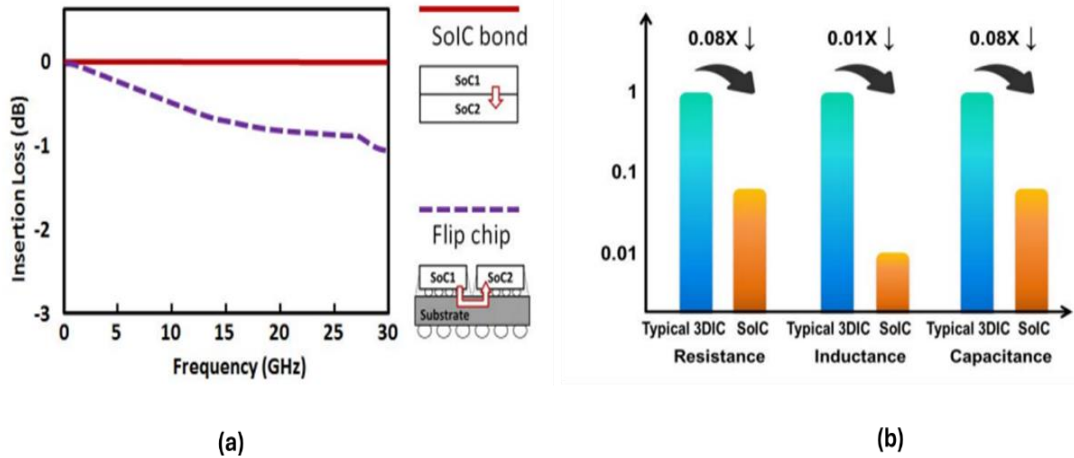


Figure 6 (a) Hybrid bonding (TSMC’s SoIC) vs. flip-chip using bump interconnect technology comparison for (a) Insertion loss; (b) RLC parasitics [21]

Given the array of economic and performance-centric benefits outlined above, heterogeneous integration, which optimizes space utilization and I/O scaling to achieve unparalleled yield, performance, and power efficiency, is being considered the way forward.

1.3 Recent Advancements in Heterogeneous Integration

1.3.1 Enhanced 2D Heterogeneous Integration (2.5D Heterogeneous integration)

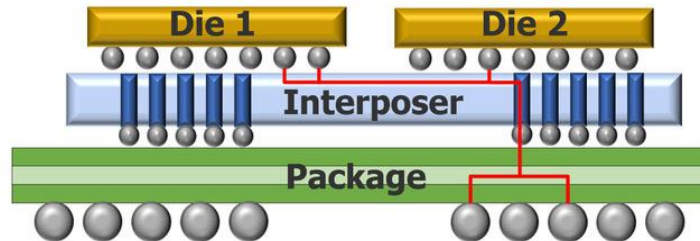
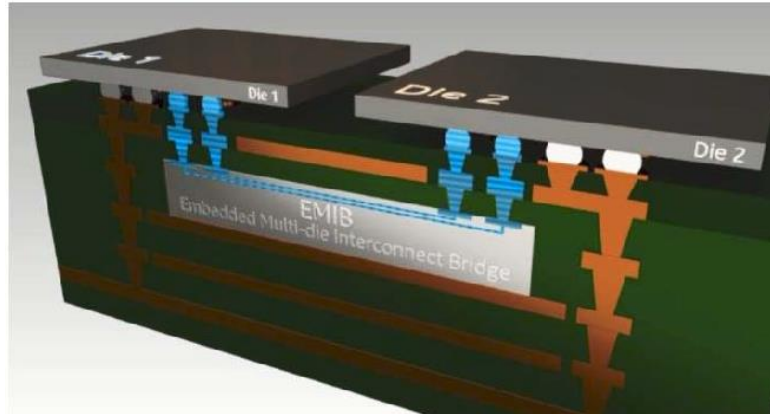


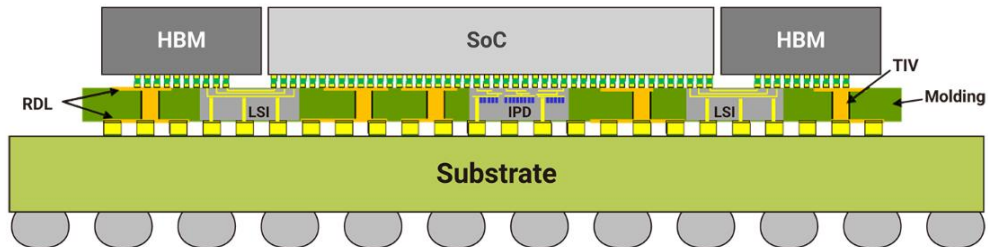
Figure 7 Conventional 2.5D architecture using an interposer [24]

The 2.5D architecture is an advancement from the traditional multi-chip module (MCM) configuration, where an additional interposer with through silicon vias (TSVs) and high-density RDLs is used between the chips and the package, as shown in Figure 7. Inter-chiplet communication occurs through the high-density metal traces on the interposer while power is delivered through the TSVs. Over time, several enhancements have been made to the conventional 2.5D HI, of which we discuss the following two: embedded silicon bridge and fan-out wafer-level packaging (FO-WLP).

(a) Silicon Bridge – Intel’s EMIB & TSMC’s CoWoS-L



(a)



(b)

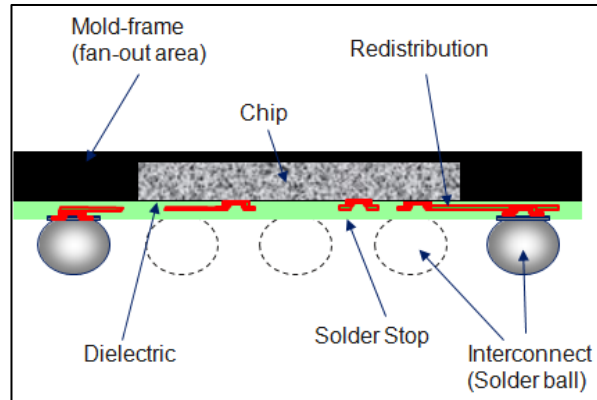
Figure 8 (a) Intel’s Embedded multi-die interconnect bridge (EMIB) using silicon bridge embedded in package itself; (b) TSMC’s Chip-on-wafer-on-substrate (CoWoS-L) using a silicon bridge embedded in an interposer [25] [26]

As the number of components that need to be interconnected increases, so does the size of the interposer. Due to the multiple processes involved in interposer fabrication, such as damascene, bumping, TSV fill, etc, the high cost makes it relatively unfit for high-volume manufacturing [27]. Hence, the silicon-bridge-chip emerged as a technology

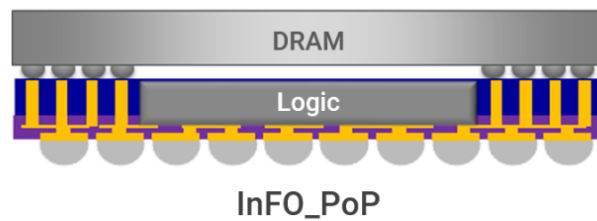
that enables the use of high-density I/Os while maintaining viable costs. The silicon bridge architecture utilizes a localized silicon bridge with high-density I/Os embedded into a low-cost organic laminate or a silicon interposer. Intel adopted the silicon bridge architecture for its embedded multi-die interconnect bridge (EMIB) technology, which can provide a bandwidth of 2 Gbps with 1.2 pJ/bit power consumption [28]. TSMC's Chip-on-wafer-on-substrate-L (CoWoS-L) packaging technology also uses the silicon bridge. However, a major distinction from EMIB was the bridge chip being embedded in a molding-based interposer rather than the package itself [26]. Using a local silicon interconnect (LSI) or silicon bridge in CoWoS-L reduces the cost for the interposer; however, the technology still uses TSVs.

(b) Fan-out Wafer-level Packaging – TSMC's InFO-WLP

Building on our previous discussions, it's clear that the I/O count significantly impacts bandwidth. In conventional wafer-level packaging, the integrated dice are diced out into packages, limiting the I/Os to the die area. However, in fan-out wafer-level packaging, the wafer is diced, and individual dice are reconstituted with an epoxy molding compound (EMC). After reconstitution, redistribution layers are built, extending through the package and allowing the I/Os to extend beyond the die area. This approach significantly increases the I/O count and enhances communication and bandwidth. [29].



(a)



(b)

Figure 9 (a) Traditional FO-WLP with the die embedded in mold; (b) Commercial application of FO-WLP in TSMC's InFO-WLP technology [30] [31]

FO-WLP has several advantages, such as reduced package thickness and, hence, thermal resistance, higher performance through shorter interconnects, etc [32]. TSMC used the FO-WLP for its Integrated fan-out wafer-level packaging (InFO-WLP) to stack DRAM over logic, as shown in Figure 9.b, reporting a bandwidth density of 10 Tbps/mm for a logic-logic integration using ultra-high-density InFO technology [33].

1.3.2 3D Heterogeneous Integration

The 3DHI involves vertical stacking of chiplets using off-chip interconnects such as micro-bumps, hybrid bonding, etc., and through silicon vias (TSV). 3D stacking enables reduced interconnect lengths, hence reduced parasitics with more power efficiency [34]. Conventionally, 3D stacking has been used using microbumps and underfill such as non-conductive film, epoxy etc. However, next-gen 3DHI technologies such as TSMC's SoIC, Intel's Foveros Direct, and Xperi's DBI Ultra are more inclined towards "bump-less" hybrid bonds, which can enable extremely fine-pitch I/Os and smaller form factor. These 3DHI technologies are detailed further in the next chapter.

1.4. Challenges with Contemporary Interconnect Technologies in Heterogeneous Integration

The interconnects in an HI system can be classified as on-chip interconnects, which include through silicon vias, re-distribution layers, etc., and off-chip interconnects, which include microbumps, solder bumps, copper pillars, hybrid bonds, etc., and are used for inter-chiplet communication.

The off-chip bandwidth is limited by I/O or pin count [35]. Hence, the off-chip bandwidth bottleneck lies at the off-chip I/Os, and continuous pitch scaling is required to increase the I/O count within the limited available chip area.

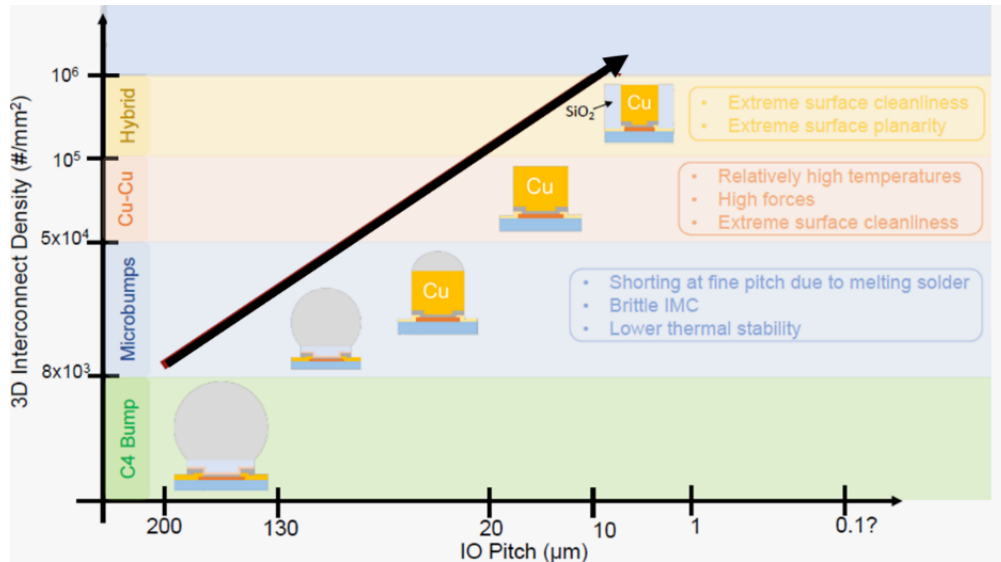


Figure 10 Bandwidth density and I/O pitch for different interconnect technologies [36]

Figure 10 demonstrates the off-chip I/Os scaling trends and the associated interconnect technology. Conventionally, solder and microbump technologies have been standard interconnect technologies in the industry. However, the formation of brittle intermetallic compounds (IMCs) along with bridging during assembly inhibits scaling of I/O pitch below 20 μm [37]. Direct Cu-Cu and hybrid bonding are two of the contemporary bonding schemes for fine pitch integration. Direct copper bonding enables fine-pitch assembly with lower inductance and improved thermal conductivity at the bonding interface compared to solder bumps [38] [39] [40]. Also, direct copper bonding can achieve low-temperature bonding via methods such as surface activation, surface passivation, etc. However, a post-assembly underfill is required in direct copper bonding for stress distribution and heat dissipation [41]. Due to its inherent viscosity, conventional epoxy underfills with filler constituents pose challenges at fine-pitch

dimensions. The viscous nature of the underfill affects its flow, and the uneven settling of the underfill leads to stress points within the package, which are detrimental to its reliability [42]. In addition, the viscous underfill requires a $10\ \mu\text{m} - 30\ \mu\text{m}$ inter-chip vertical gap [43] for penetration, which necessitates longer off-chip I/Os, which increase electrical parasitics.

Hybrid bonding is the present-day interconnect technology that is being extensively used for fine-pitch integration. Due to the presence of an innate dielectric layer at the bonding interface, the bonding scheme does not require an epoxy infill [43]. Also, hybrid bonding is surface-surface bonding, and the off-chip I/O lengths are minimal. Hence, the bonding scheme enables low parasitic signal transmission, lower thermal resistance, and reduced form factor for the overall package. Hybrid bonding is a highly mature commercial integration scheme and is used extensively in CMOS image sensors with wafer-to-wafer integration. However, applications such as HBM with tens of DRAMs (Dynamic Random Access Memory) stacked in a 3D architecture require die-to-die (D2D)/die-to-wafer (D2W) integration for enhanced yield and modularity [44]. However, particle control is a major challenge for D2D/D2W integration. The pre-bond dicing in D2D/D2W causes distorted edges and particles on the bonding surface in [45]. A particle of a few nanometers in diameter can induce a void 20x in diameter [44]. Given the potential impact of even a small particle, stringent particle control is not just a requirement but a critical aspect of the process. This need becomes more pronounced as the pitch decreases, making it a key consideration in fine-pitch integration.

Further, over recent years, glass, a low-loss, cost-efficient material, has emerged as a substrate technology for high-speed data transmission in the IC industry [46]. State-of-the-art hybrid bonding utilizes plasma activation of homogeneous dielectric surfaces such as polymer-polymer or oxide-oxide for bonding [47]. However, while silicon substrates use inorganic dielectrics such as oxide/nitrides for RDL due to process compatibility [48], emerging substrate technologies such as glass with tunable CTE [46] generally consist of cost-effective polymer-based RDL. However, hybrid bonding with non-homogeneous surfaces, such as polymer-oxide in Si-glass applications, can be difficult to realize.

The various interconnect technologies and subsequent challenges will be detailed further in the forthcoming chapters. However, it can be inferred that there are several challenges with contemporary integration schemes, and hence, a need for new bonding frameworks that not only push the boundaries of pitch scaling but also broaden the application spectrum to cater to upcoming substrate technologies such as glass.

1.5 Scope of Research

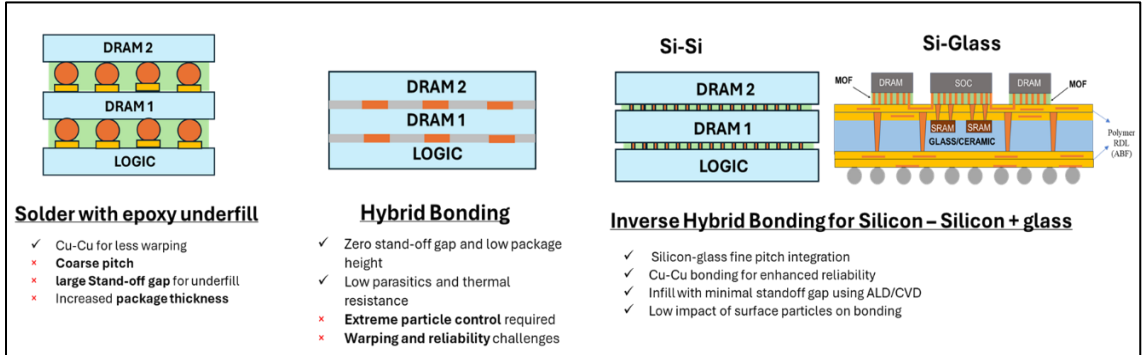


Figure 11 Challenges with contemporary fine pitch bonding schemes and research objective for the current work

We present IHB as a novel D2D/D2W integration scheme that utilizes direct copper bonding for the off-chip I/O interconnects with a post-bond, ALD-CVD deposited MOF as infill. Through this work, we aim to drive pitch scaling to smaller dimensions for D2D/D2W bonding by mitigating the stringent particle control requirements in contemporary hybrid bonding technology and reducing the impact of particles by utilizing Cu-Cu bonding as the first step against surface activation as used in hybrid bonding. However, unlike hybrid bonding, Cu-Cu bonding requires a considerable stand-off gap between integrated dice for conventional underfill post bonding, which renders large I/O lengths and is detrimental to both power efficiency and signal integrity. In this work, we aim to address the underfill challenge using ALD-CVD-based infill, which can provide uniform and full coverage with a minimal stand-

off gap between the dice (500 nm – 1 μ m). With comparable I/O length due to a small stand-off or die-to-die gap along with low-k MOF use as infill, we aim to achieve improved signal integrity with a low-form factor for the IHB integrated systems. Further, with IHB, we aim to enable fine-pitch silicon to non-silicon (Glass) integration which can be challenging with hybrid bonding due to the reasons discussed above such as material compatibility, CTE mismatch etc.

The present work demonstrates effective Cu-Cu bonding while exploring methods to reduce bonding temperature, as detailed in Chapter 3. Chapters 4 and 5 focus on the infill, where a sequential study of the ALD/CVD infill deposition parameters through a series of experiments has been presented. The insights from the experiments are then utilized to successfully demonstrate IHB in Chapter 5 with a 140 μ m I/O pitch sample having a 5 mm x 5 mm effective bonded area and a 1 μ m die-to-substrate vertical gap at first. Finally, the results and findings are presented for inverse hybrid bonding along with a fine pitch of 5 μ m and future work.

CHAPTER 2: LITERATURE SURVEY OF CURRENT INTERCONNECT TECHNOLOGIES

2.1 Need for High-density I/Os

Effective inter-module communication is the key to achieving high-performance computing, driving the need for high bandwidth density and low energy communication. A typical AI accelerator such as Intel's Gaudi 3, which consists of FPGAs, GPUs, ASIC, etc., may require up to a few Tbsp bandwidth [49], and with the advent of technology, the demand for bandwidth doesn't seem to be slowing down. While increasing the data rate does help increase bandwidth, it is accompanied by increased power consumption and signal losses [50]. Hence, increasing the I/O density is a better alternative to achieving higher bandwidth while maintaining feasible power consumption. Figure 12 demonstrates the I/O density trends, which suggest $\sim 10^8$ interconnects/mm² by 2030. Hence, new interconnect technologies that can enable extreme pitch scaling will be needed to meet this demand.

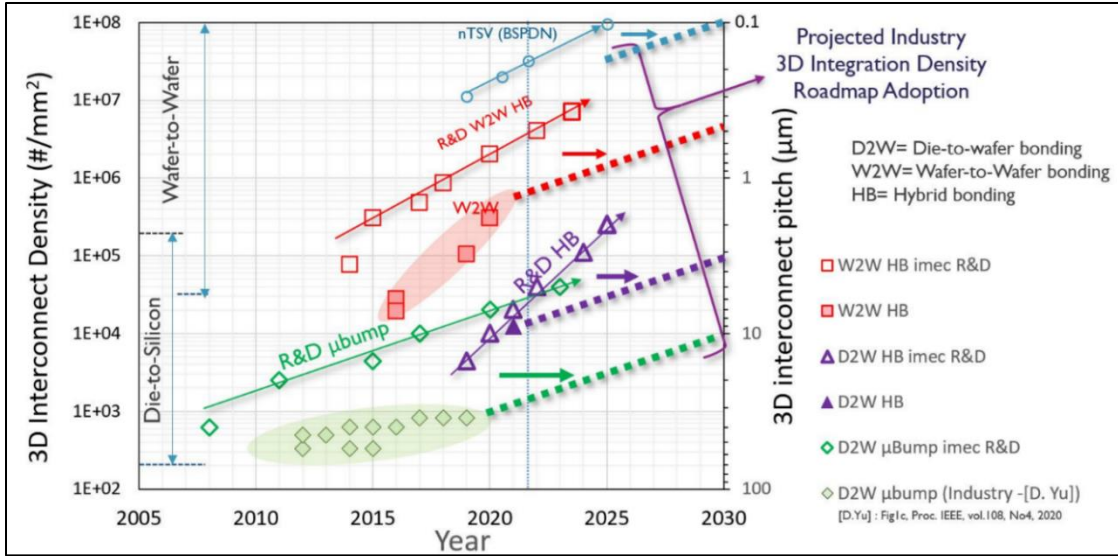


Figure 12 Projected I/O density and pitch scaling trends for different integration schemes such as Die-to-Die, Die-to-wafer etc. [51]

Conventional solder-based interconnect technologies have long been used as off-chip interconnects. However, bridging at fine pitch dimensions and reliability concerns due to the formation of brittle intermetallics are major challenges with solder-based interconnects, which inhibit pitch scaling below 20 µm. Two of the extensively used interconnect technologies for fine-pitch applications are direct copper bonding and hybrid bonding, which are discussed in detail below.

2.2 Direct Copper Bonding

Cu-Cu bonding emerged as a major interconnect technology to push the pitch scaling boundaries towards 10 µm. Compared to solder bumps, copper pillars allow pitch scaling without bridging and ensure better interconnect reliability due to the

absence of brittle intermetallic. In addition, properties such as electrical conductivity, resistance to electromigration, high thermal conductivity, process compatibility, etc. make copper an ideal fit for interconnects [52].

Conventional direct copper bonding or thermocompression bonding utilizes high temperature and force to support the inter-diffusion of copper interconnects. Due to its affinity towards oxygen, copper forms non-limiting oxides that impede inter-Cu diffusion and hence require 300-400 °C of bonding temperature [53]. The bonding process is then followed by a high-temperature anneal to further strengthen the bonds and mitigate stress. However, high bonding temperature induces warping in the package and affects the performance of temperature-sensitive devices such as DRAMs [44]. Hence, several methods to lower bonding temperature have been explored. These methods can be broadly classified into surface activation, passivation, and modification, which are discussed in detail below.

2.2.1 Surface Activated Bonding (SAB)

Surface activation has been demonstrated as a viable method to lower bonding temperature close to room temperature. The copper surface is activated in a high vacuum through the physical bombardment of argon ions. The ions physically remove oxides and, at the same time, activate the surface, which then can be bonded using cohesive forces at room temperature [54]. With no high temperatures required, surface-activated bonding mitigates challenges such as thermal stress and expansion-induced bonding misalignment [55]. However, the high vacuum requirement as a bonding condition makes the process commercially expensive [56].

2.2.2 Surface Modification

Different surface modification methods have been explored to achieve low-temperature Cu-Cu bonding. Direct copper bonding with (111) oriented nano-twinned copper is one such method where the high diffusivity of copper along the (111) plane is utilized to achieve effective bonding at temperatures as low as 150°C and under non-vacuum conditions [57]. Nano-twinned (nt) copper poses exceptional electrical and mechanical properties such as high resistance to electromigration and oxidation, better diffusivity at low temperatures, etc. However, fabrication challenges pertaining to the controlled orientation of (111) copper exist. Moreover, the considerably long electroplating time (10-30 min) [58], along with the requirement of costly additive stabilizers for nt-Cu electroplating renders the process difficult for high-volume manufacturing [59].

2.2.3 Surface Passivation

The copper surface is usually passivated with a thin metal capping to prevent oxidation. Different metal cappings such as Ag, Pd, Ti, and Au have been utilized in the past to demonstrate low-temperature Cu-Cu bonding. While titanium forms oxides easily and hence impedes effective bonding, a study demonstrated that Ag and Au are better passivating films than Pd, considering bonding strength and electrical properties [59]. However, a limitation of Ag and Au is the high cost and process compatibility

[60]. Hence, due to its properties, such as low resistivity and oxidation resistance, ruthenium has been explored in the present work as a capping material over copper I/Os.

2.3 Challenges with Underfill for Direct Copper Bonding

Having discussed the various advantages of direct copper bonding, such as non-bridging, better reliability, etc., a major challenge that impedes the pitch scaling capabilities of direct copper bonding below 10 μm is the underfill. An underfill is an epoxy-based insulating material that is inserted between the interconnects and consists of fillers for CTE tuning, adhesion promoters, etc. With its primary objective being stress distribution, an underfill serves various other purposes, such as electrical insulation, heat dissipation, etc.

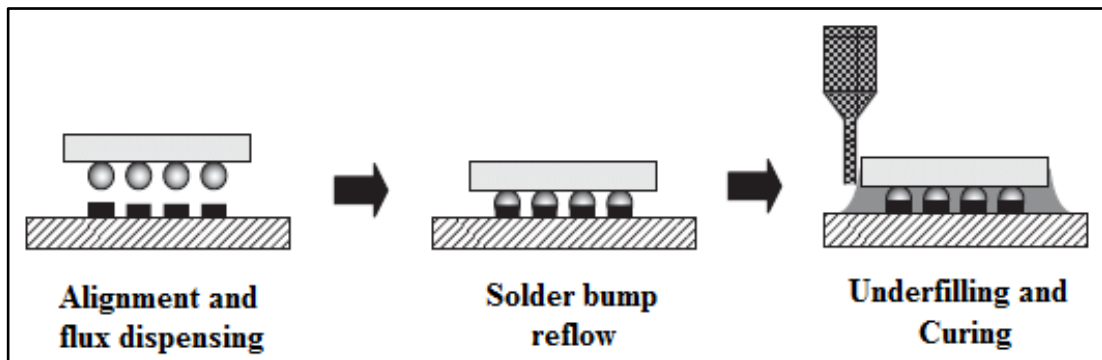


Figure 13 Capillary flow-based underfill dispensing process [61]

The underfill is generally dispensed using a weight-controlled needle along one side of the package and can flow through capillary action, covering the area between the interconnects [62]. Post dispensation, it is cured into a stiff material, which holds the

interconnects together. Material properties of the underfill play an important role in determining the package's reliability. Some of the preferred properties of an underfill material are as follows

Table 1 - Ideal parameters for an underfill [63]

Property	Viscosity [Pa s]	Thermal conductivity [$\text{W m}^{-1} \text{K}^{-1}$]	Tg [K]	CTE below Tg [ppm K^{-1}]	Volume resistivity [$\Omega \text{ cm}$]	Modulus [GPa]	Dielectric constant
Specification	<20	>1.0	>398	25-30	> 10^{12}	5-10	<4

One key underfill material property is the thermal expansion coefficient affecting the stress distribution capabilities of underfill. Silica fillers are an integral part of underfill, which are used to tune the latter's CTE and minimize the impact of stress. These fillers add to the viscosity of the underfill, which leads to uneven settling at fine-pitch dimensions, causing stress points within the package [42]. Further, the vertical die-substrate gap decreases with continuous miniaturization and shortening of off-chip I/Os for low-loss transmission. With a reduced gap and fine interconnect pitch, the viscous underfill material becomes difficult to flow through and deposit uniformly between the interconnects [43].

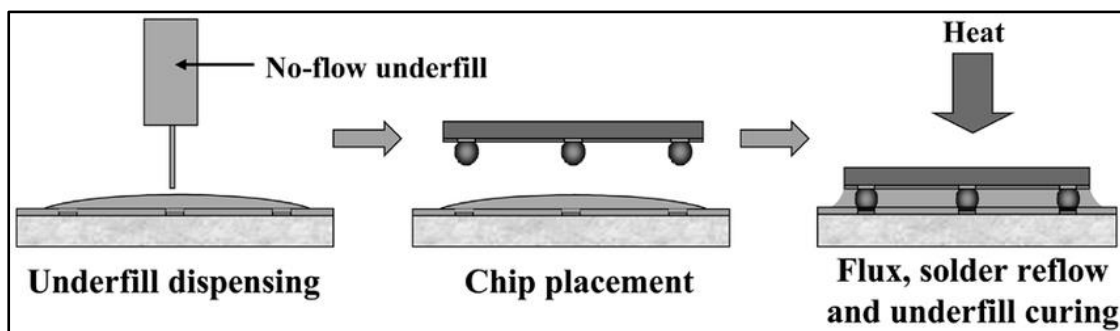


Figure 14 No-Flow underfill process [64]

An alternate underfill deposition method is no-flow underfill, where the underfill is dispensed before assembly. The underfill material is squeezed away from the I/O contact area on applying force during assembly and then cured. While advantageous for applications with small flow gaps, the method leads to enhanced voiding and high CTE of non-filler-based underfill used for no-flow dispersion [65].

Hence, the current work explores an ALD/CVD deposition process for underfill involving gaseous precursors against the conventional epoxy-based, highly viscous underfill for fine-pitch applications.

2.4 Hybrid Bonding

Hybrid bonding is a state-of-the-art bonding scheme with a demonstrated ability to scale I/O pitch to sub-micron dimensions. In hybrid bonding, dielectrics such as silicon dioxide or silicon carbon nitride and recessed copper are present on the surface. The wafer/die surfaces are activated using plasma, increasing the hydrophilicity. Thereafter, the dielectric surfaces form hydrogen bonds to facilitate initial dielectric-dielectric bonding when brought into contact at room temperature. Further, a high-temperature anneal is used to (a) convert temporary hydrogen bonds into strong covalent bonds, (b) expand and then allow diffusion of copper I/Os to form electrical connection [66]. The presence of innate dielectric at the surface eliminates the need for an underfill, and hence, extreme pitch dimensions can be achieved with this bonding

scheme [43]. Other advantages of hybrid bonding include reduced latency, higher bandwidth density, reduced power consumption due to shorter interconnects, lower form factor, etc.

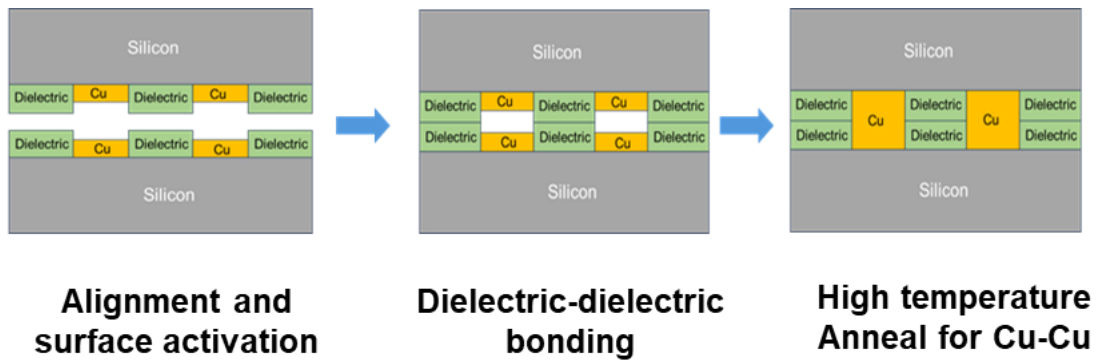


Figure 15 Process sequence for hybrid bonding

Initially used for CMOS image sensors (CIS) to stack image processors and sensors [67], today, hybrid bonding is being extensively used in applications ranging from high-end processors such as AMD’s EPYC [68] to High-Bandwidth-Memory (HBM) [69]. Hybrid bonding has two different configurations: die-to-die (D2D), die-to-wafer (D2W), and wafer-to-wafer (W2W).

2.4.1 Wafer-to-wafer Hybrid Bonding

In W2W hybrid bonding, two wafers that are fully processed with the front and back-end of the line are bonded using surface activation, followed by a high-

temperature annealing process, and then diced out into an integrated stacks. Wafer-to-wafer assembly configuration poses advantages such as high-throughput [70], better alignment accuracy, process compatibility, etc., and hence, is already being used for high-volume manufacturing CMOS image sensors and NAND Flash [71]. Additionally, integration with I/O pitch as low as 400 nm has been demonstrated with W2W hybrid bonding, which can provide extremely high bandwidth [72].

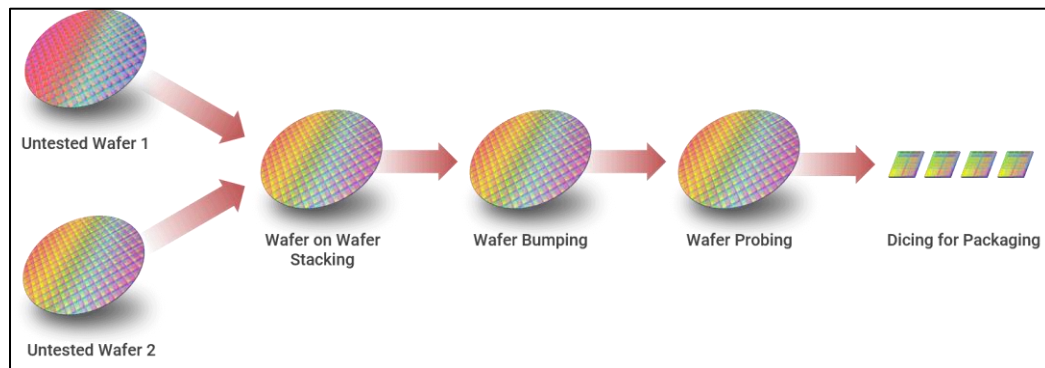


Figure 16 Wafer-to-wafer bonding [73]

However, the low yield and reduced flexibility are some of the major challenges in W2W hybrid bonding. Since the bonding occurs before dicing, the integrated dice must be of the same size. Further, the testing of integrated dice is performed after dicing, leading to low-yield [70].

2.4.2 D2D/D2W Bonding

For applications such as HBM where tens of DRAMs are stacked vertically, yield becomes a key factor, and hence D2D/D2W is preferred. Compared to W2W, where a single defect across the wafer can hamper the yield, D2D/D2W assembly utilizes tested,

known good dice (KGD) for integration, which helps increase yield [70]. Below are some of the commercial applications of D2D/D2W assembly, which signify the demand for the technology.

(a) AMD 3D V- Cache Technology

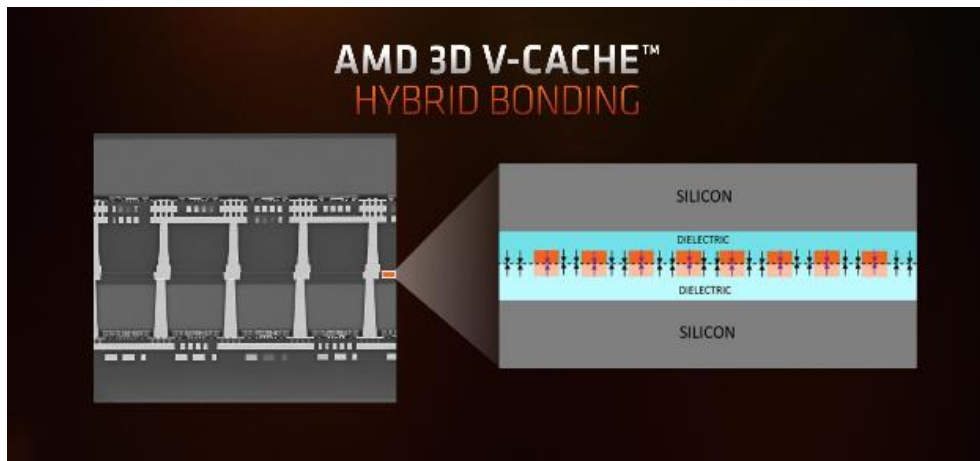


Figure 17 AMD 3D V-Cache using die-to-wafer hybrid bonding [74]

The latest 3D V-Cache technology from Advanced Micro Devices (AMD) employs hybrid bonding for stacking L3 cache over logic for its consumer products, such as AMD Ryzen 7000X3D desktop CPUs and EPYC processors for high-performance computing applications [68]. Through the adoption of hybrid bonding in the 3D V-Cache technology, an increase of 200x and 15x in interconnect density has been reported as compared to contemporary 2D and 3DIC interconnect technologies, respectively, with an achievable bandwidth of 2 TB/s [75]. In addition, the shorter, “bumpless” I/Os, as used in 3D cache hybrid bonding technology against microbumps

used in conventional 3DHI, reduces the energy consumption per bit by a factor of three. The cache-on-logic technology in the discussion here essentially utilizes TSMC's SoIC die-to-wafer hybrid bonding [76], enabling heterogeneous integration of tested cache dice over the compute chips with an I/O pitch as low as 9 μm [77].

(b) Intel's Foveros Direct

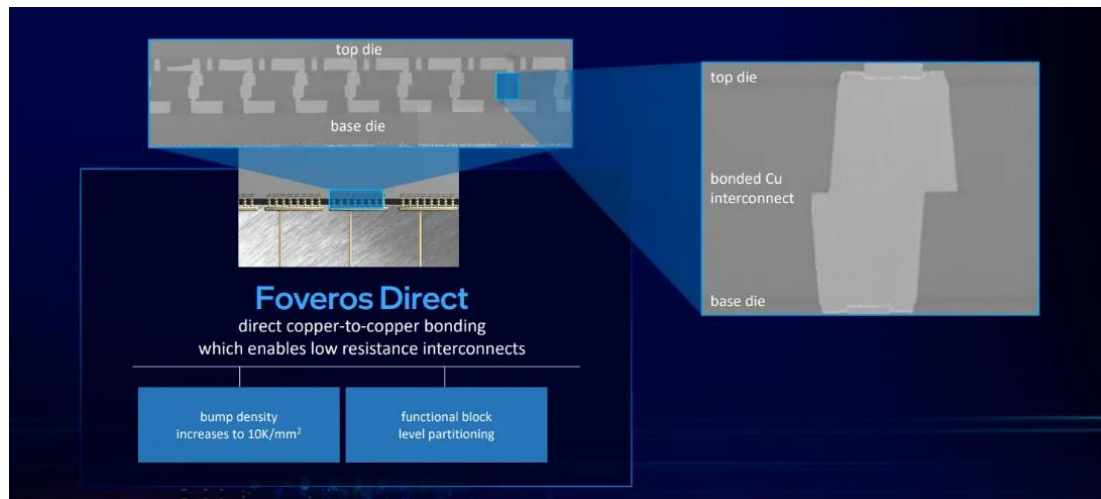


Figure 18 Foveros Direct technology by Intel using hybrid bonding [78]

The Foveros Direct is Intel's die-to-wafer hybrid bonding technology, which is being leveraged for stacking the Clearwater Forest Xeon CPUs [79]. The Foveros Direct utilizes hybrid bonding instead of thermocompression Cu-Cu bonding [80], as used in Intel's previous generations of Foveros and Omni, to achieve an I/O pitch of 10 μm with a density of 10K/mm² [78].

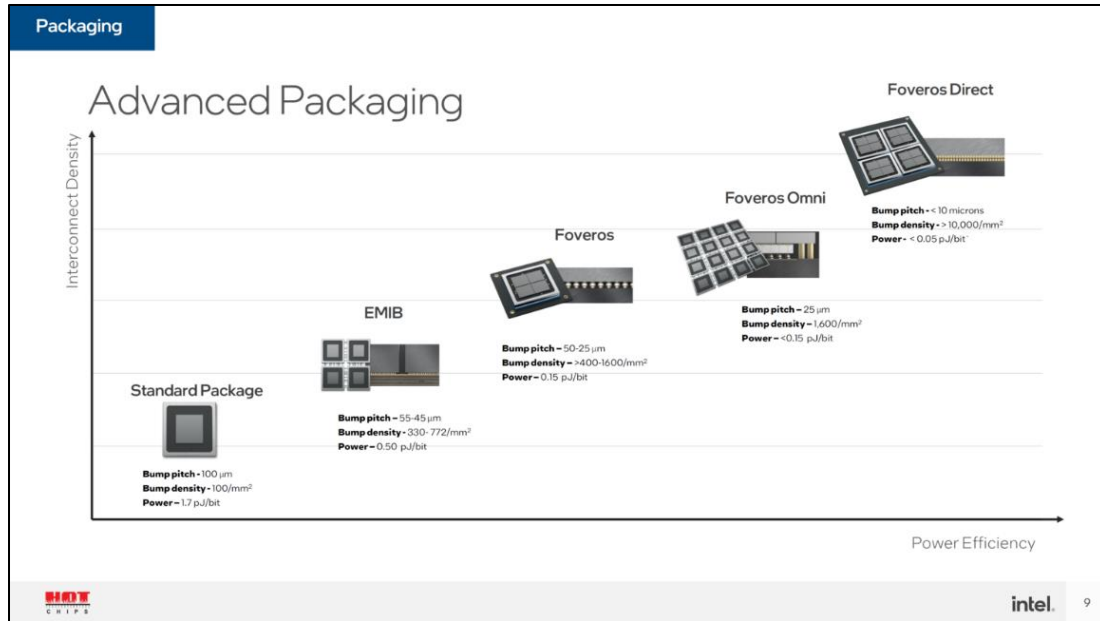


Figure 19 Evolution of Intel’s interconnect technologies [81]

With Foveros Direct, Intel expects to reduce the power consumption to 0.05 pJ/bit from 0.15 pJ/bit in Foveros Omni, accounting for a 3x reduction. For the second-generation Foveros Direct, Intel aims to scale the I/O pitch further to 3 μm [81], putting it in close competition with TSMC’s SoIC. While 3D stacking via hybrid bonding is expected to pay dividends regarding bandwidth, power efficiency, etc., challenges such as thermal management, power delivery, interference, etc., persist and must be addressed. Earlier generations of Foveros have circumvented the heat dissipation challenge in 3D stacking by allocating most computing components to the top-tier dice [82]. However, with a multifold increase in I/O density, thermal management would be a big challenge ahead.

(c) Hybrid bonding as an emerging technology in HBM

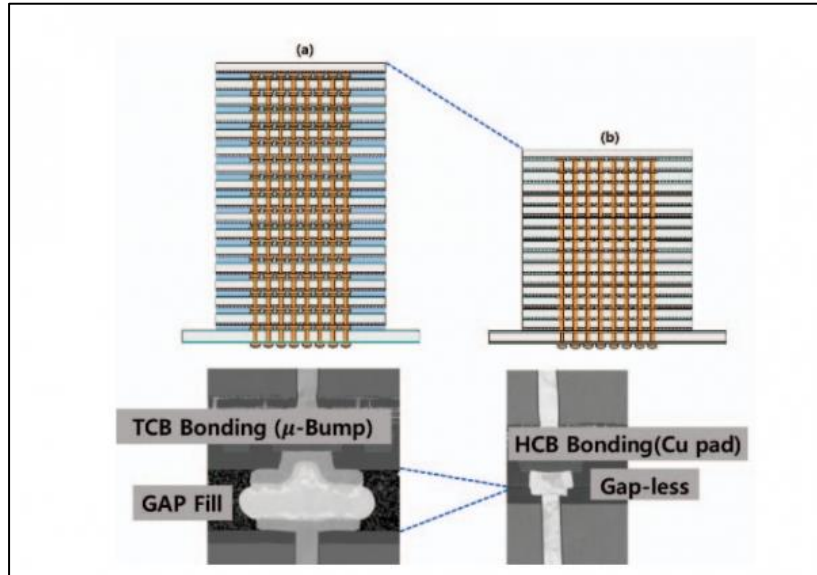


Figure 20 Low form factor advantage of using hybrid bonding over conventional TCB bonding using microbumps [83]

Performance-centric applications such as GPUs and data centers require HBM. The extremely thin DRAMs are connected through interconnects such as microbumps and sustain challenges such as warping. Thermo-compression bonding mitigates the warping issue to an extent and is one of the main reasons for using the same for 3D integration instead of hybrid bonding till now [84]. However, with increasing DRAMs in an HBM stack, the form factor or the stack height becomes an issue that the industry tends to address using hybrid bonding. Being a bump-less interconnect technology, hybrid bonding can reduce the form factor and hence is being adopted in the upcoming HBM technologies of multiple companies. SK Hynix is set to utilize hybrid bonding for their upcoming HBM 4 technology, including 16 DRAMs stacked together [84]. Along

with keeping the stack height intact at 720 μm even with 16 DRAMs in the stack [84], the company expects other benefits such as reduced thermal resistance and lower achievable pitch ($<20 \mu\text{m}$) [85]. Furthermore, when used for SK Hynix's HBM 2E with an 8-HI stack, hybrid bonding was found to pass the reliability test, which further encourages its use in HBM applications [86]. However, the low yield [87] observed with hybrid bonding is yet another challenge that needs to be resolved for adopting hybrid bonding in high-volume HBM manufacturing.

Similarly, Samsung has been utilizing thermocompression bonding with a non-conductive film (NCF) as an underfill. [88]. It helps mitigate warping in the case of thin DRAM stacking. However, as Samsung moves towards the next-gen HBM technology, hybrid bonding may become necessary, as evident from its demonstration of a 16-HI HBM sample [89].

2.4.3 Challenges in D2D/D2W Hybrid Bonding

Hybrid bonding for D2D/D2W integration has several challenges, such as misalignment, low thermal budget, pristine and defect-free bonding surface requirements, etc. The same can be inferred from Figure 21, which illustrates the roadmap for pitch scaling by IMEC for various integration schemes. As evident, while W2W integration is set to reach sub-micron pitch dimensions in the forthcoming years, in contrast, D2D/D2W I/O pitch ranges between 2-10 μm , which is comparatively large.

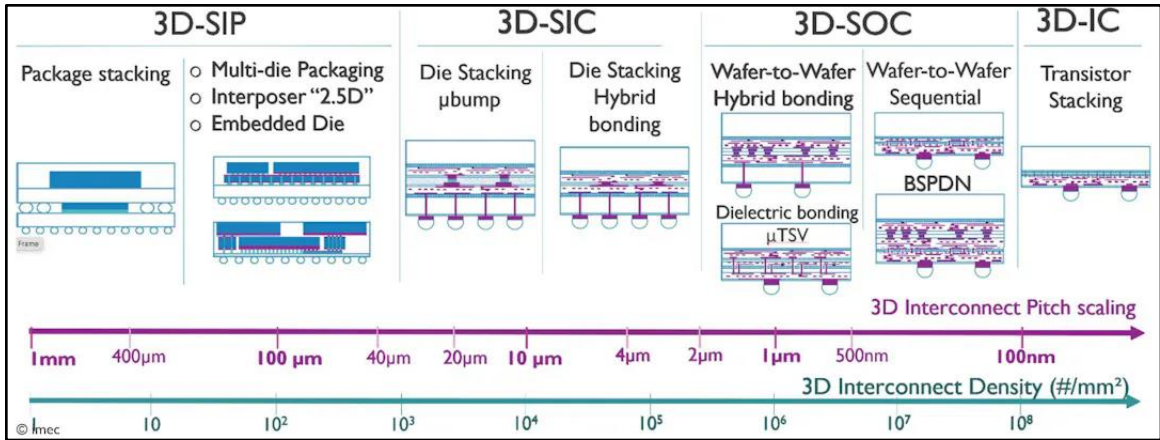


Figure 21 The IMEC 3D interconnect technology landscape [90]

(a) Alignment accuracy

Alignment is a major challenge, especially for D2D/D2W assembly, which is limited by the tool's capabilities. For D2D/D2W, flip-chip bonders are used, where individual dice are assembled over another die/wafer one at a time, as shown in Figure 22. This D2D/D2W assembly method provides less throughput, and the limited placement accuracy of the tools makes the process challenging [91].



(a)



Figure 22 D2D/D2W assembly using (a) pick and place tool; (b) Carrier wafer [91]

Alternatively, for D2W, improved placement accuracy is achieved by assembling individual dice on a carrier wafer with the help of thermal adhesives, followed by wafer-carrier hybrid bonding. The carrier wafer is released after bonding, so die-to-wafer assembly is achieved through an intermediate W2W step. While collective die-to-wafer bonding does improve bonding alignment, the limited thermal budget of the adhesive is a major challenge. Hence, novel methods such as self-assembly using water droplets are being explored to improve alignment accuracy for D2D/D2W processes [92]. For self-assembly, the placement areas are selectively wetted using surface treatment, making them more hydrophilic. Next, the capillary force of the water droplets converts the initial coarse alignment into fine alignment, as shown in Figure 23.

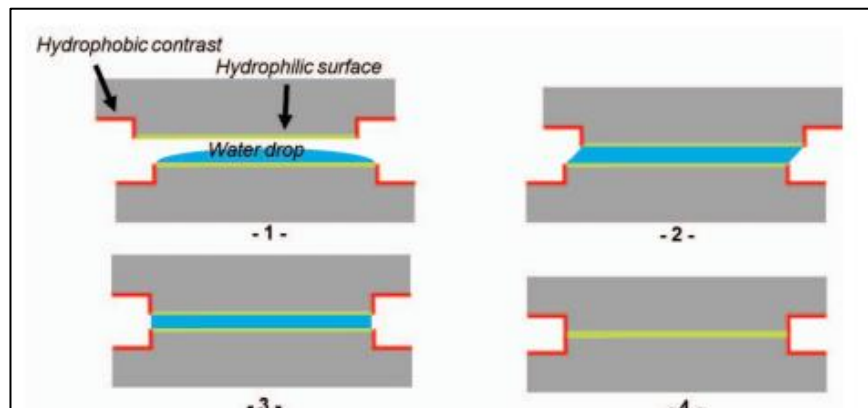


Figure 23 Self-alignment process using water droplets [93]

(b) Low-thermal budget

More than the stress and performance degradation effects of high bonding temperature, the thermal budget in the case of D2D/D2W assembly is limited by the stability of the thermal adhesive/temporary bond material used to attach individual dice onto the carrier wafer for collective bonding [94]. The temporary bond material is usually an organic glue with a thermal budget of 250°C, which is insufficient for effective bonding. Further, the instability of this organic glue at higher temperatures causes non-clean release with residues.

Memory devices such as DRAMs also sustain performance degradation during high-temperature anneal in hybrid bonding. Hence, low-temperature direct copper bonding methods are being explored for D2D/D2W.

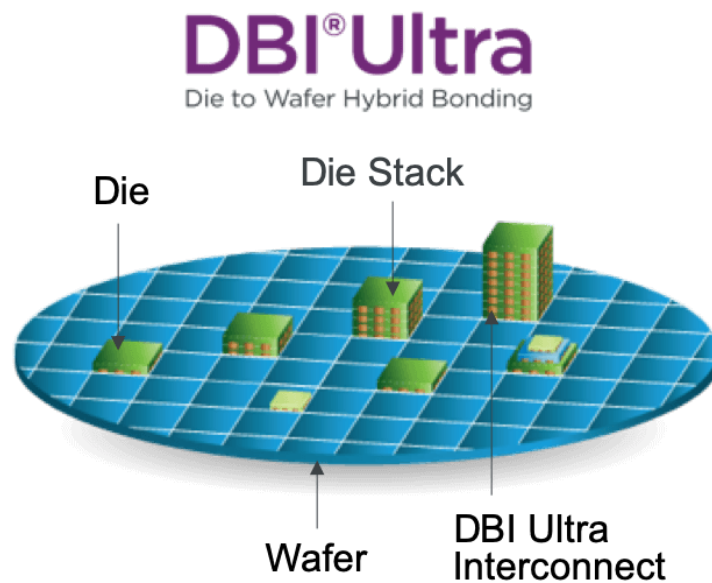


Figure 24 Xperi's DBI Ultra D2D/D2W technology [95]

Xperi's Direct Bond Interconnect Ultra (DBI Ultra) is one such D2D/D2W bonding technology that can achieve effective bonding at a temperature as low as 200 °C. DBI Ultra utilizes an optimized CMP process to achieve low-temperature bonding. [72] and has been licensed by SK Hynix for its next-generation memory devices [73]. While the technology is a big step toward commercializing D2D/D2W hybrid bonding, challenges such as extreme surface cleanliness persist as a major blockage, which will be discussed in this section.

(c) Surface cleanliness

In the context of hybrid bonding, the importance of maintaining an extremely clean and pristine surface cannot be overstated. Fine pitch assembly via hybrid bonding requires an ISO 3 environment with particle control up to 0.2 μm and surface roughness of 1 nm. Even a single particle on the surface can lead to a void of 20x the diameter of the particle, and the particle sensitivity worsens with smaller pitch dimensions. Given that an individual die undergoes multiple processing steps, it is challenging to eliminate particles. Therefore, the maintenance of a clean surface is a critical aspect of the die-to-die and die-to-wafer assembly processes.

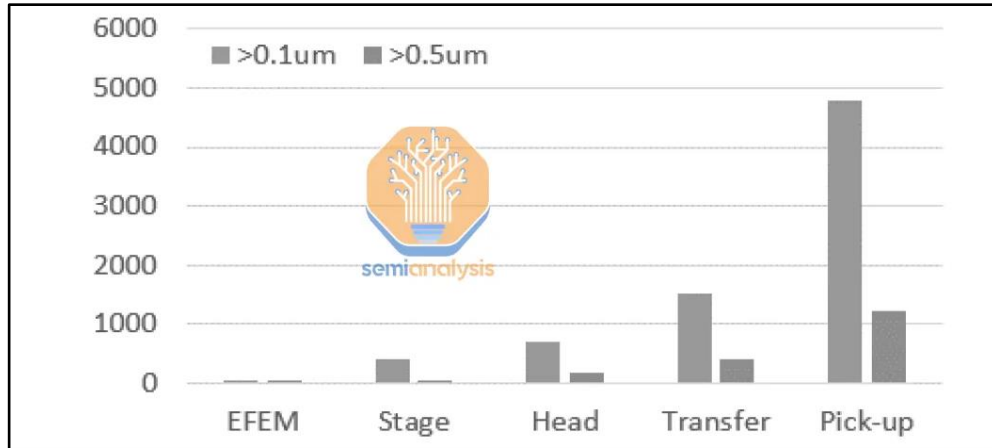


Figure 25 Number of particles generated by various sources during C2W hybrid bonding [76]

While most of the particles come from the process tools, as shown in Figure 25, dicing is a major contributor to particles in D2D/D2W assembly. In contrast to W2W, the D2D/D2W integration involves a pre-bond dicing step which conventionally is blade-based dicing. The blade dicing induces particles on the surface and edge chipping along the dicing line, which is detrimental to the bonding process. Alternatively, stealth laser dicing is used for minimal chipping, but it comes with a higher tooling cost, and the particle generation issue persists [96]. Hence, particle sensitivity is one of the biggest hurdles in pitch scaling with hybrid bonding for D2D/D2W bonding.

2.5 Conclusion

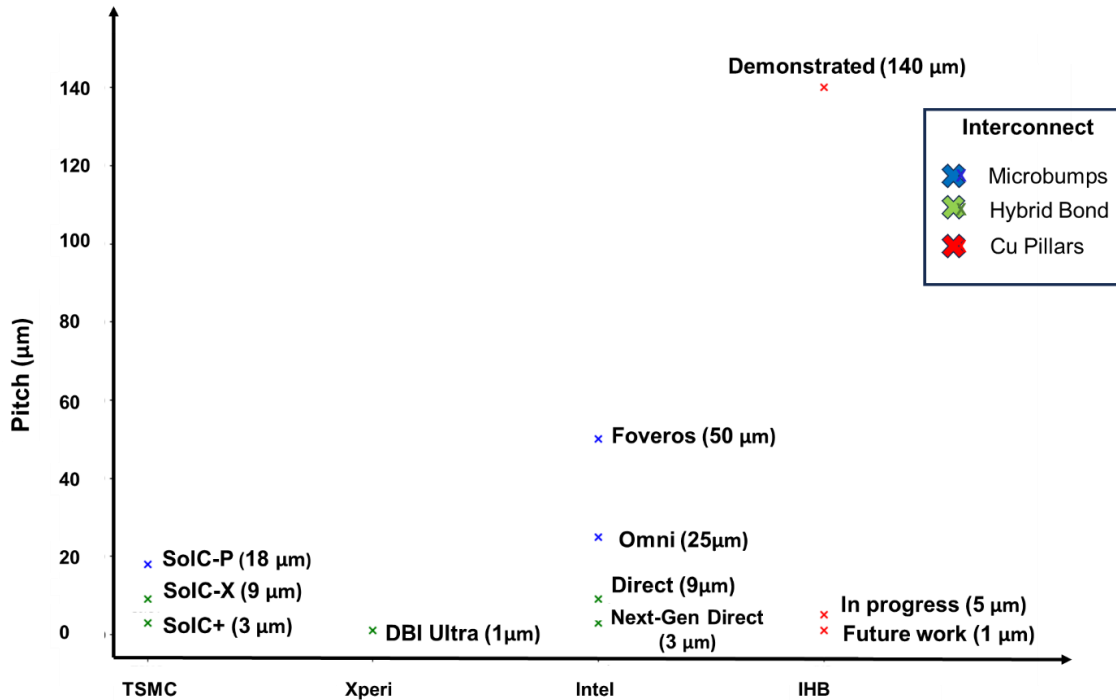


Figure 26 IHB in comparison to state-of-the-art D2D/D2W bonding technologies

Through this chapter, we have discussed two state-of-the-art fine pitch interconnect technologies: direct copper bonding and hybrid bonding. We identify low-temperature bonding and underfill coverage with a small stand-off gap as major challenges with Cu-Cu bonding. Similarly, we have focused on existing D2D/D2W hybrid bonding technologies in high-volume manufacturing, and we infer particle control and low-temperature Cu-Cu bonding as major hurdles impeding pitch scaling for D2D/D2W bonding. The challenges identified align with our research objectives, where we aim to (a) explore methodologies to lower bonding temperatures, such as using formic acid pre-treatment and ruthenium capping, and (b) provide an efficient underfill deposition with a low stand-off gap.

CHAPTER 3: DIRECT COPPER BONDING

Direct copper bonding has been used in present work to form off-chip I/O connections and is the first step of IHB. The following sections demonstrate direct Cu-Cu thermocompression bonding using formic acid pre-treatment on a fabricated testbed. The bonds were characterized with four probe resistance measurements, SEM cross-sectioning, and shear testing. Further, we explored ruthenium as a viable passivation layer over copper as a method to achieve low-temperature Cu-Cu bonding. While the bonding with ruthenium as capping could not be achieved successfully, the findings have been discussed, and future direction are illustrated.

3.1 Testbed for Direct Copper Bonding

A testbed was fabricated to evaluate the direct copper bonding (DCB) process. As shown in Figure 27, the testbed included a top (5 mm x 5 mm) and bottom (11 mm x 5 mm) die. On the front surface, the top die consisted of 15 nm/150 nm tall Ti/Cu pads on top of which were 350 nm tall copper pillars, rendering the total copper pillar thickness as 500 nm. Similarly, the bottom die consisted of 500 nm tall copper pillars with metal traces leading to contact pads on the periphery which were used for four-probe resistance measurements. The copper I/Os were distributed over an array of 80 x 80 with a pitch of 50 μm . Figure 28 depicts the fabrication steps used for the top die. The fabrication steps for the bottom die were similar, wherein the first metal layer, i.e., Ti/Cu (15nm/150 nm), was deposited using PVD for the metal traces, contact pads, and pillars, followed by elevation of Cu pillars by 350 nm through the second PVD-metal deposition step leading to 500 nm tall pillars.

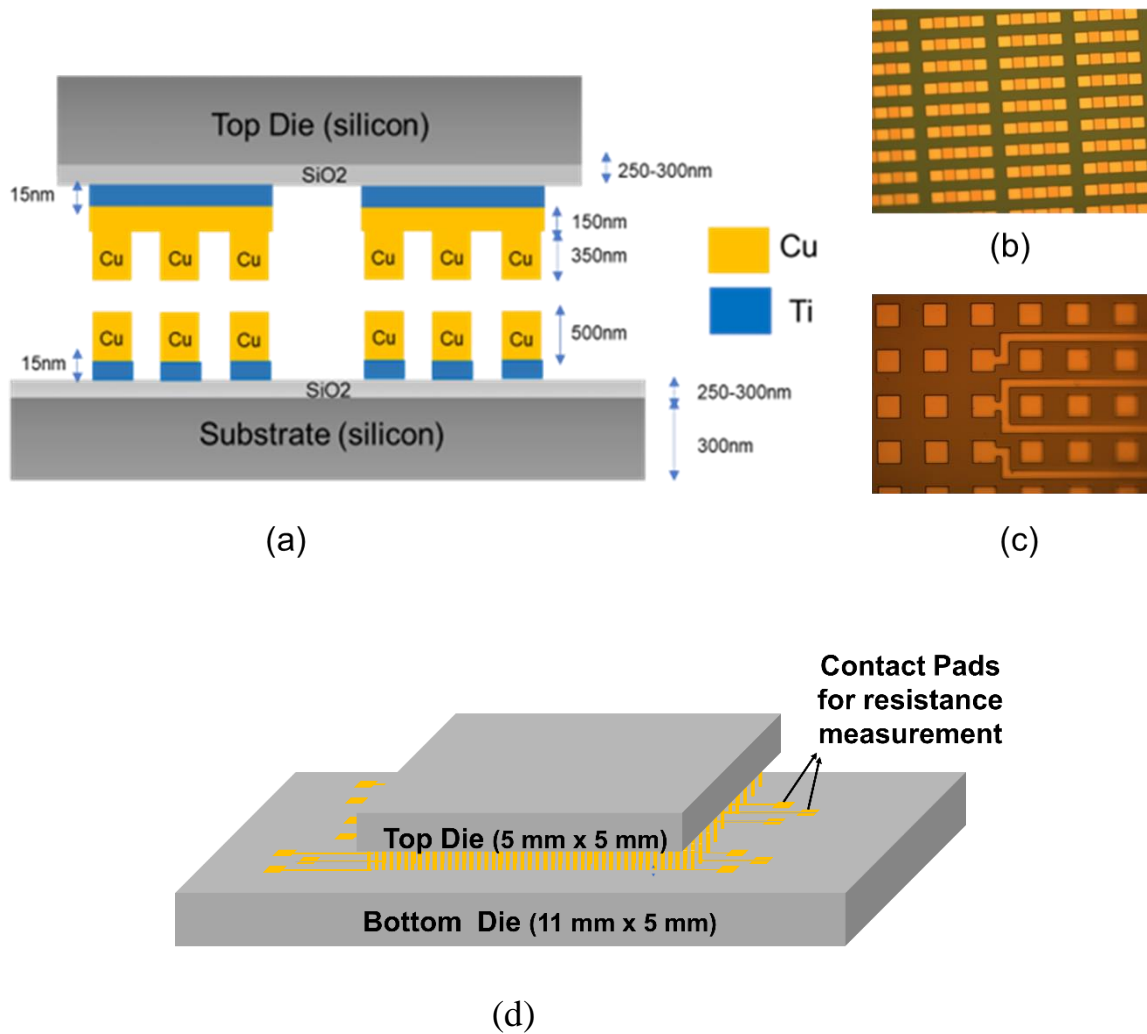


Figure 27 (a) schematic of the testbed depicting the cross-section; (b) microscope images of top die depicting copper pillars on top of pads; (c) microscope image of bottom die depicting copper pillars and metal traces; (d) schematic of the integrated testbed used for Cu-Cu bonding study

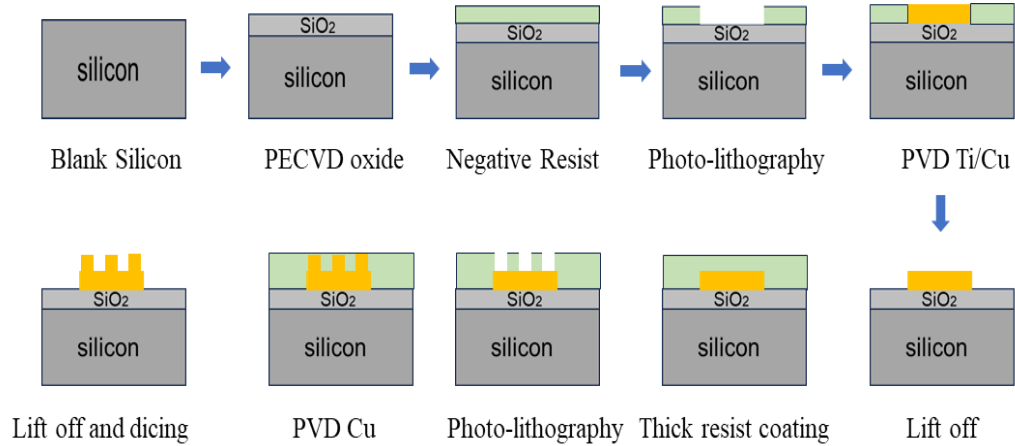
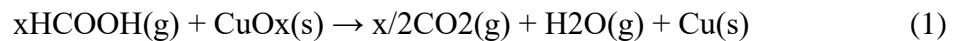


Figure 28 Fabrication steps for Cu-Cu bonding testbed

3.1.1 Flip-chip bonding process

The Cu-Cu bonding was performed using Finetech’s Lambda 2 flip-chip bonder with an in-built formic acid module. The bonding process was carried out in two steps i.e. pre-clean using formic acid and bond/mount. The formic acid pre-clean was used to remove the oxides on the copper surface. Copper has a high affinity towards oxygen and is prone to oxidation, which is detrimental to the bonding process. For the pre-clean step, the top and bottom dice were aligned using Finetech Lambda 2 flip-chip bonder, and while being kept at 0.5 mm (about 0.02 in) apart, the dice were exposed to high-flowing (3 L/min) formic acid gas at 200 °C for 5 min to obtain an oxide free surface. Formic acid reacts with surface copper oxide to form copper formate, which disassociates into CO₂ and H₂O above 180 °C, yielding a clean bonding interface [97] :



The residues were purged out through the in-built exhaust system in the flip-chip bonder. Thereafter, the copper I/Os on the dice were brought into contact and bonded at 300 °C using a force of 400 N for 5 min. Table 2 illustrates the bonding parameters used for the flip-chip assembly.

Table 2 - List of bonding parameters used for flip-chip assembly

Bonding parameters								
Parameters	Pre-clean				Main profile			
	Chip	Substrate	Nitrogen flow	Formic acid flow	Chip	substrate	Nitrogen flow(L/min)	Formic acid flow(L/min)
Temperature(°C)	200	200	3	3	300	300	3	1
Time	5min				5min			
Force	x				400N			

The bonded copper I/Os array in the testbed also included Kelvin structures, as shown in Figure 29, which were used for resistance measurements. Four probe resistance measurements using multiple Kelvin structures were used to analyze the bond uniformity and electrical resistance. An average resistance of $5\text{m}\Omega \pm 5\%$ per I/O was achieved over a 50+ sample size. Further, the bonds were optically characterized using an SEM cross-sectioning, as shown in Figure 29.b . The absence of a distinct boundary at the bonding interface depicts the effective Cu-Cu diffusion and hence, good bond quality.

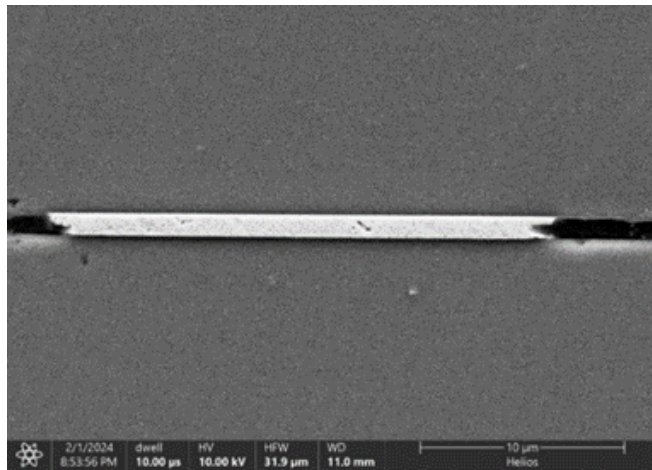
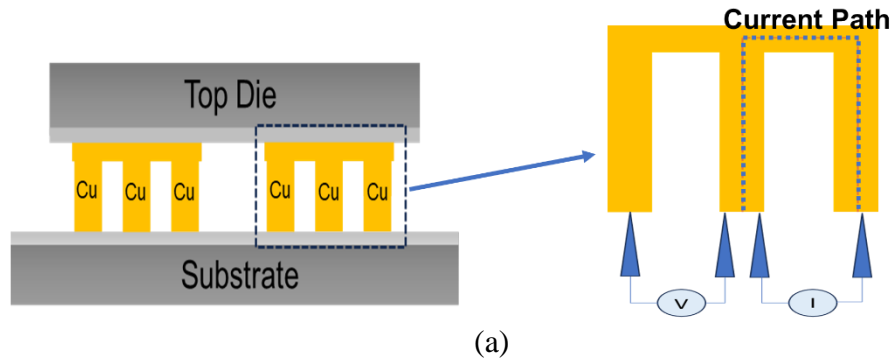


Figure 29 (a) Kelvin structure for per I/O resistance measurement; (b) SEM Cross-section of the bonded copper pillar

3.2 Ruthenium for Low-Temperature Cu-Cu Bonding

Due to its resistance to oxidation and good electrical conductivity, ruthenium was explored as a capping layer over copper I/Os to achieve low-temperature bonding. A testbed, as shown in Figure 30, was fabricated.

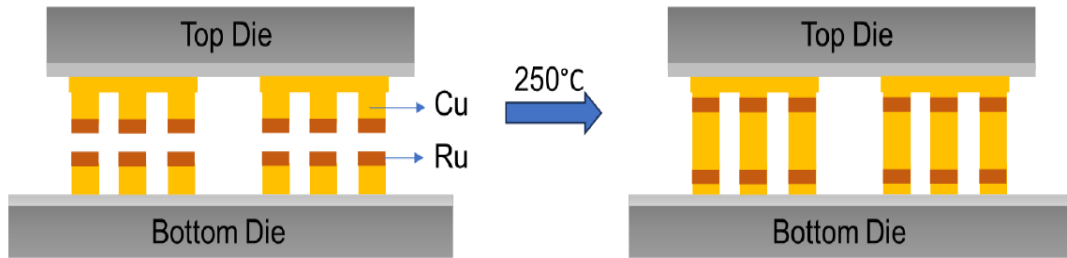


Figure 30 Testbed for Cu-Cu bonding with Ru capping

The testbed consisted of 50 μm pitch copper pillars capped with 5 nm ruthenium and was similar to the one used for evaluating direct copper thermocompression bonding in the previous section. A thin film of ruthenium was sputtered on Cu I/Os with 100 W argon plasma at 3 mtorr. The sputtered ruthenium was then reacted with ruthenium cyclopentadienyl ethyl tantalum (CpET) in an ALD chamber and annealed under forming a gas to achieve uniform ruthenium capping on copper, as shown in Figure 31. Selective deposition of ruthenium over Cu I/Os was achieved using a photoresist mask, which was lifted off post-ruthenium deposition. Due to differences in surface energy, copper was expected to diffuse through ruthenium, leading to pristine, oxide-free copper available for bonding at the surface.

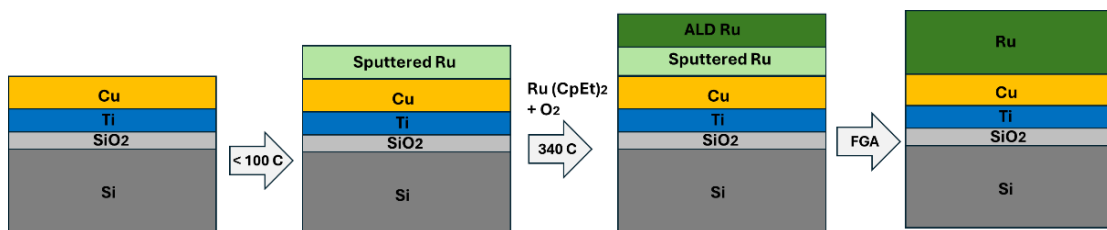


Figure 31 Deposition process of ruthenium on copper

To evaluate the diffusion of copper through ruthenium, even before bonding, our colleagues at UCSD performed an XPS analysis on a bare ruthenium-on-copper sample first. While negligible copper was detected at temperatures below 300°C, after a 30-minute anneal at 350°C with formic acid, only 34 % copper was detected, as shown in Figure 32. The sample was further annealed at 450°C to observe the grain growth of copper on Ru, and while the copper on the surface was more evident, pinholes were observed.

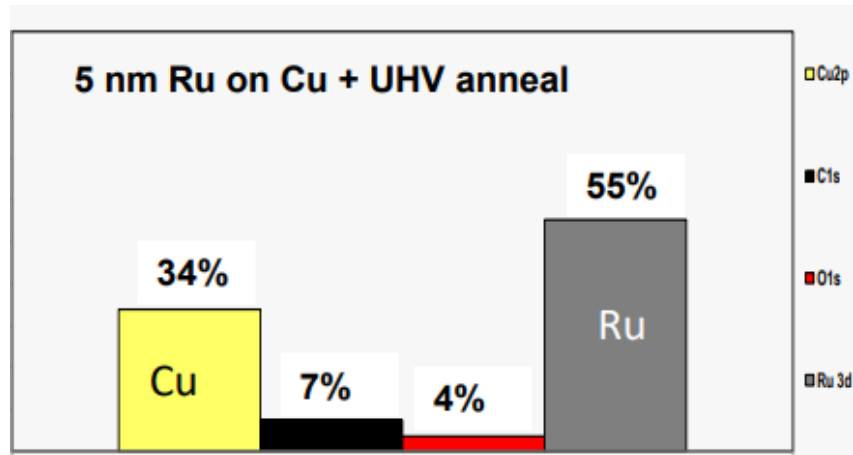


Figure 32 Percentage of Cu and Ru on the surface after 30 min FGA (Forming Gas Annealing) anneal [98]

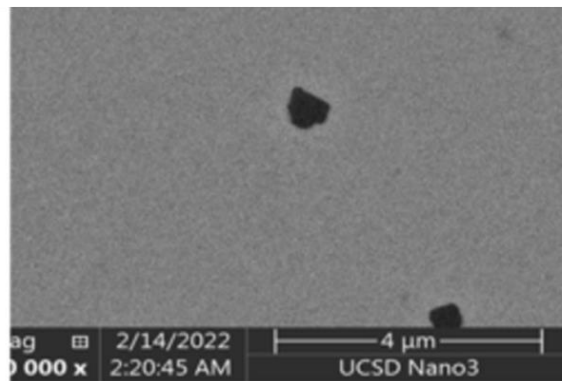


Figure 33 Diffused Cu on Ru surface with pinholes after 450°C FGA anneal

Due to the 350°C temperature observed to facilitate Cu diffusion through ruthenium, which was higher than 300°C (temperature for existing thermocompression bonding technology), it was concluded that the idea of ruthenium as a capping layer, while potent, needs more work to conceptualize. A probable cause for the insufficient diffusion of copper through the ruthenium surface can be the high carbon content later found on the evaporated copper in the test samples. A cyclic clean method was hence developed by our colleagues at UCSD and has been found effective in reducing the carbon content on evaporated copper at Georgia Tech. We plan to utilize the same as part of future work.

CHAPTER 4: STUDY OF ALD DEPOSITION PARAMETERS FOR ENHANCED INFILL COVERAGE

Objective

After Cu-Cu bonding, the second step in IHB is the infill. Infill is an integral part of a package that provides mechanical support, helps dissipate heat, and has electrical implications. The present works explore two different infill materials – Aluminum oxide and ZIF-8 Metal Organic Framework (MOF). The infill deposition process was thoroughly studied for coverage, uniformity, etc., with continuous optimization of the ALD/CVD deposition. The objective of the work concerning infill is to (a) achieve full coverage of the effective bonded area (the overlapping area between the top and bottom dice where interconnects exist), (b) achieve full infill for the die-to-substrate vertical gap, (c) scale down the die-to-substrate gap and I/O pitch without compromising on infill uniformity and coverage.

4.1 ALD-based Aluminum Oxide as Infill for IHB

With ever-increasing power density in current electronic packages, thermal management is a major challenge. A thermally conductive underfill can aid in heat dissipation and prevent thermal hotspots, which cause performance degradation in the integrated devices. Hence, due to its high thermal conductivity of $30 \text{ Wm}^{-1}\text{K}^{-1}$ [99], the aluminum oxide was initially studied as a potential infill for IHB. A testbed with top and bottom dice with copper I/Os was fabricated using the fabrication steps illustrated in Figure 34. Along with copper I/Os, the top die also included $75 \mu\text{m}$ diameter through

holes with a pitch of 140 μm , which were used as a pathway for the ALD gases to flow through. The top view of the upper die, as shown in Figure 22.b, depicts the distribution of high density through holes throughout the die area except in the middle of the die.

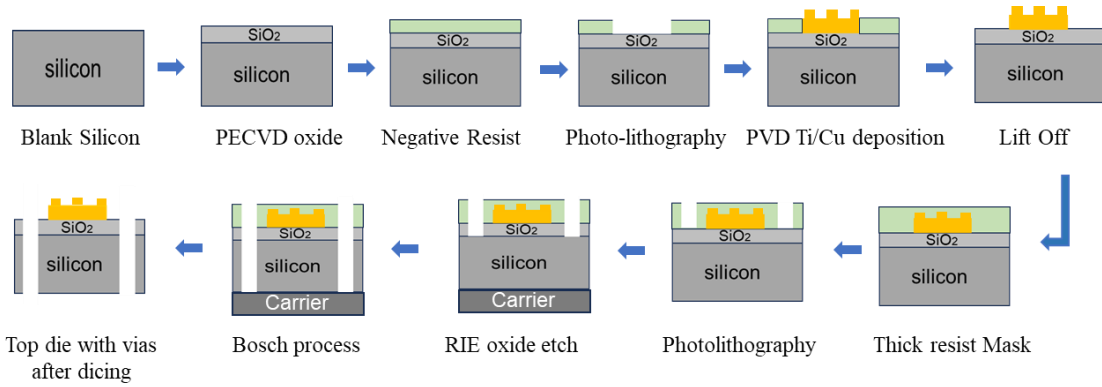
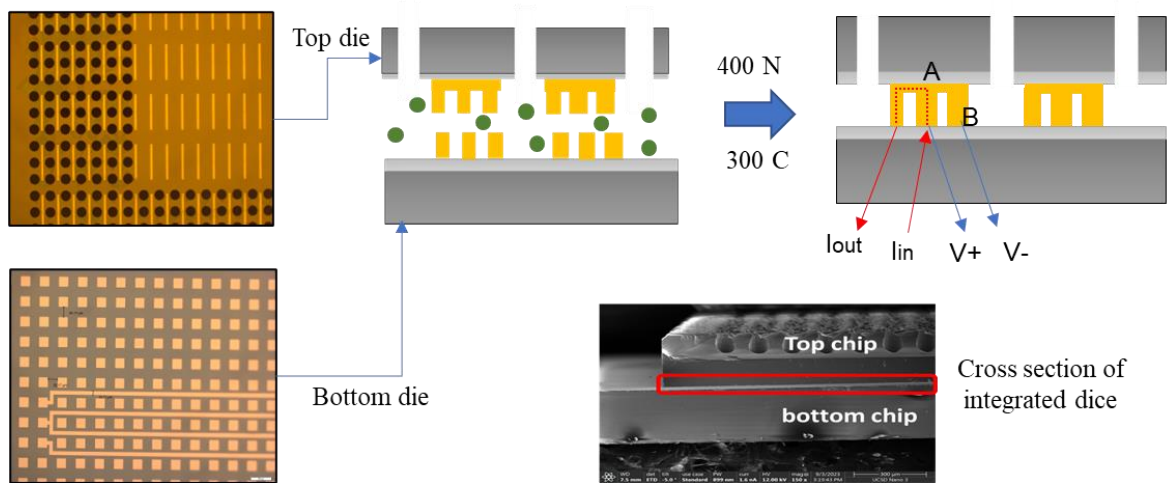
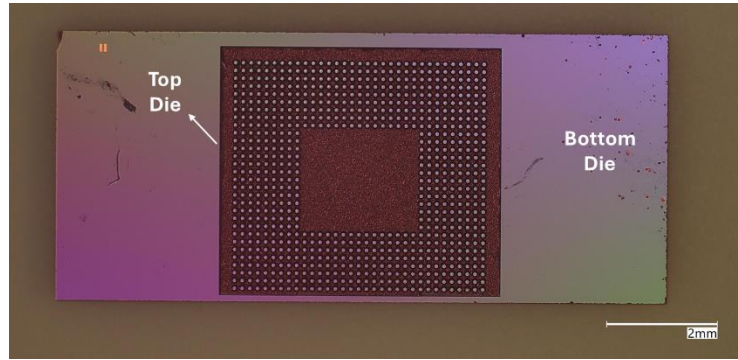


Figure 34 Fabrication flow for testbeds used for study of infill



(a)



(b)

Figure 35 (a) Schematic of cross-section of testbed used for infill analysis; (b) Top view of the integrated structure using Keyence 7000 Microscope depicting the through hole density on the top die

As detailed in the last chapter, the individual dice were integrated using flip-chip bonding. The integrated structure was then processed with 200 cycles of aluminum oxide ALD at first with a deposition rate of ~ 1 nm/cycle. Between the cycles, nitrogen purge was used to purge out excessive precursors. After 200 cycles ALD, the dice were de-bonded, and EDX was performed on the bottom die. As evident from the EDX mapping results shown in Figure 36.c, aluminum and oxygen were detected in areas beneath the through holes, with relatively no deposition on the areas of the bottom dice away from the through holes.

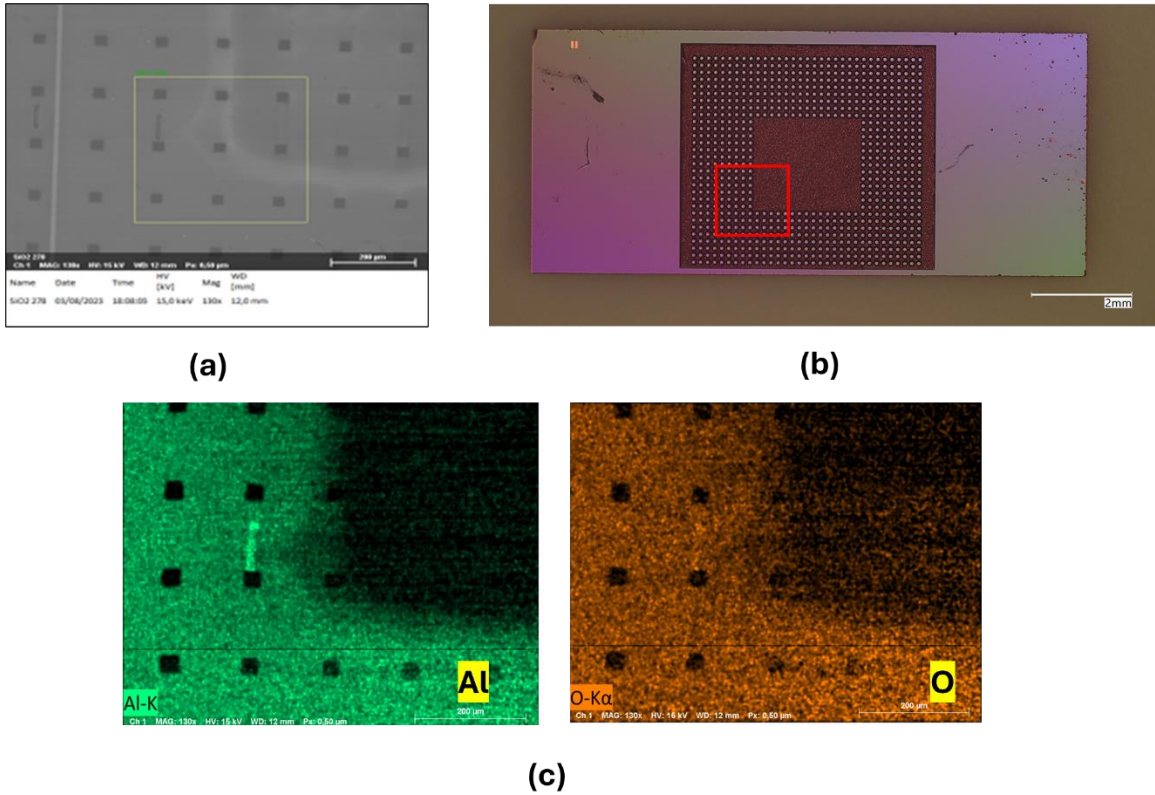


Figure 36 (a) SEM image of the area scanned for EDX mapping post 200 cycle aluminum oxide ALD; (b) Top view of the integrated structure highlighting the area under study for infill coverage; (c) EDX mapping results depicting the areas where aluminum and oxygen peaks were observed and hence the coverage of aluminum oxide

Further, the infill's thickness when analyzed with a forced ion beam (FIB) cross-section as shown in Figure 37, was observed to be 161.9 and well in accordance with the deposition rate of ~ 1 nm/cycle.

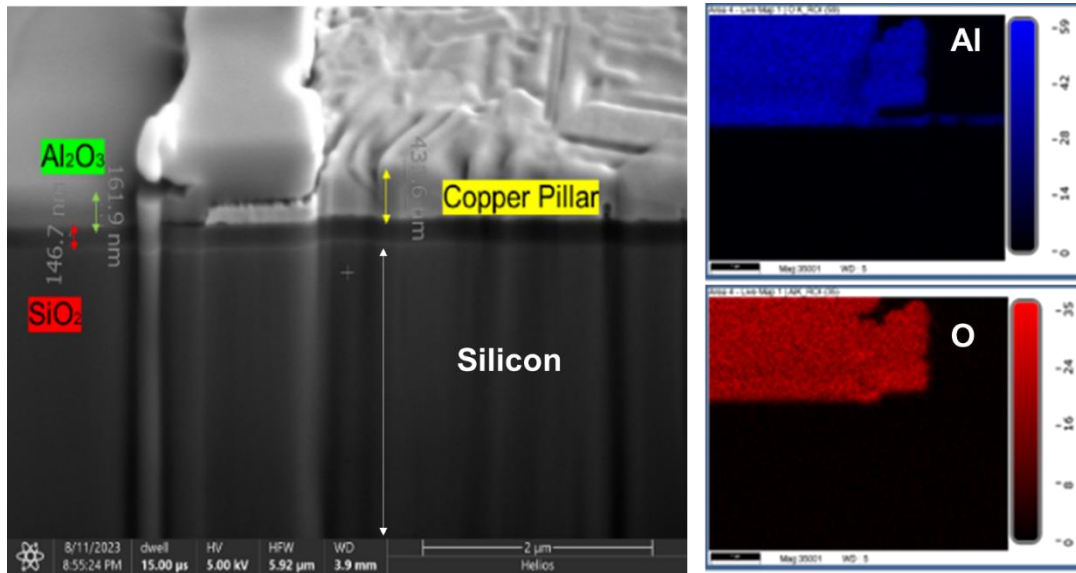


Figure 37 FIB cross-section of the de-bonded bottom die after 200 cycles of aluminum oxide ALD depicting

Next, 200 cycles of Al_2O_3 were repeated, but this time with a vacuum purge instead of a nitrogen purge. Post-ALD, EDX characterization of the three different areas of disassembled dice, as depicted in Figure 38, was performed, and similar Al/Si ratios were observed in all three areas, suggesting the presence of aluminum oxide in places away from the through holes as well.

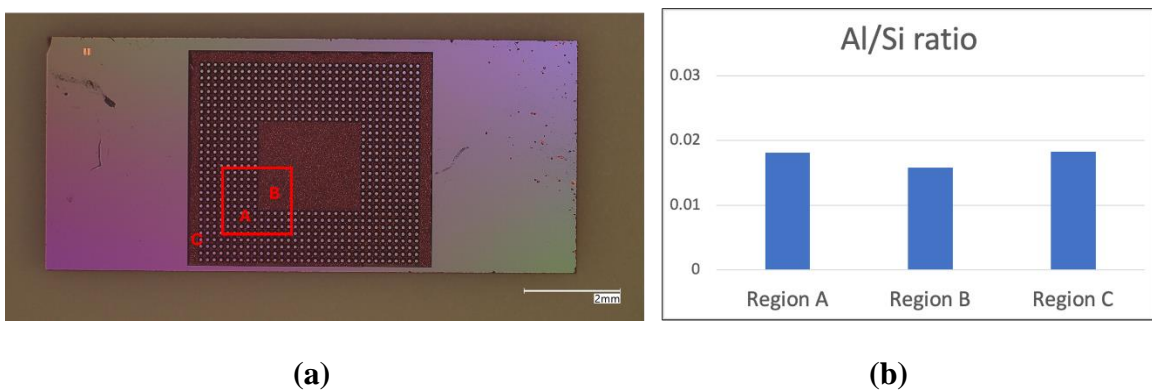
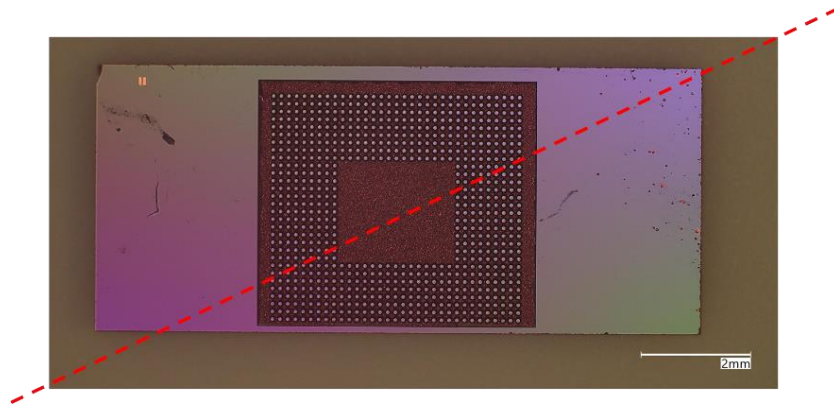


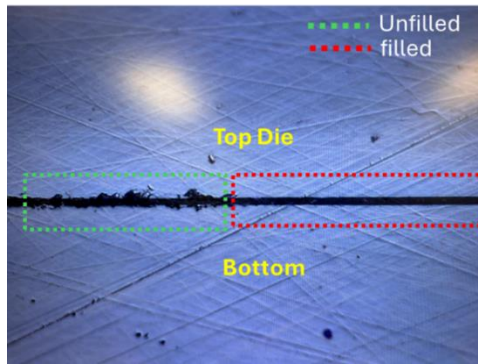
Figure 38 (a) Different regions of EDX study after 200 cycles ALD with vacuum

purge – areas of bottom die right beneath through holes (Region A), areas 300 μm from the nearest through hole (Region B), Areas near the edge (Region C); (b) Al/Si ratio in the three regions

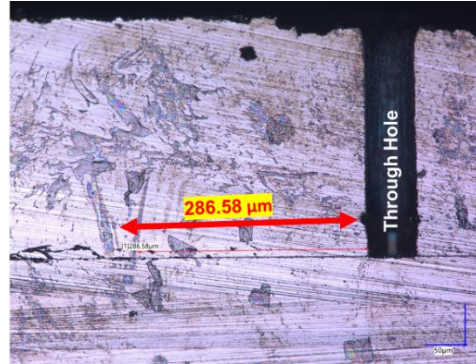
An increased coverage (up to 300 μm from the nearest through hole) was hence achieved using vacuum purge against nitrogen purge. The testbed was then further processed for complete infill of the 1 μm die-substrate gap with the help of 6000 ALD cycles of aluminum oxide and was characterized for the infill. Figure 39 depicts the cross-section of a filled 1 μm die-substrate gap with coverage up to 300 μm away from the nearest via. The cross-section shown in was obtained by polishing a diced integrated and infilled structure as shown in Figure 39.a. It was observed that – (a) areas beyond 300 μm from the nearest via were comparatively damaged and had no infill, (b) areas within 300 μm from the nearest via had infill and were smooth and undamaged even after polishing which highlights the presence of brittle aluminum oxide within 300 μm of the nearest through hole. The high rating of aluminum oxide, i.e., nine on the Mohs hardness scale [100], supports aluminum oxide being unaffected through dicing, as observed in the experiment. It was realized that the hardness of the aluminum oxide can, hence, potentially generate more stress in the package rather than stress distribution, thus negating its usefulness as an infill. Furthermore, the 6000 cycles of ALD to fill a 1 μm die-to-substrate gap took days to process, rendering it unsuitable for high-volume manufacturing setup.



(a)



(b)



(c)

Figure 39 (a) Diagonal Dicing of the integrated structure; (b) Close view of the cross-section of the integrated structure after depicting smooth infilled area and damaged area with no infill; (c) non-magnified image of cross-section depicting infill up to 300 μm from nearest, border through hole

Through the experiment, IHB was conceptually demonstrated, albeit with a need to – (1) enhance the infill’s coverage area, (2) reduce the through-hole density, (3) reduce the infill deposition time. The following section describes the study of infill spread concerning process parameters such as pulse time and temperature for enhanced

coverage. Also, based on the improved coverage achieved with vacuum purge against nitrogen purge, it was imperative to use vacuum purge for the ALD process going forward.

4.2 Metal Organic Framework for Infill using ALD-CVD Deposition

Metal organic frameworks have been explored in this work as infill material due to their exceptional electrical and mechanical properties. In addition, a 1 μm die-to-substrate infill using ALD/CVD-based MOF deposition takes a few hours compared to days for a conventional ALD process.

While MOF as infill exhibits benefits as delineated above, achieving full and uniform coverage over the effective bonded area was a significant challenge that has been experimentally addressed in this section. The course of action pursued to study and enhance coverage was – (a) to evaluate the dependency of coverage against ALD deposition parameters such as pulse time, temperature, etc., (b) using the experimental findings to optimize the deposition process to achieve infill while reducing through hole density, die-to-substrate gap, and I/O pitch.

4.2.1 MOF Deposition Process

MOF deposition is a two-step process wherein step one, a 20 nm thick zinc oxide (ZnO) is deposited as a precursor using ALD. The ZnO film is then converted to MOF in a

sealed isothermal vessel filled with 2-methylimidazole powder by exposing it to a vapor phase linker above 120 °C, as shown in Figure 40.b.

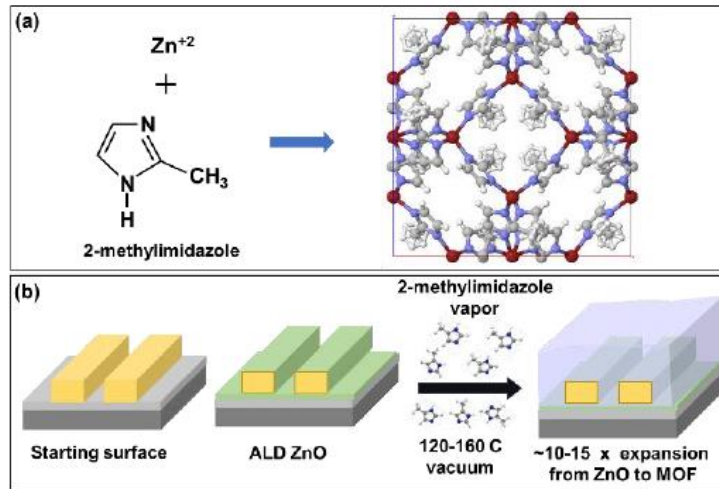


Figure 40 (a) ZIF-8 MOF chemistry and structure; (b) Schematic represents ALD ZnO and conversion to vapor-deposited MOF with around 10-15x volume expansion and gap fill.

With a 70% open fraction and 10-15 folds expansion in film thickness, MOF enables enhanced coverage with shorter deposition time compared to conventional ALD.

Table 3 depicts the properties of ZiF-8 MOF used in the current work.

Table 3 - Properties of ZiF-8 Metal Organic Framework [101] [102]

Dielectric Constant (100 kHz)	2.33
Thermal conductivity (W/m K)	0.33
Leakage Current (A cm²)	10 ⁸
Elastic Modulus (GPa)	>3
Breakdown Voltage (MV cm⁻¹)	2

4.2.2. Coverage vs ALD deposition parameters

A testbed with a top die consisting of high-density through holes everywhere except in the middle, the same as that used for aluminum oxide infill and shown in Figure 35.b, was used for MOF coverage analysis. Five cycles of MOF deposition (ZnO ALD + CVD conversion to MOF) were used to fill the 1 μm die-to-substrate gap fully. Post-infill, the integrated structure was diced diagonally from the middle and polished to obtain a smooth cross-section. Similar to aluminum oxide infill, the MOF coverage was observed to be $\sim 300 \mu\text{m}$ from the nearest via, which was also confirmed through the EDX mapping of the region 300 μm from the nearest via depicting the presence of Zn, O, and N (MOF constituents), as shown in figure 41.b.

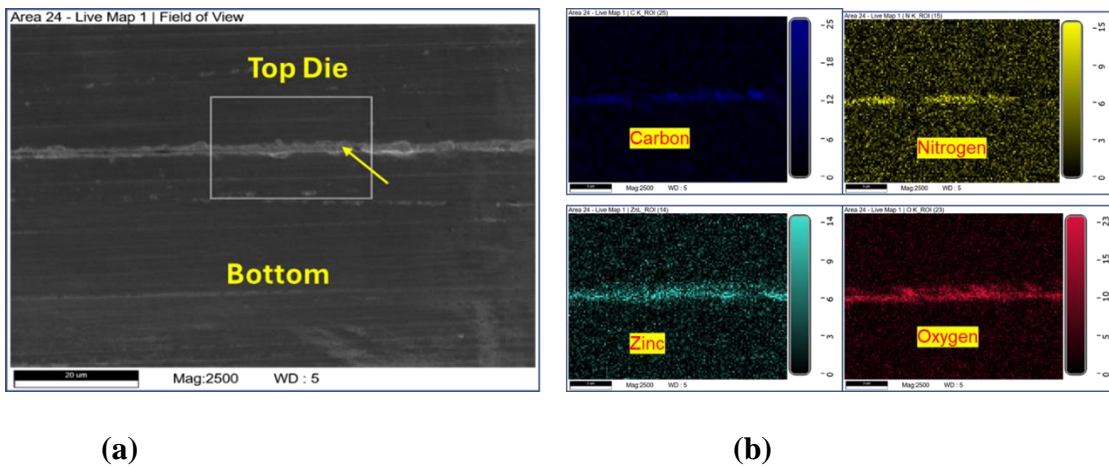


Figure 41 EDX mapping of the cross-section of integrated structure 300 μm from nearest through hole after full infill of die-to-substrate gap

Further, to ensure that the infill observed was MOF and not the epoxy used for polishing the diced structure, the through holes in the diced/polished structure were analyzed for the presence of an epoxy mold, as shown in Figure 42.

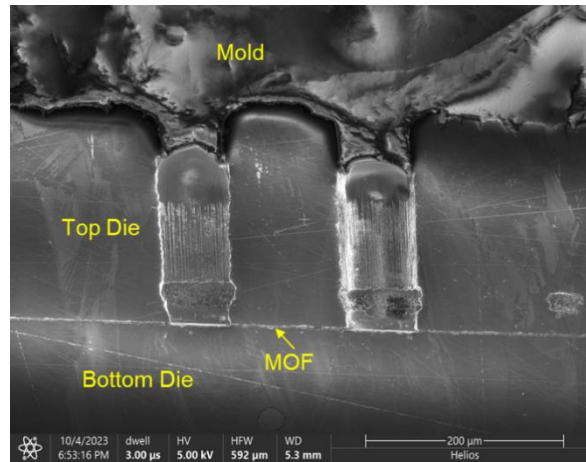


Figure 42 SEM image of cross-section depicting non-penetration of polishing mold into the vias and die-to-substrate gap rendering the results discussed above credible

The coverage of $\sim 300 \mu\text{m}$ from the nearest through hole for both aluminum oxide and MOF suggested that the infill's area coverage depends on deposition parameters rather than infill material. Hence, the next experiment focused on studying deposition parameters such as pulse time, temperature, etc., to achieve enhanced coverage.

A different testbed, as shown in Figure 43, with fewer through holes, was fabricated and assembled. Assuming that a uniform coverage of ZnO throughout the effective bonded area would enable the same for MOF, the testbed was studied for the coverage of ZnO.

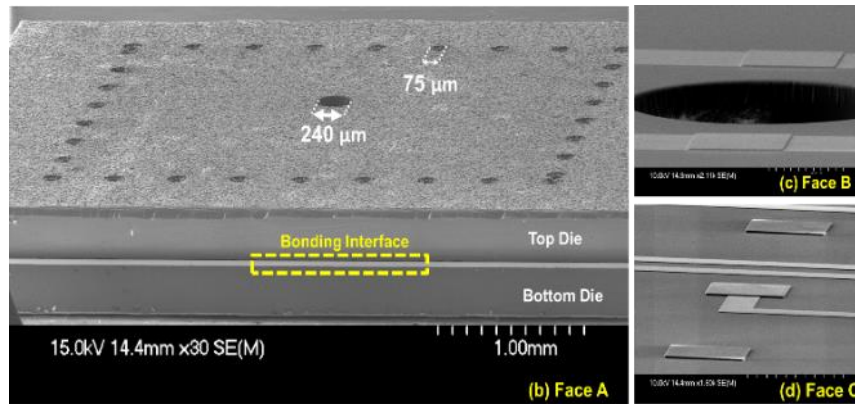
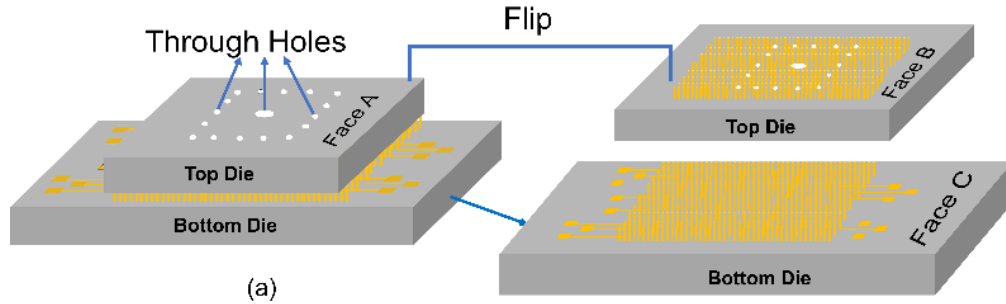


Figure 43 (a) Schematic of the testbed with a top die having through holes Cu-Cu bonded to a bottom substrate; (b) SEM cross-section of the Cu-Cu bonded test structure (Face A); (c) copper pad/pillar on the surface of the top die before bonding (Face B); (d) copper pillar with metal traces on the bottom die's surface before bonding (Face C)

While both the bottom and top die consisted of 140 μm pitch copper I/Os, the top die also consisted of 165 μm pitch and 75 μm diameters through holes. In addition, there was a 240 μm diameter through hole in the center of the top die which was included to see the feasibility of using one large through hole instead of multiple small ones. A 20 nm ZnO ALD film was deposited using the through-holes on the test structure with a pulse time of 250 ms and deposition temperature of 120°C. Post 20

nm infill, the integrated structure was de-assembled using a bond shear tool and the bottom die was then analyzed for the coverage of ZnO through SEM/EDX.

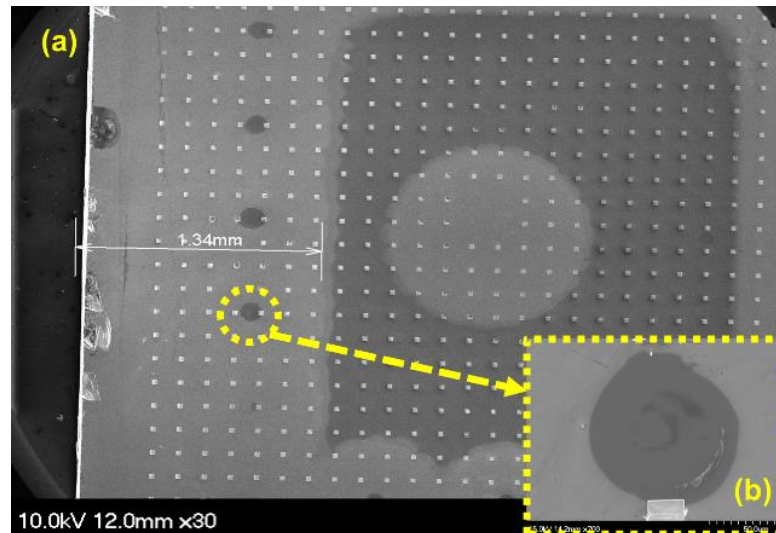
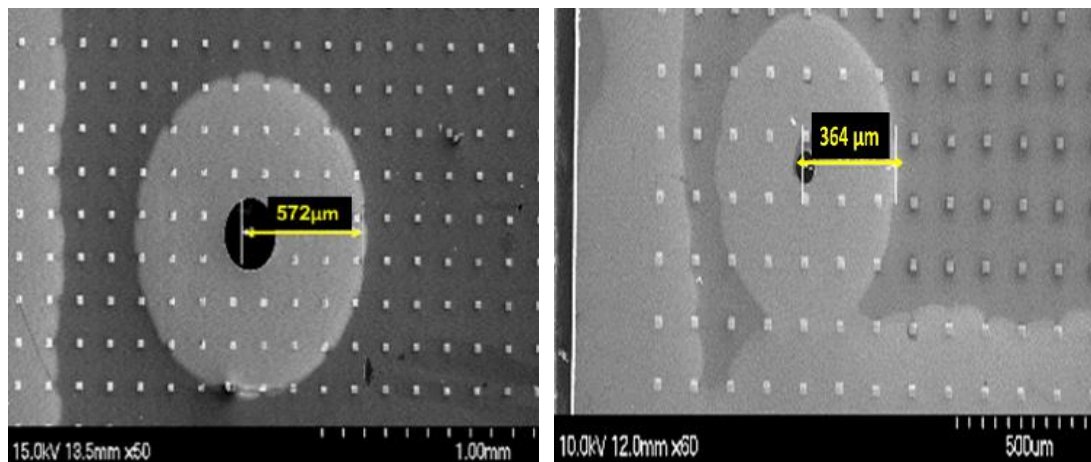


Figure 44 De-bonded bottom die overview after 20 nm ZnO ALD; (b) Regions of undeposited infill beneath the through holes

Incomplete coverage of ZnO film was achieved over the effective bonded area accompanied with undeposited ZnO on areas of the bottom die right beneath the through hole as shown in Figure 44, which was counterintuitive. Along with low deposition temperature inhibiting effective reactions in areas right beneath the through holes, another probable cause for the observation could be the pressure gradient for the incoming ALD gases as they enter through a 75 μm diameter through-hole and seep into a 1 μm die-substrate gap. Hence, to mitigate the pressure gradient effect and provide enough reaction energy, keeping a constant pulse time, the deposition temperature was increased from 120°C to 150 °C, which helped in infill deposition even beneath the through holes, as shown in Figure 46.a. Figures 45.a and Figure 45.b

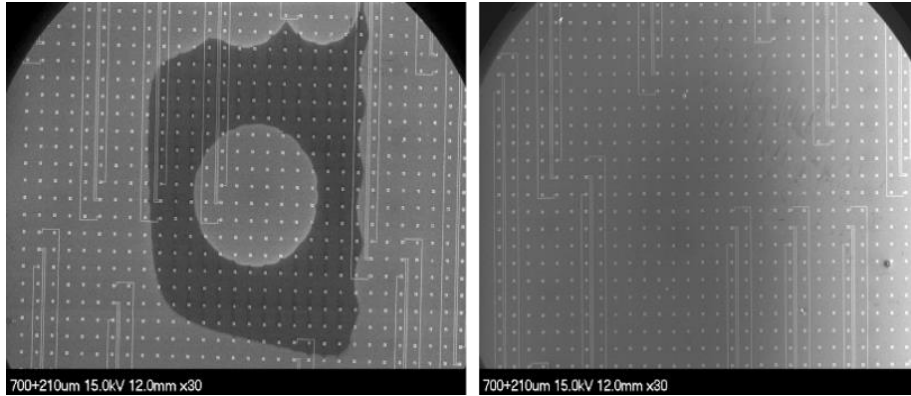
show the through hole on the top die above the bottom die depicted in Figure 44. A spread ratio (radius of ZnO spread: hole radius) of $\sim 1:10$ and $\sim 1:4$ was observed for the smaller vias ($75\ \mu\text{m}$ diameter) on the sides and the larger ($240\ \mu\text{m}$ diameter) via in the middle respectively. The spread radius of $364\ \mu\text{m}$ for the smaller through hole was consistent with that observed for the aluminum oxide ALD in the last section. The spread radius as observed was noted, and keeping the temperature same at $150\ ^\circ\text{C}$, the pulse time was now increased from $250\ \text{ms}$ to $1\ \text{s}$, and complete coverage of the infill over the effective bonded area was observed, as shown in Figure 46.b.



(a)

(b)

Figure 45 Area on the top die near through hole depicting spread radius of (a) $572\ \mu\text{m}$ for $240\ \mu\text{m}$ diameter via; (b) $364\ \mu\text{m}$ for $75\ \mu\text{m}$ diameter via



(a)

(b)

Figure 46 De-bonded bottom die overview after 20 nm ZnO ALD with a) pulse time 250 ms & temperature 150°C; (b) pulse time 1 sec & temperature 150°C

Hence, full coverage of 5 mm x 5 mm effective bonded area was achieved through 150°C temperature and a pulse time of 1 sec for ALD parameters. Through the experiment, we were able to (a) assess the right infill material for inverse hybrid bonding, (b) study the effect of ALD deposition parameters on coverage area, and (c) reduce the through hole density via ALD process optimizations. The experiments prove to be pivotal in our understanding of the ALD/CVD process required to achieve enhanced coverage, and the learnings will be used to demonstrate IHB in the next chapter while eliminating through holes for the penetration of gases.

CHAPTER 5: Inverse Hybrid Bonding

5.1 First Demonstration of Inverse Hybrid Bonding

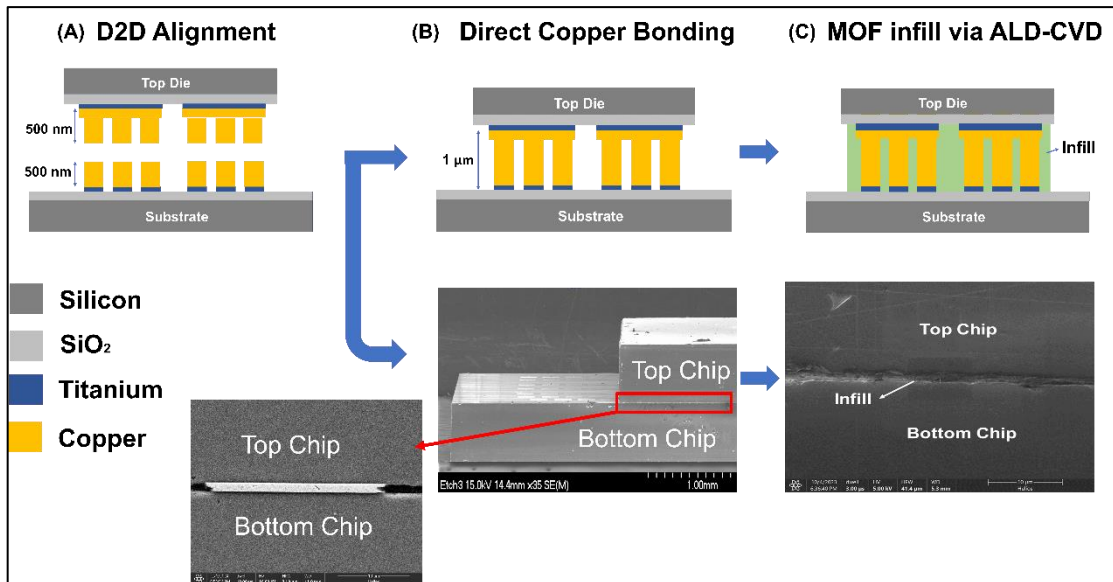
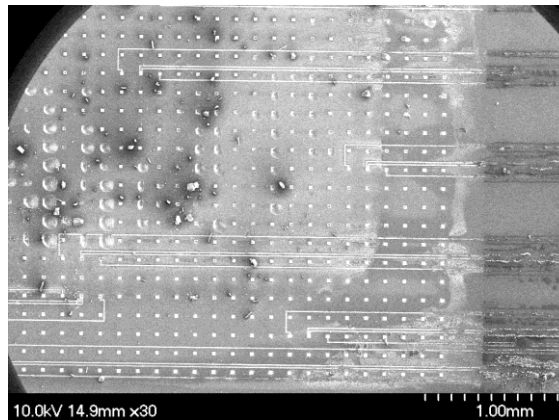


Figure 47 Process sequence for Inverse Hybrid Bonding

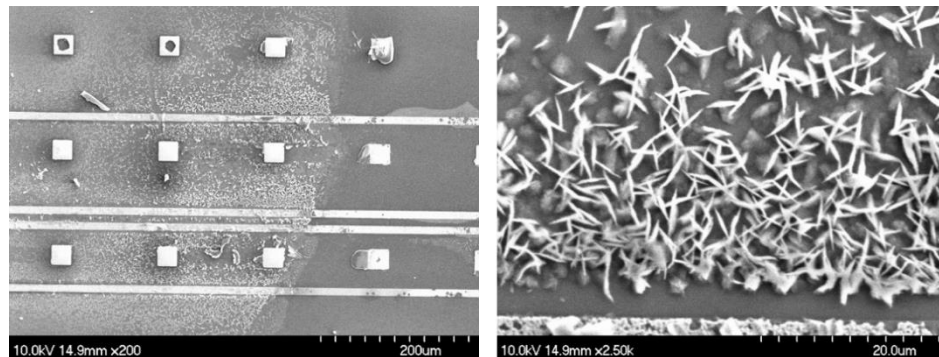
As mentioned earlier, Inverse Hybrid Bonding (IHB) is a two-step process. In step 1, we use direct copper bonding for off-chip I/O connections, followed by an ALD-CVD deposition for the MOF infill. We successfully demonstrated IHB without any through holes on the top die, using process parameters such as pulse time and deposition temperature derived from our experimental findings.

Thermocompression Cu-Cu bonding with a force of 400 N and 300°C was used for integrating a 5 mm x 5 mm top die over an 11 mm x 5 mm bottom die with copper I/Os on a 140 μ m pitch, as shown in step B of Figure 47. Post bonding, the integrated

structure was processed with a single cycle of MOF deposition using the 1 sec pulse time and 150°C temperature (as learned in the last chapter) and analyzed for coverage. It was observed that the pulse time and temperature used were sufficient to propel full penetration of MOF across the 5 mm x 5 mm effective bonded area, as shown in Figure 48.a. The presence of MOF in the middle of the bonded structure was visually confirmed by observing grains, as shown in Figure 48.b.



(a)



(b)

Figure 48 (a) Overview of the de-bonded bottom die after 1 cycle of MOF; (b) MOF grains observed right in the middle of bottom die depicting complete penetration

Four more cycles of MOF deposition were done to ensure full infill of the 1 μm die-to-substrate gap. To confirm the full penetration of MOF throughout the effective bonded area and 1 μm die-to-substrate vertical gap, the integrated, infilled structure was sheared, and a FIB cross-section right in the middle of the sheared bottom die was made to analyze the thickness of MOF, as shown in Figure 49.

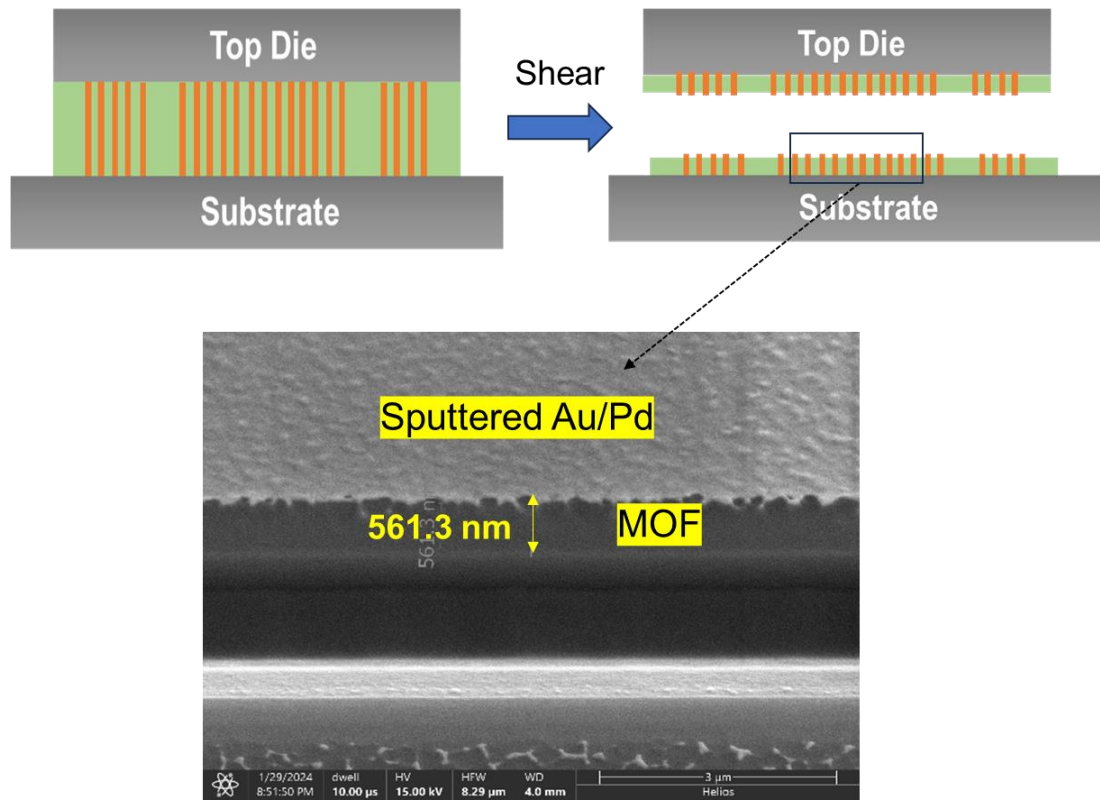
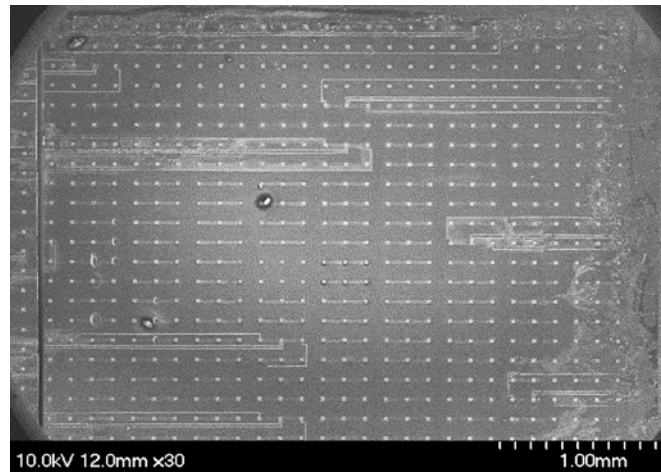


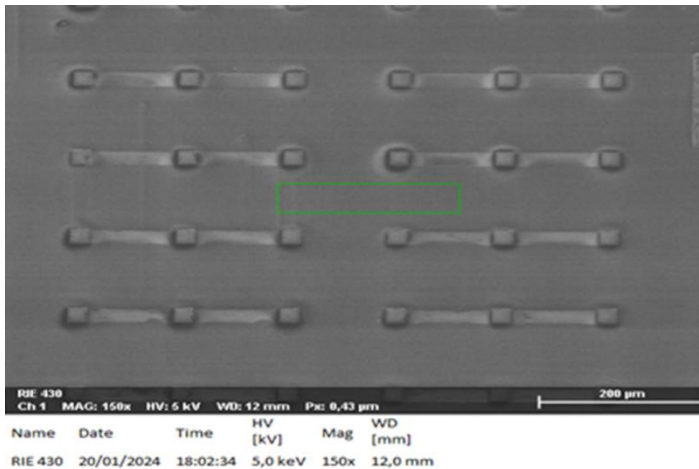
Figure 49 FIB cross-section near the center of the bottom substrate depicting 500 nm MOF as expected

The ~ 500 nm thickness of MOF, as observed on the FIB cross-section, indicated the full infill of the 1 μm die-to-substrate gap. Further, Figure 50.a depicts the overview of the de-bonded bottom die wherein impressions of the copper pads from the top die

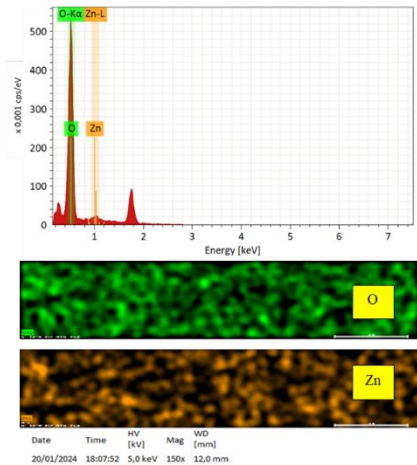
were observed throughout the bottom die area, suggesting that the MOF did infiltrate the substrate center. Full penetration and reach of the MOF infill even to the middle of the substrate was also confirmed using energy dispersive x-ray (EDX), as shown in Figure 50.c, which depicted Zn and O peaks.



(a)



(b)



(c)

Figure 50 (a) Overview of the de-bonded bottom die after 5 cycles of MOF infill; (b) Analysis area at the center of the bottom substrate after de-bonding; (c) Spectrum and EDX mapping depicting Zn and O peaks and hence the presence of MOF in the substrate’s center

However, a higher Zn atomic percentage on edges was detected, which accounts for the loading effect. The loading effect in this respect refers to more deposition or accumulation of ZnO at the edges due to the narrow die-to-substrate gap leading to unconverted ZnO, which is detrimental to the electrical properties of infill. The ZnO is a semiconductor and unless converted fully into MOF, it would be difficult to term the infill as “low-k”. As part of future work, efforts will be made to mitigate this effect through optimization of process parameters such as pulse time, flow rate etc.

Hence, with the full coverage of 5 mm x 5 mm effective bonded area and 1 μm die-to-substrate gap confirmed through the FIB cross section, IHB as a concept was proved. Further, peak shear value of the integrated structure was measured before and after MOF infill to see the impact on package strength, as depicted in figure 51.b. Post infill, a fourfold increase in the shear strength observed is a testament to the mechanical properties of MOF and its potential use as an infill for electronic packages.

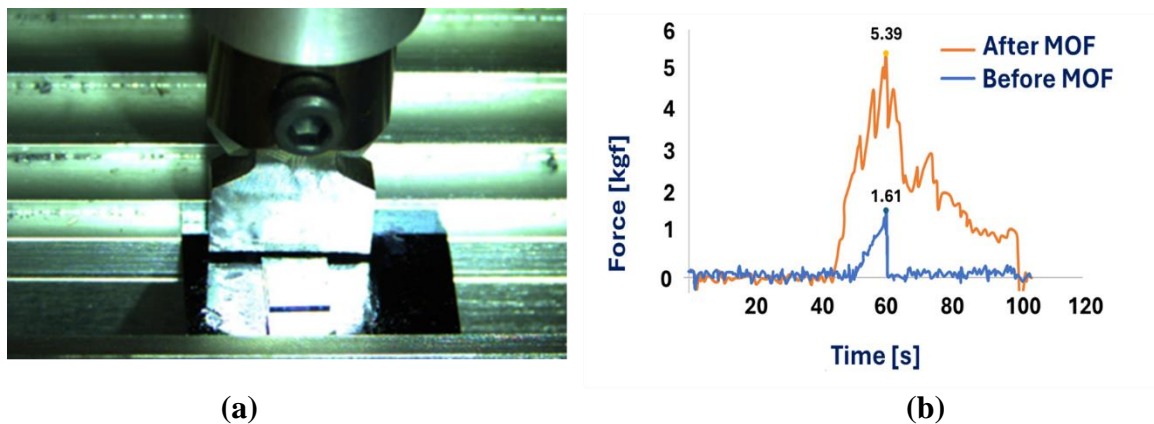
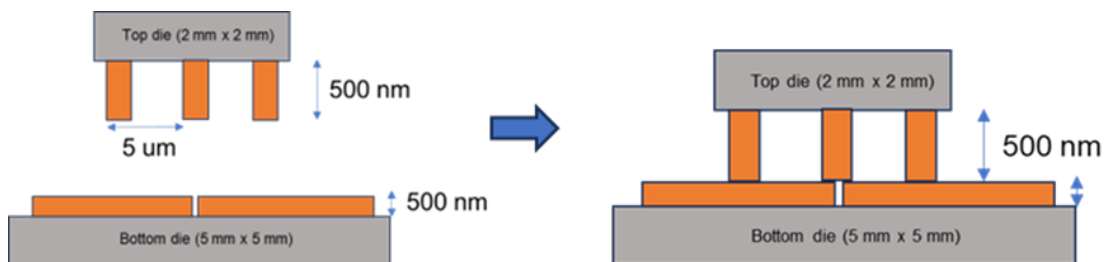


Figure 51 (a) Bond shear using Xyztec’s shear tester; (b) Comparison of Peak shear

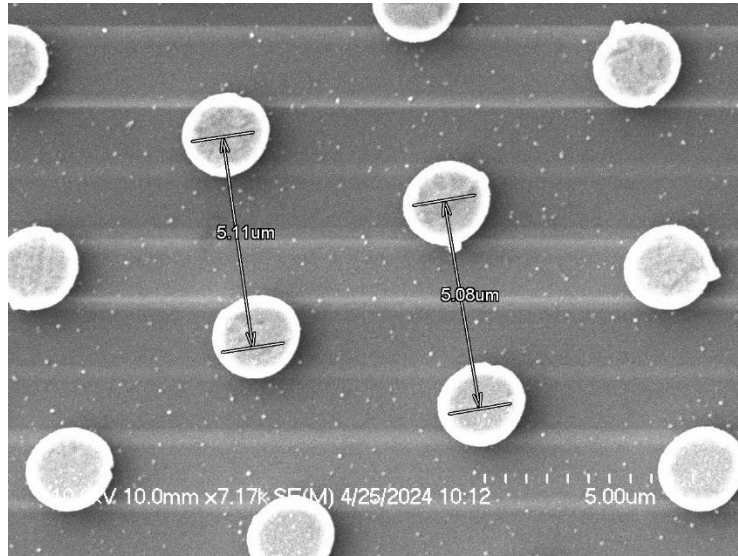
force for the bonded dice before and after MOF infill

5.2 Inverse Hybrid Bonding with Fine Pitch Dimensions – 5 μm Pitch

With pitch scaling, the inter-I/O gaps become narrower, making the penetration of gases difficult. While we anticipated coverage and uniformity as a challenge, the learnings from previous experiments equip us with the understanding of process parameters that need to be twitched to enhance coverage. A testbed, as shown in Figure 52.a, was hence fabricated to demonstrate infill for an integrated structure with a top die having 5 μm pitch copper pillars (shown in Figure 52.b) and a bottom die consisting of 100 μm x 100 μm , 500 nm thick copper pads. Pillar-to-pad bonding was chosen instead of pillar-pillar as previously to (a) simplify the testbed as pillar-pillar bonding would be limited by the alignment accuracy of the flip chip bonder tool and (b) create a reduced die-to-substrate gap of 500 nm against 1 μm previously. The reduced die-to-substrate gap and pitch reaffirm our aim to push toward sub-micron pitch scaling for D2D/D2W with I/O lengths comparable to existing hybrid bond technology while addressing challenges such as particle control at the same time.



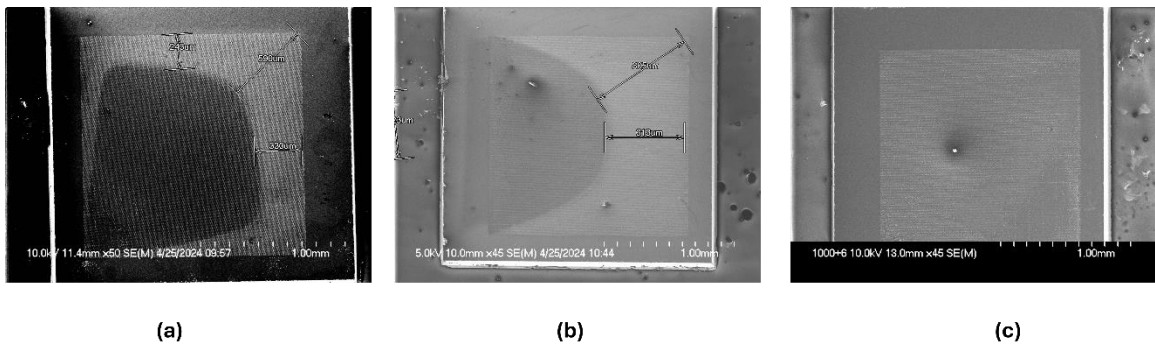
(a)



(b)

Figure 52 (a) Schematic of the cross-section of testbed used for infill analysis with 5 μm pitch I/Os; (b) SEM image of top die depicting 5 μm pitch copper I/Os

Similar to the infill coverage analysis performed previously, the new testbed was processed with 20 nm ZnO ALD at first with a pulse time of 250 ms, 1 sec, and 8 sec at a temperature of 160°C, and the spread was observed as shown in Figure 53.



(a)

(b)

(c)

Figure 53 Overview of the top die after 20 nm ZnO deposition with a temperature of 160°C and pulse time of (a) 250 ms (b) 1 sec (c) 8 sec

It was observed that with the pulse time increased from 250 ms to 8 sec, ZnO could cover the effective bonded area of 2 mm x 2 mm. The coverage was further confirmed by characterizing the top die de-bonded sample with an 8-second pulse time for the atomic percentage of Zn, Si, and O at multiple points via EDX, which depicted a similar Si-Zn ratio at all the tested points.

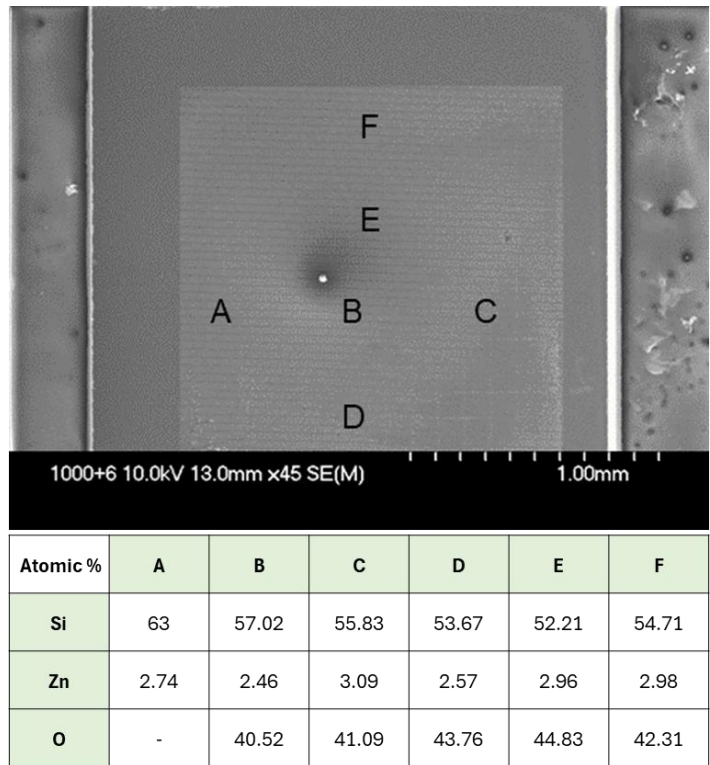


Figure 54 Atomic percentage of Zn,O and Si at different regions of the de-bonded top die with 8 sec pulse time to analyze for coverage uniformity

With the required pulse time of 8 sec for full area coverage, 3 cycles of ALD-CVD MOF deposition were performed on the testbed to analyze the die-to-substrate gap infill coverage with MOF. The infilled sample was again sheared, and various areas of the de-bonded top die at the center and edges were analyzed through EDX for an atomic

percentage of Zn, as shown in Figure 55. The low Zn percentage at the edges suggested the conversion of ZnO into MOF and, hence, its presence of the latter, while the higher Zn atomic percentage at the center suggested the presence of unconverted ZnO.

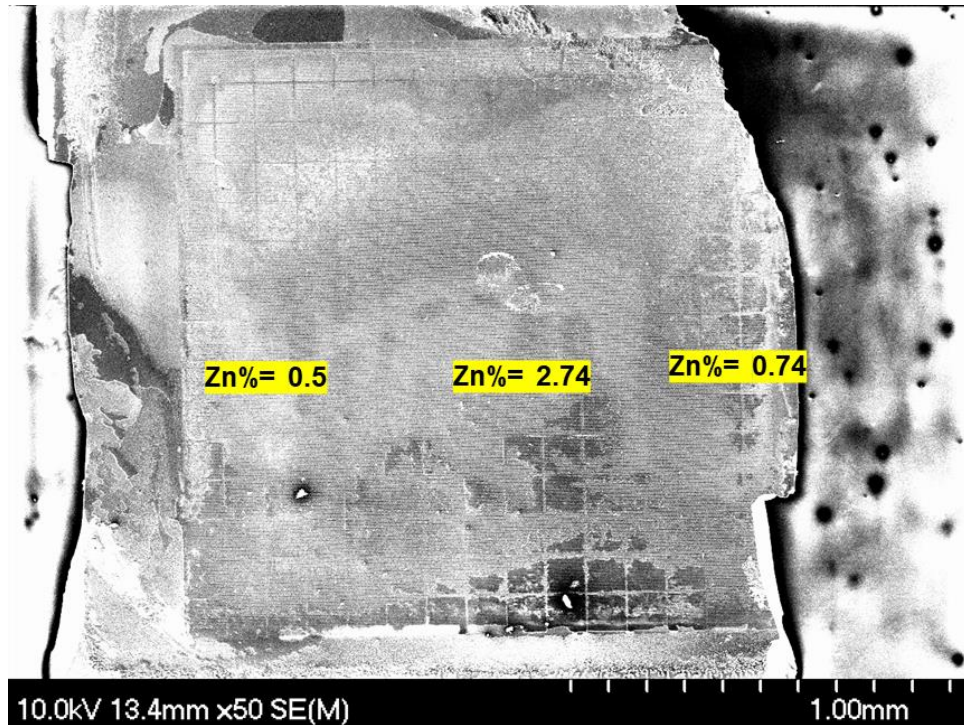


Figure 55 Atomic percentage of Zn at different regions of the de-bonded top die after 3 cycles of ALD/CVD MOF deposition

However, to confirm the inference, a tilted SEM characterization of both the de-bonded top and bottom die after 3 cycles of ALD/CVD-based MOF processing was performed. Figure 56 shows the SEM images of the top die at the center and edge regions, which confirm the presence of MOF at the edges, as can be seen near the copper pillars, while the same is absent in the middle region.

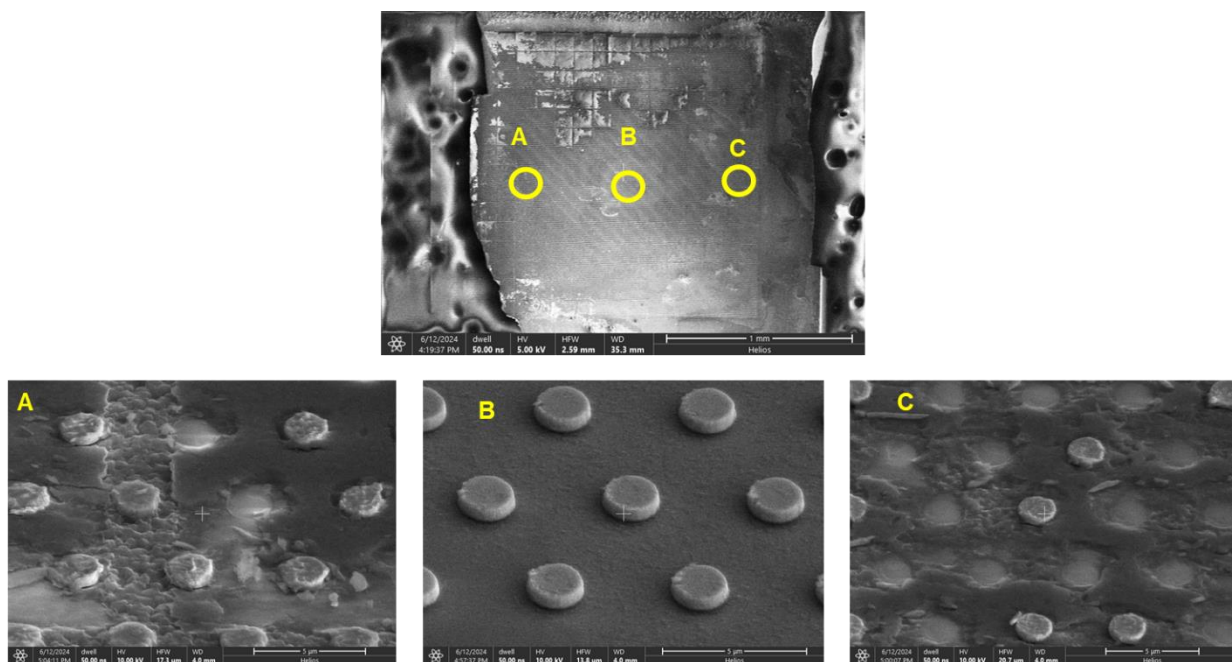


Figure 56 Tilted SEM images of the top die at edges (A and C) and at the center after 3 cycles of MOF infill

The possibility of MOF being transferred from the top to the bottom die during shearing was also considered; hence, a FIB cross-section was made on the de-bonded bottom die at the center region corresponding to the top die. As evident in Figure 57, the FIB cross section on the bottom die's copper pad depicted the top surface thickness of ~500 nm, which was essentially the copper pad itself. However, the surface morphology did suggest a partial presence of MOF at the center.

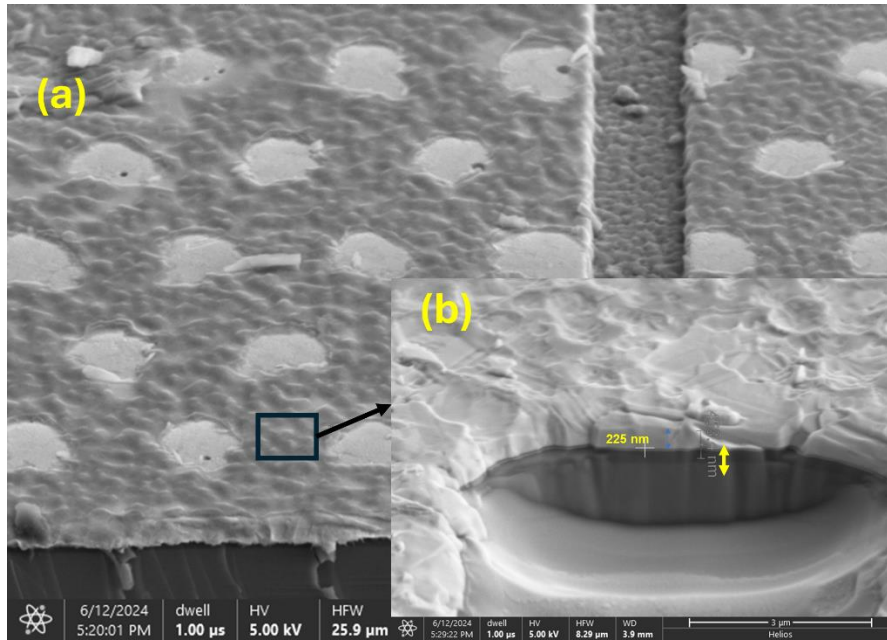


Figure 57 (a) The region where FIB cross section was made (b) FIB cross-section of a 100 μm x 100 μm copper pad on de-bonded bottom die after 3 cycles of MOF

Hence it was concluded that while ZnO could penetrate all through the effective bonded area of 2 mm x 2 mm with 5 μm pitch I/Os, the same was not true for MOF. The probable cause behind the observation can be the reduced die-top-substrate gap of 500 nm (against 1 μm as used previously) being inadequate for full penetration of MOF. Hence, while we did address the coverage challenge for our testbed with 5 μm fine pitched copper I/Os for ZnO using enhanced pulse time, we tend to pursue further optimization of ALD/CVD process parameters, such as pressure, pulse time of subsequent deposition cycles of ZnO, etc. to demonstrate inverse hybrid bonding with fine-pitch interconnects and ultra-low die-to-substrate gaps

CHAPTER 6: FUTURE WORK

6.1 Direct Copper Bonding

The current bonding temperature being used for inverse hybrid bonding is 300°C. Ruthenium as a capping layer over copper was explored to achieve low-temperature bonding. When a copper-on-ruthenium sample was annealed, it was observed that the copper atomic percentage on the surface after 350°C anneal was merely 34 %. The high carbon content on the evaporated copper samples was considered a probable cause, which may have been blocking the diffusion of copper through ruthenium. Hence, along with the formic acid pre-clean, which essentially removes oxides, an additional pre-clean step using hydrogen peroxide (developed at Dr. Andrew Kummel's lab, UCSD) is being planned for testing to reduce carbon on the copper samples. Being a strong oxidant, HOOH(g) oxidizes the carbides to CuO, which then is converted to Cu using hydrazine (N₂H₄). Hence, repeated oxidation-reduction cycles – “cyclic clean” could be employed to (a) affect low-temperature Cu-Cu bonding at low temperature without Ru capping, (b) allow reevaluation of Ru as capping layer by potentially enabling increased copper diffusion through ruthenium and the same needs to be tested.

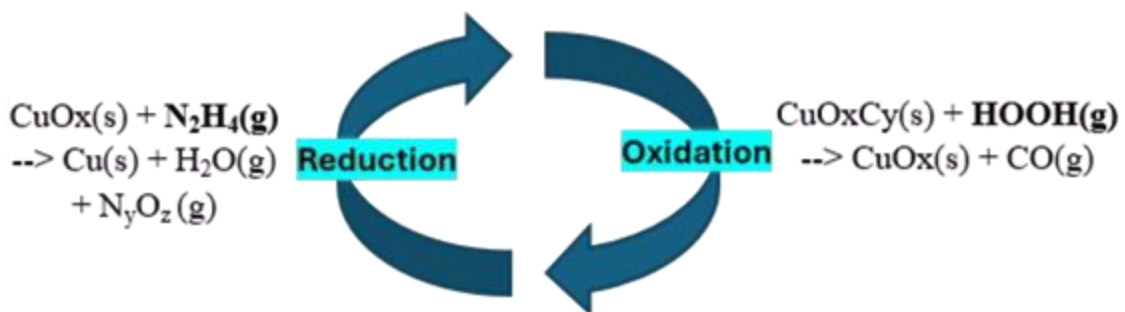


Figure 58 Cyclic clean process by using HOOH as oxidant to remove carbides over copper and followed by N_2H_4 as reductant to remove oxides.

6.2 Infill for Inverse Hybrid Bonding for Fine Pitch Heterogeneous Integration

The current results, as discussed in the last chapter, suggest incomplete infiltration of MOF over the 2 mm x 2 mm effective bonded area with a pitch of 5 μm for I/Os. While full coverage was achieved for ZnO (precursor for MOF), the presence of ZnO, but not MOF at the die's center, suggests that the 500 nm die-to-substrate gap may not be sufficient for efficient MOF penetration. In a nutshell, the incomplete conversion of ZnO to MOF near the middle areas of the bonded structure is a hurdle that needs to be overcome. Hence, the ALD-CVD deposition process must be further optimized for higher pulse time, lower deposition or purge pressure, etc.

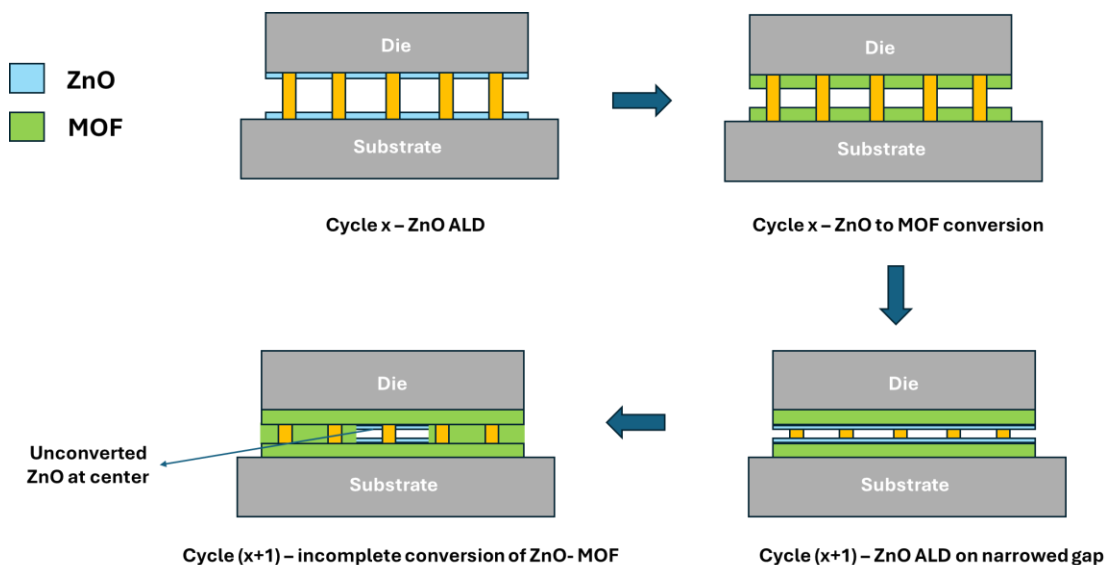
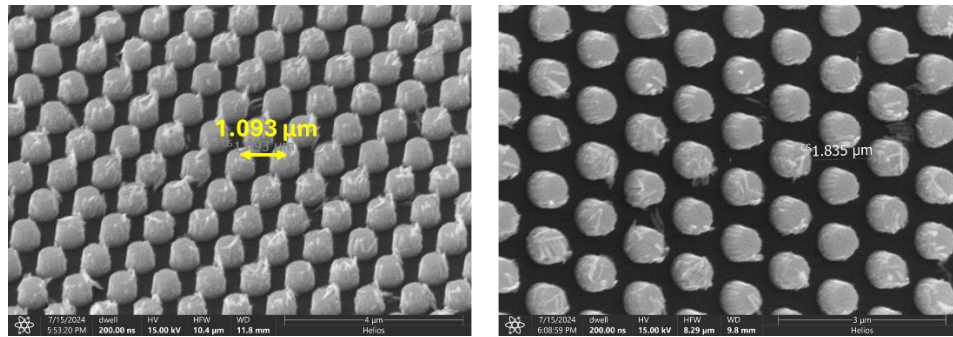


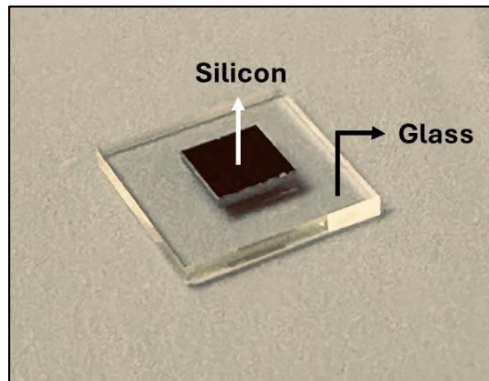
Figure 59 Unconverted ZnO at the center for 5 μm I/O pitch sample with 500 nm die-to-substrate gap

Further, efforts similar to the demonstrated fine-pitch silicon-silicon integration are required to extend the developed bonding scheme for silicon-glass assembly. Preliminary work on the silicon-glass integration includes a similar testbed as shown in Figure 52.a, but with the bottom substrate of glass. The top silicon die used in the testbed included copper pillars with a pitch of 1 μm , 2 μm , and 5 μm , assembled on a glass substrate with copper pads using thermocompression flip-chip bonding, as shown in Figure 60. However, the integrated testbed is yet to be processed with infill and the same must be undertaken going forward.



(a)

(b)



(c)

Figure 60 Top silicon die with (a) 1 μm ; (b) 2 μm Cu pillars; (c) Silicon-to-glass assembled testbed for fine pitch IHB

6.3 Modeling and Simulation for the IHB Integrated Structures

A system-level modeling for the bonded structures with infill is required to assess the impact of infill on the thermo-mechanical strength of the package as well as its electrical characteristics. Through thermo-mechanical simulations, one can determine the integrated structure's response to successive thermal cycles, stress, and fatigue; conversely, a detailed study of the low-k infill on electrical parameters such as insertion loss, cross talk, etc., is required.

References

- [1] A. Mallik *et al.*, "Economics of semiconductor scaling - a cost analysis for advanced technology node," in *2019 Symposium on VLSI Technology*, Kyoto, Japan: IEEE, Jun. 2019, pp. T202–T203. doi: 10.23919/VLSIT.2019.8776521.
- [2] G. E. Moore, "Progress in digital integrated electronics [Technical literature, Copyright 1975 IEEE. Reprinted, with permission. Technical Digest. International Electron Devices Meeting, IEEE, 1975, pp. 11-13.]," *IEEE Solid-State Circuits Soc. Newsl.*, vol. 11, no. 3, pp. 36–37, Sep. 2006, doi: 10.1109/N-SSC.2006.4804410.
- [3] D. O’Laughlin, "The Rising Tide of Semiconductor Cost." Accessed: Jun. 26, 2024. [Online]. Available: <https://www.fabricatedknowledge.com/p/the-rising-tide-of->

semiconductor

- [4] S. Hong, “ISSCC: Roadmap on 3D Interconnect Density,” EE Times Asia. Accessed: Jun. 26, 2024. [Online]. Available: <https://www.eetasia.com/isscc-roadmap-on-3d-interconnect-density/>
- [5] “From Complexity to Cost: Unveiling the Semiconductors’ Journey as the New Gold | LinkedIn.” Accessed: Jun. 26, 2024. [Online]. Available: <https://www.linkedin.com/pulse/semiconductor-new-gold-balancing-technology-cost-azhar-md-nayan/>
- [6] C. Mack, “The Multiple Lives of Moore’s Law,” *IEEE Spectr.*, vol. 52, no. 4, pp. 31–31, Apr. 2015, doi: 10.1109/MSPEC.2015.7065415.
- [7] G. H. Loh, S. Naffziger, and K. Lepak, “Understanding Chiplets Today to Anticipate Future Integration Opportunities and Limits,” in *2021 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, Grenoble, France: IEEE, Feb. 2021, pp. 142–145. doi: 10.23919/DATE51398.2021.9474021.
- [8] “TSMC Announces 2x Reticle CoWoS For Next-Gen 5nm HPC Applications – WikiChip Fuse.” Accessed: Jun. 26, 2024. [Online]. Available: https://fuse.wikichip.org/news/3377/tsmc-announces-2x-reticle-cowos-for-next-gen-5nm-hpc-applications/#google_vignette
- [9] S. Wong, A. El-Gamal, P. Griffin, Y. Nishi, F. Pease, and J. Plummer, “Monolithic

3D Integrated Circuits,” in *2007 International Symposium on VLSI Technology, Systems and Applications (VLSI-TSA)*, Hsinchu: IEEE, Apr. 2007, pp. 1–4. doi: 10.1109/VTSA.2007.378923.

[10] “🌱 E13: Chiplets - How Chip Lego is Driving AI Progress | LinkedIn.”

Accessed: Jun. 26, 2024. [Online]. Available: <https://www.linkedin.com/pulse/e13-chiplets-how-chip-lego-driving-ai-progress-lundy-bryan-frsa/>

[11] R. P. Filho, “What is a Chiplet, and Why Should You Care? | Keysight Blogs.”

Accessed: Jun. 26, 2024. [Online]. Available:

<https://www.keysight.com/blogs/en/tech/sim-des/2024/2/8/what-is-a-chiplet-and-why-should-you-care>

[12] “1.1 Semiconductor Industry: Present & Future | IEEE Conference Publication |

IEEE Xplore.” Accessed: Jun. 30, 2024. [Online]. Available:

https://ieeexplore.ieee.org/abstract/document/10454358?casa_token=0TMRDSb8-mMAAAAAA:_d8jqOhh8XzPmxY906ni2vhFeVpw1x-WIcbE28KufI0gIH7sSAtnbilELs3LpSn4KIBQsZP_

[13] “All About Chiplet Technology.” Accessed: Jun. 26, 2024. [Online]. Available:

<https://resources.pcb.cadence.com/blog/2023-all-about-chiplet-technology>

[14] “Understanding the Impact of Memory Access Patterns in Intel Processors | IEEE

Conference Publication | IEEE Xplore.” Accessed: Jun. 29, 2024. [Online].

Available: <https://ieeexplore.ieee.org/document/9307131>

- [15] A. Gholami, "AI and Memory Wall," riselab. Accessed: Jun. 30, 2024. [Online]. Available: <https://medium.com/riselab/ai-and-memory-wall-2cb4265cb0b8>
- [16] "Heterogeneous Integration Sets Up an Inflection in Panel Processing Equipment (Drivers)," Lam Research Newsroom. Accessed: Jun. 29, 2024. [Online]. Available: <https://newsroom.lamresearch.com/heterogeneous-integration-panel-processing>
- [17] "2019 Edition - IEEE Electronics Packaging Society." Accessed: Jun. 30, 2024. [Online]. Available: <https://eps.ieee.org/technology/heterogeneous-integration-roadmap/2019-edition.html>
- [18] "Low Temperature SoIC Bonding and Stacking Technology for 12-/16-Hi High Bandwidth Memory (HBM) | IEEE Journals & Magazine | IEEE Xplore." Accessed: Jun. 30, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9195779?casa_token=wfs3tkQ_QiAAAAAA:XVCvEDZbjc0MYWYii0x720BiuyyXuEnPtfVAMJbhzw sCepPnWL-sCpbmJo-SB86X5QmVSame
- [19] S. S. Iyer, "Heterogeneous Integration for Performance and Scaling," *IEEE Transactions on Components, Packaging and Manufacturing Technology*, vol. 6, no. 7, pp. 973–982, Jul. 2016, doi: 10.1109/TCPMT.2015.2511626.

- [20] A. Elsherbini, S. Liff, J. Swan, K. Jun, S. Tiagaraj, and G. Pasdast, "Hybrid Bonding Interconnect for Advanced Heterogeneously Integrated Processors," in *2021 IEEE 71st Electronic Components and Technology Conference (ECTC)*, Jun. 2021, pp. 1014–1019. doi: 10.1109/ECTC32696.2021.00166.
- [21] M.-F. Chen, F.-C. Chen, W.-C. Chiou, and D. C. H. Yu, "System on Integrated Chips (SoIC(TM) for 3D Heterogeneous Integration," in *2019 IEEE 69th Electronic Components and Technology Conference (ECTC)*, May 2019, pp. 594–599. doi: 10.1109/ECTC.2019.00095.
- [22] "IEEE Solid-States Circuits Magazine - Fall 2021 - 82." Accessed: Jun. 30, 2024. [Online]. Available: https://www.nxtbook.com/nxtbooks/ieee/mssc_fall2021/index.php?startid=82#/p/82
- [23] "High Performance and Energy Efficient Computing with Advanced SoIC™ Scaling | IEEE Conference Publication | IEEE Xplore." Accessed: Jun. 30, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9816724?casa_token=SjIofWesCTkAAAAA:KVGLn8ci6Ve-zvZuK_HUKb6XplxAxugyR6TbbM0-HS_KuTN3W5m0vdkOQMYqQLdoZHJnQ9KY
- [24] "Interposers," Semiconductor Engineering. Accessed: Jun. 30, 2024. [Online]. Available: https://semiengineering.com/knowledge_centers/packaging/advanced-packaging/2-5d-ic/interposers/

- [25] “Embedded Multi-die Interconnect Bridge (EMIB) -- A High Density, High Bandwidth Packaging Interconnect | IEEE Conference Publication | IEEE Xplore.” Accessed: Jun. 30, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/7545486?casa_token=FtT-bUpiO_kAAAAA:_AUQS15F9VsJg5HrhjuXc5BV_ZgUIodvSzo7YEFohlo95YLM52_-OVDe6aYVb8Oo-NZ9s5yJ
- [26] “CoWoS® - Taiwan Semiconductor Manufacturing Company Limited.” Accessed: Jun. 30, 2024. [Online]. Available: <https://3dfabric.tsmc.com/english/dedicatedFoundry/technology/cowos.htm>
- [27] “Interposer Technologies for High-Performance Applications | IEEE Journals & Magazine | IEEE Xplore.” Accessed: Jun. 30, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/7883901?casa_token=jW_j3jvazCgAAAA:4T9BB6uK9qFbolmGRCg80UhJJwoh1HsYdf3Hcuhzbdp6wmAUmodUlhMXzg8nDJbVqgt-ON
- [28] D. Schor, “TSMC Demonstrates A 7nm Arm-Based Chiplet Design for HPC,” WikiChip Fuse. Accessed: Jun. 30, 2024. [Online]. Available:<https://fuse.wikichip.org/news/2446/tsmc-demonstrates-a-7nm-arm-based-chiplet-design-for-hpc/>
- [29] “Design, Materials, Process, Fabrication, and Reliability of Fan-Out Wafer-Level Packaging | IEEE Journals & Magazine | IEEE Xplore.” Accessed: Jun. 30, 2024.

[Online]. Available: https://ieeexplore.ieee.org/abstract/document/8353150?casa_token=e9z1vz8s_3EAAAAA:3x5lnlh13gOTNjWUEKMDrDfd5nQN2fpJFB60eTd5Fn-hsVX24Oc-MATRk7ahve6ZG0fB7lK

- [30] “Fan-out wafer-level packaging,” *Wikipedia*. Mar. 29, 2024. Accessed: Jun. 30, 2024. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Fan-out_wafer-level_packaging&oldid=1216234553
- [31] “InFO (Integrated Fan-Out) Wafer Level Packaging - Taiwan Semiconductor Manufacturing Company Limited.” Accessed: Jun. 30, 2024. [Online]. Available: <https://3dfabric.tsmc.com/english/dedicatedFoundry/technology/InFO.htm>
- [32] T. Braun *et al.*, “Opportunities of Fan-out Wafer Level Packaging (FOWLP) for RF applications,” in *2016 IEEE 16th Topical Meeting on Silicon Monolithic Integrated Circuits in RF Systems (SiRF)*, Jan. 2016, pp. 35–37. doi: 10.1109/SIRF.2016.7445461.
- [33] “Signal Integrity of Submicron InFO Heterogeneous Integration for High Performance Computing Applications | IEEE Conference Publication | IEEE Xplore.” Accessed: Jun. 30, 2024. [Online]. Available: <https://ieeexplore.ieee.org/document/8811046>
- [34] B. Bailey, “True 3D Is Much Tougher Than 2.5D,” *Semiconductor Engineering*. Accessed: Jun. 30, 2024. [Online]. Available: <https://semiengineering.com/true-3d-is->

much-tougher-than-2-5d/

- [35] P. Gratz, C. Kim, R. McDonald, S. W. Keckler, and D. Burger, "Implementation and Evaluation of On-Chip Network Architectures," in *2006 International Conference on Computer Design*, San Jose, CA, USA: IEEE, Oct. 2006, pp. 477–484. doi: 10.1109/ICCD.2006.4380859.
- [36] M. Manley, "Towards Selective Cobalt Atomic Layer Deposition for Chip-to-Wafer 3D Heterogeneous Integration," Center of Heterogeneous Integration of Micro Electronic Systems.
- [37] L. Wang *et al.*, "Direct Bond Interconnect (DBI®) for fine-pitch bonding in 3D and 2.5D integrated circuits," in *2017 Pan Pacific Microelectronics Symposium (Pan Pacific)*, Feb. 2017, pp. 1–6. Accessed: Jun. 26, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/7859568?casa_token=6TGVAXJZTisAAAA:98gCKNRACqd_niSfRsuR_WbZBBvAjwKO6oajo1n6ufR07_qNM_dxUXn9b6F8PC9ZkakGm1hm
- [38] Y. Ma, A. Roshanghias, and A. Binder, "A comparative study on direct Cu–Cu bonding methodologies for copper pillar bumped flip-chips," *J Mater Sci: Mater Electron*, vol. 29, no. 11, pp. 9347–9353, Jun. 2018, doi: 10.1007/s10854-018-8965-8.
- [39] T. H. Kim, M. M. R. Howlader, T. Itoh, and T. Suga, "Room temperature Cu–Cu

- direct bonding using surface activated bonding method,” *Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films*, vol. 21, no. 2, pp. 449–453, Mar. 2003, doi: 10.1116/1.1537716.
- [40] D. Liu *et al.*, “Investigation of Low-Temperature Cu–Cu Direct Bonding With Pt Passivation Layer in 3-D Integration,” *IEEE Trans. Compon., Packag. Manufact. Technol.*, vol. 11, no. 4, pp. 573–578, Apr. 2021, doi: 10.1109/TCPMT.2021.3069085.
- [41] Z. Zhang and C. P. Wong, “Flip-Chip Underfill: Materials, Process and Reliability,” in *Materials for Advanced Packaging*, D. Lu and C. P. Wong, Eds., Boston, MA: Springer US, 2009, pp. 307–337. doi: 10.1007/978-0-387-78219-5_9.
- [42] K. Darbha, J. H. Okura, and A. Dasgupta, “Impact of underfill filler particles on reliability of flip-chip interconnects,” *IEEE Trans. Comp., Packag., Manufact. Technol. A*, vol. 21, no. 2, pp. 275–280, Jun. 1998, doi: 10.1109/95.705475.
- [43] “Advancing Cu-Cu Hybrid Bonding: The Future of Semiconductor Packaging,” IDTechEx. Accessed: Jun. 26, 2024. [Online]. Available: <https://www.idtechex.com/en/research-article/advancing-cu-cu-hybrid-bonding-the-future-of-semiconductor-packaging/29847>
- [44] “Gearing Up For Hybrid Bonding.” Accessed: Jun. 26, 2024. [Online]. Available: <https://semiengineering.com/gearing-up-for-hybrid-bonding/>

- [45] R. Barnett, "Plasma Dicing Enables Challenging Applications Including D2W Hybrid Bonding," 3D InCites. Accessed: Jun. 26, 2024. [Online]. Available: <https://www.3dincites.com/2023/06/plasma-dicing-enables-challenging-applications-including-d2w-hybrid-bonding/>
- [46] A. Shorey *et al.*, "Advancements in fabrication of glass interposers," in *2014 IEEE 64th Electronic Components and Technology Conference (ECTC)*, Orlando, FL, USA: IEEE, May 2014, pp. 20–25. doi: 10.1109/ECTC.2014.6897261.
- [47] "Wafer-level polymer/metal hybrid bonding using a photosensitive permanent bonding material featured in Chip Scale Review - Brewer Science." Accessed: Jun. 26, 2024. [Online]. Available: https://www.brewerscience.com/csr_wlp_permanentbonding/
- [48] P. Nimbalkar, P. Bhaskar, M. Kathaperumal, M. Swaminathan, and R. R. Tummala, "A Review of Polymer Dielectrics for Redistribution Layers in Interposers and Package Substrates," *Polymers*, vol. 15, no. 19, p. 3895, Sep. 2023, doi: 10.3390/polym15193895.
- [49] P. Mah, "Intel Launches Gaudi 3 AI Accelerator," CDOTrends. Accessed: Jun. 26, 2024. [Online]. Available: <https://www.cdotrends.com/story/3922/intel-launches-gaudi-3-ai-accelerator>

- [50] D. G. Kam *et al.*, “Is 25 Gb/s On-Board Signaling Viable?,” *IEEE Trans. Adv. Packag.*, vol. 32, no. 2, pp. 328–344, May 2009, doi: 10.1109/TADVP.2008.2011138.
- [51] “A view on the 3D technology landscape.” Accessed: Jun. 26, 2024. [Online]. Available: <https://www.imec-int.com/en/articles/view-3d-technology-landscape>
- [52] H.-W. Hu and K.-N. Chen, “Development of low temperature Cu Cu bonding and hybrid bonding for three-dimensional integrated circuits (3D IC),” *Microelectronics Reliability*, vol. 127, p. 114412, Dec. 2021, doi: 10.1016/j.microrel.2021.114412.
- [53] B. Rebhan and K. Hingerl, “Physical mechanisms of copper-copper wafer bonding,” *Journal of Applied Physics*, vol. 118, no. 13, p. 135301, Oct. 2015, doi: 10.1063/1.4932146.
- [54] S. L. Chua, G. Y. Chong, Y. H. Lee, and C. S. Tan, “Direct copper-copper wafer bonding with Ar/N₂ plasma activation,” in *2015 IEEE International Conference on Electron Devices and Solid-State Circuits (EDSSC)*, Singapore, Singapore: IEEE, Jun. 2015, pp. 134–137. doi: 10.1109/EDSSC.2015.7285068.
- [55] T. Suga, R. He, G. Vakanas, and A. La Manna, “Direct Cu to Cu Bonding and Alternative Bonding Techniques in 3D Packaging,” in *3D Microelectronic Packaging*, vol. 64, Y. Li and D. Goyal, Eds., in Springer Series in Advanced Microelectronics, vol. 64. , Singapore: Springer Singapore, 2021, pp. 201–231. doi:

10.1007/978-981-15-7090-2_8.

- [56] M. S. Jeong *et al.*, “Unraveling diffusion behavior in Cu-to-Cu direct bonding with metal passivation layers,” *Sci Rep*, vol. 14, no. 1, p. 6665, Mar. 2024, doi: 10.1038/s41598-024-57379-2.
- [57] C.-M. Liu *et al.*, “Low-temperature direct copper-to-copper bonding enabled by creep on (111) surfaces of nanotwinned Cu,” *Sci Rep*, vol. 5, no. 1, p. 9734, May 2015, doi: 10.1038/srep09734.
- [58] K.-X. Chen, L.-Y. Gao, Z. Li, R. Sun, and Z.-Q. Liu, “Research Progress of Electroplated Nanotwinned Copper in Microelectronic Packaging,” *Materials*, vol. 16, no. 13, p. 4614, Jun. 2023, doi: 10.3390/ma16134614.
- [59] Y.-C. Huang, Y.-X. Lin, C.-K. Hsiung, T.-H. Hung, and K.-N. Chen, “Cu-Based Thermocompression Bonding and Cu/Dielectric Hybrid Bonding for Three-Dimensional Integrated Circuits (3D ICs) Application,” *Nanomaterials*, vol. 13, no. 17, p. 2490, Sep. 2023, doi: 10.3390/nano13172490.
- [60] S. W. Park, S. K. Hong, S. E. Kim, and J. K. Park, “First Demonstration of Enhanced Cu-Cu Bonding at Low Temperature With Ruthenium Passivation Layer,” *IEEE Access*, vol. 12, pp. 82396–82401, 2024, doi: 10.1109/ACCESS.2024.3409622.
- [61] A. Abas, Z. L. Gan, M. H. H. Ishak, M. Z. Abdullah, and S. F. Khor, “Lattice

- Boltzmann Method of Different BGA Orientations on I-Type Dispensing Method,”
PLoS ONE, vol. 11, no. 7, p. e0159357, Jul. 2016, doi:
10.1371/journal.pone.0159357.
- [62] M. K. Schwiebert and W. H. Leong, “Underfill flow as viscous flow between parallel plates driven by capillary action,” *IEEE Trans. Comp., Packag., Manufact. Technol. C*, vol. 19, no. 2, pp. 133–137, Apr. 1996, doi: 10.1109/3476.507149.
- [63] Y. Wen *et al.*, “Advances on Thermally Conductive Epoxy-Based Composites as Electronic Packaging Underfill Materials—A Review,” *Advanced Materials*, vol. 34, no. 52, p. 2201023, Dec. 2022, doi: 10.1002/adma.202201023.
- [64] Z. Zhang and C. P. Wong, “Recent Advances in Flip-Chip Underfill: Materials, Process, and Reliability,” *IEEE Trans. Adv. Packag.*, vol. 27, no. 3, pp. 515–524, Aug. 2004, doi: 10.1109/TADV.2004.831870.
- [65] “Hybrid No-flow Underfill for Flip Chip | Semiconductor Digest.” Accessed: Jun. 26, 2024. [Online]. Available: <https://sst.semiconductor-digest.com/2007/07/hybrid-no-flow-underfill-for-flip-chip/>
- [66] H. Ren, Y.-T. Yang, G. Ouyang, and S. S. Iyer, “Mechanism and Process Window Study for Die-to-Wafer (D2W) Hybrid Bonding,” *ECS J. Solid State Sci. Technol.*, vol. 10, no. 6, p. 064008, Jun. 2021, doi: 10.1149/2162-8777/ac0a52.

- [67] J. H. Lau, “Hybrid Bonding,” in *Semiconductor Advanced Packaging*, Singapore: Springer Singapore, 2021, pp. 379–411. doi: 10.1007/978-981-16-1376-0_8.
- [68] “Role of 3D Cu-Cu Hybrid Bonding in Powering Future HPC & AI Products,” IDTechEx. Accessed: Jun. 26, 2024. [Online]. Available: <https://www.idtechex.com/en/research-article/role-of-3d-cu-cu-hybrid-bonding-in-powering-future-hpc-and-ai-products/30902>
- [69] S. Wang *et al.*, “Optimization of Cu protrusion of wafer-to-wafer hybrid bonding for HBM packages application,” *Materials Science in Semiconductor Processing*, vol. 152, p. 107063, Dec. 2022, doi: 10.1016/j.mssp.2022.107063.
- [70] A. Palesko and C. Palesko, “Cost and Yield Comparison of Wafer-to-Wafer, Die-to-Wafer, and Die-to-Die Bonding,” *International Symposium on Microelectronics*, vol. 2014, no. 1, pp. 000624–000629, Oct. 2014, doi: 10.4071/isom-WP16.
- [71] J. A. Theil, L. Mirkarimi, G. Fountain, G. Gao, and R. Katkar, “Recent Developments in Fine Pitch Wafer-To-Wafer Hybrid Bonding with Copper Interconnect,” in *2019 International Wafer Level Packaging Conference (IWLPC)*, Oct. 2019, pp. 1–6. doi: 10.23919/IWLPC.2019.8913862.
- [72] S. A. Chew, J. De Vos, and E. Beyne, “Wafer-to-wafer hybrid bonding at 400-nm interconnect pitch,” *Nat Rev Electr Eng*, vol. 1, no. 2, pp. 71–72, Feb. 2024, doi: 10.1038/s44287-024-00019-8.

- [73] “TSMC-SoIC® - Taiwan Semiconductor Manufacturing Company Limited.”
Accessed: Jun. 28, 2024. [Online]. Available:
https://3dfabric.tsmc.com/english/dedicatedFoundry/technology/SoIC.htm#SoIC_WoW
- [74] B. Solca, “AMD presents more details on Zen 3 3D V-Cache and the future of 3D stacking,” Notebookcheck. Accessed: Jul. 01, 2024. [Online]. Available:
<https://www.notebookcheck.net/AMD-presents-more-details-on-Zen-3-3D-V-Cache-and-the-future-of-3D-stacking.556527.0.html>
- [75] J. Wuu *et al.*, “3D V-Cache: the Implementation of a Hybrid-Bonded 64MB Stacked Cache for a 7nm x86-64 CPU,” in *2022 IEEE International Solid-State Circuits Conference (ISSCC)*, Feb. 2022, pp. 428–429. doi:
10.1109/ISSCC42614.2022.9731565.
- [76] D. Patel, “Hybrid Bonding Process Flow - Advanced Packaging Part 5.”
Accessed: Jun. 26, 2024. [Online]. Available:
<https://www.semianalysis.com/p/hybrid-bonding-process-flow-advanced>
- [77] “AMD Unveils More Ryzen 3D Packaging and V-Cache Details at Hot Chips | Tom’s Hardware.” Accessed: Jun. 28, 2024. [Online]. Available:
<https://www.tomshardware.com/news/amd-unveils-more-ryzen-3d-packaging-and-v-cache-details-at-hot-chips>

- [78] D. Schor, “Intel Unveils Foveros Omni And Foveros Direct; Leveraging Hybrid Bonding,” WikiChip Fuse. Accessed: Jun. 28, 2024. [Online]. Available: <https://fuse.wikichip.org/news/5949/intel-unveils-foveros-omni-and-foveros-direct-leveraging-hybrid-bonding/>
- [79] H. Mujtaba, “Intel Clearwater Forest Xeon CPUs With Up To 288 E-Cores To Utilize Foveros Direct 3D Stacking Technology,” Wccftech. Accessed: Jun. 28, 2024. [Online]. Available: <https://wccftech.com/intel-clearwater-forest-xeon-cpus-288-e-cores-utilize-foveros-direct-3d-stacking-tech/>
- [80] J. H. Lau, “Recent Advances and Trends in Chiplet Design and Heterogeneous Integration Packaging,” *Journal of Electronic Packaging*, vol. 146, no. 1, p. 010801, Mar. 2024, doi: 10.1115/1.4062529.
- [81] “Intel’s new 3D Foveros packaging tech: LEGO-like chiplets for CPUs,” TweakTown. Accessed: Jun. 28, 2024. [Online]. Available: <https://www.tweaktown.com/news/88105/intels-new-3d-foveros-packaging-tech-lego-like-chiplets-for-cpus/index.html>
- [82] A. Pek, “Intel Beyond 2025: 5 Nodes In 4 Years (Part 2),” Value Investing Substack. Accessed: Jun. 28, 2024. [Online]. Available: <https://valueinvesting.substack.com/p/intel2>

- [83] “Samsung maintains hybrid bonding needed for HBM 16H,” THE ELEC, Korea Electronics Industry Media. Accessed: Jun. 28, 2024. [Online]. Available: <http://thelec.net/news/articleView.html?idxno=4868>
- [84] A. F. Asia Taipei; Jack Wu, DIGITIMES, “As HBM competition heats up, introduction schedule of ‘hybrid bonding’ receiving attention,” DIGITIMES. Accessed: Jun. 28, 2024. [Online]. Available: <https://www.digitimes.com/news/a20230920PD206/hbm-hybrid-bonding-samsung-sk-hynix.html>
- [85] K.-I. Moon, H.-Y. Son, and K. Lee, “Advanced Packaging Technologies in Memory Applications for Future Generative AI Era,” in *2023 International Electron Devices Meeting (IEDM)*, San Francisco, CA, USA: IEEE, Dec. 2023, pp. 1–4. doi: 10.1109/IEDM45741.2023.10413890.
- [86] “SK Hynix’s Next-Gen HBM Embraces Hybrid Bonding Tech - Electronic Components Distributor - SMBOM.COM,” SK Hynix’s Next-Gen HBM Embraces Hybrid Bonding Tech - Electronic Components Distributor - SMBOM.COM. Accessed: Jun. 28, 2024. [Online]. Available: <https://www.smbom.com/news/13001>
- [87] “[News] Samsung Considers Hybrid Bonding a Must for 16-stack HBM,” TrendForce Insights. Accessed: Jun. 28, 2024. [Online]. Available: <https://www.trendforce.com/news/2024/06/12/news-samsung-considers-hybrid-bonding-a-must-for-16-stack-hbm/>

- [88] T. C. Daily, “SK Hynix, Samsung unveil key differences in HBM techniques,” The Chosun Daily. Accessed: Jun. 28, 2024. [Online]. Available: <https://www.chosun.com/english/industry-en/2024/04/25/OOX326SCQ5FTFKIXTB7LTTSXKQ/>
- [89] “Samsung completes 16-stack HBM sample,” THE ELEC, Korea Electronics Industry Media. Accessed: Jun. 28, 2024. [Online]. Available: <http://thelec.net/news/articleView.html?idxno=4788>
- [90] “Wafer-to-wafer hybrid bonding | imec.” Accessed: Jun. 26, 2024. [Online]. Available: <https://www.imec-int.com/en/articles/wafer-wafer-hybrid-bonding-pushing-boundaries-400nm-interconnect-pitch>
- [91] T. Uhrmann, “Die-to-Wafer Bonding Steps into the Spotlight on a Heterogeneous Integration Stage,” 3D InCites. Accessed: Jun. 26, 2024. [Online]. Available: <https://www.3dincites.com/2021/05/die-to-wafer-bonding-steps-into-the-spotlight-on-a-heterogeneous-integration-stage/>
- [92] J. Amandine *et al.*, “Self-Assembly Process for 3D Die-to-Wafer using Direct Bonding: A Step Forward Toward Process Automatisation,” in *2019 IEEE 69th Electronic Components and Technology Conference (ECTC)*, Las Vegas, NV, USA: IEEE, May 2019, pp. 225–234. doi: 10.1109/ECTC.2019.00041.

- [93] “Hybrid Bonding Moves Into The Fast Lane.” Accessed: Jun. 26, 2024. [Online]. Available: <https://semiengineering.com/hybrid-bonding-moves-into-the-fast-lane/>
- [94] H. Ishida, S. Sood, C. Rosenthal, and S. Lutter, “Temporary bonding/de-bonding and permanent wafer bonding solutions for 3D integration,” in *2012 2nd IEEE CPMT Symposium Japan*, Kyoto, Japan: IEEE, Dec. 2012, pp. 1–4. doi: 10.1109/ICSJ.2012.6523416.
- [95] “Nomination #83707,” 3D InCites. Accessed: Jul. 01, 2024. [Online]. Available: <https://www.3dincites.com/awardnominee/83707/>
- [96] W. Zhou *et al.*, “Critical Challenges with Copper Hybrid Bonding for Chip-to-Wafer Memory Stacking,” in *2023 IEEE 73rd Electronic Components and Technology Conference (ECTC)*, Orlando, FL, USA: IEEE, May 2023, pp. 336–341. doi: 10.1109/ECTC51909.2023.00063.
- [97] W. Yang, H. Shintani, M. Akaike, and T. Suga, “Low temperature Cu-Cu direct bonding using formic acid vapor pretreatment,” in *2011 IEEE 61st Electronic Components and Technology Conference (ECTC)*, Lake Buena Vista, FL, USA: IEEE, May 2011, pp. 2079–2083. doi: 10.1109/ECTC.2011.5898804.
- [98] K. Wang, Victor Chenghsuan, “Effects of Inverse Templating on Ru ALD grown on sputtered Ru,” presented at the SRC CHIMES Annual Review ’23,

- [99] “Properties: Alumina - Aluminium Oxide - Al₂O₃ - A Refractory Ceramic Oxide,” AZoM. Accessed: Jun. 26, 2024. [Online]. Available: <https://www.azom.com/properties.aspx?ArticleID=52>
- [100] “Aluminum Oxide - Panadyne.” Accessed: Jun. 26, 2024. [Online]. Available: <https://panadyne.com/aluminum-oxide/>
- [101] S. Eslava *et al.*, “Metal-Organic Framework ZIF-8 Films As Low- κ Dielectrics in Microelectronics,” *Chem. Mater.*, vol. 25, no. 1, pp. 27–33, Jan. 2013, doi: 10.1021/cm302610z.
- [102] B. Cui *et al.*, “Thermal Conductivity of ZIF-8 Thin-Film under Ambient Gas Pressure,” *ACS Appl. Mater. Interfaces*, vol. 9, no. 34, pp. 28139–28143, Aug. 2017, doi: 10.1021/acsami.7b06662.