A High Throughput Two-Stage Die-to-Wafer Thermal Compression Bonding Scheme for Heterogeneous Integration

Krutikesh Sahoo
University of California, Los Angeles
Los Angeles, USA, 90095
krutikesh@ucla.edu

Haoxiang Ren
University of California, Los Angeles
Los Angeles, USA, 90095
haoxiang.ren@ucla.edu

Subramanian S. Iyer
University of California, Los Angeles
Los Angeles, USA, 90095
s.s.iyer@ucla.edu

Abstract—In this work, we demonstrate for the first time a two-stage high throughput fine-pitch die-to-wafer Copper-Copper (Cu) thermal compression bonding (TCB) technique, which has a throughput of > 300 units-per-hour (UPH), with the potential to increase the throughput to 1100 UPH. We have optimized the bonding for high throughput, high overlay accuracy and low contact resistance with a die-to-substrate bump pitch of ≤ 10 μm extendible to ~ 7 μm pitch. The average shear force per 2x2 mm^2 die after high throughput TCB is > 110 N. The average specific contact resistance of the Cu-Cu contact is 1.24×10^{-9} Ω·cm^2, comparable to the lowest reported in Cu/SiO$_2$ hybrid bonding. The simplicity of TCB compared to hybrid bonding (HB) makes it a preferred approach for heterogeneous integration for 3D stacking, interposers, and the Silicon Interconnect Fabric (Si-IF) down to ~ 7 μm bonding pitches.

Keywords— advanced packaging, high-throughput bonding, thermal compression bonding, Silicon-Interconnect Fabric (Si-IF), multi-reticle interposer assembly

I. INTRODUCTION

Advanced packaging aims to integrate multiple heterogeneous dies close to one another at fine bonding pitches (≤ 10 μm) — mimicking a monolithic chip, but with heterogeneity and scale. A key requirement is to achieve bonding pitches down to the "fat wire/via levels" of a monolithic chip [1], effectively increasing the interconnect density by orders of magnitude over conventional μ-bumps. While there are reports of solder assisted sub 10 μm bumping [2], it is generally accepted that direct Cu-Cu bonding is the preferred approach due to better electrical and thermal performance, especially in stacked 3D architectures [3]. Two approaches for direct Cu-Cu bonding are thermal compression bonding (TCB) and hybrid bonding (HB). Die-to-substrate TCB at ≤ 10 μm bump pitches has been pursued in our lab [4], while hybrid bonding has been pursued by others [5, 6]. In this work, we present a novel two-stage TCB approach that compares favorably with the HB approach. An important application of two-stage TCB will be in the assembly of wafer-scale systems, which have been proposed for graph processing, data analytics and artificial intelligence-based applications [7].

As described by Ren *et al.* [8], die-to-wafer (D2W) hybrid bonding (HB) offers the potential benefit of higher throughput and room temperature die tacking (placement), but at the cost of lower process tolerance and stringent processing requirements as shown in Fig. 1. These requirements involve: (1) extremely flat (roughness ≤ 1 nm) dielectric surfaces for strong attraction
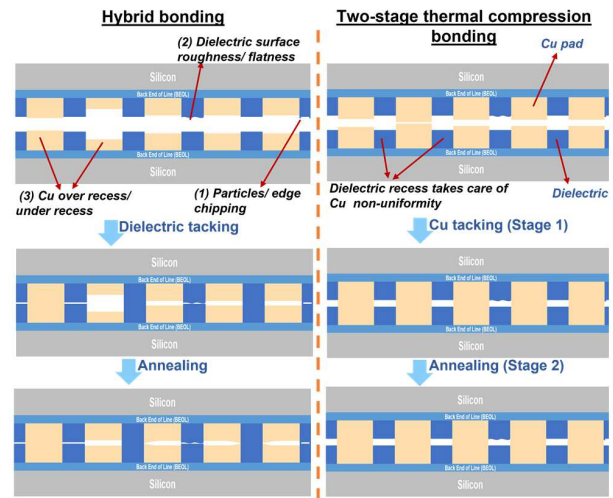
Fig. 1. Comparison between D2W HB and D2W two-stage Cu-Cu TCB which highlights challenges with HB due to metal recess, dielectric roughness, and particles. Most of these challenges are overcome in D2W two-stage TCB bonding by introducing small amount of metal recess.

forces during initial dielectric attach, (2) meticulous control of metal recess to ensure successful stress-free metal-to-metal connections during annealing, and (3) complete particle control of bonding environment to avoid failed connections. Previously [5], Intel reported that a major revision of today's die stacking process is needed to enable HB on a manufacturable scale. Extensive chemical mechanical polishing (CMP) process optimization to address metal recess requirements, use of particle-free dicing techniques (e.g., plasma dicing), and special pick-and-place tools for highly accurate die placement are a few required changes. The summary is that D2W-HB process development is expensive and additionally, has a narrow process window.

A D2W-TCB process is significantly simpler. In TCB, bondable pads/pillars are recessed to achieve metal-to-metal contact. Since there is no dielectric bonding, dielectric roughness requirements are not critical. Surface asperities on bonding pads are flattened by temperature and pressure during thermal compression bonding. D2W-TCB is independent of type of dicing used, so blade dicing is applicable. Furthermore, level of particle control obtained through standard wet cleaning

processes is adequate for successful assembly. Previously, we have successfully demonstrated ≤ 10 μm bump-pitch Cu-Cu [4] and Cu-Au [9] D2W-TCB assemblies, including the first functional assembly at 9.8 μm pitch – a scalable digital signal processor [10, 11] on a silicon-based packaging substrate – Silicon Interconnect Fabric (Si-IF) [1, 4]. Si-IF substrate consists of back end of line Cu metal layers terminating with Cu pillars. Bare dies terminating with Cu pads can be directly thermal compression bonded to the Cu pillars on Si-IF using Cu-Cu TCB.

However, one of the limitations of conventional TCB which makes it less attractive from a manufacturing perspective, is its lower bonding throughput. As a case study, our previous bonding times of ≥ 30 s/die [4] presented throughput limitations and furthermore used high temperatures (~240 °C) during alignment, which in turn limited alignment accuracy to ±1 μm. There is also the need for passivation of exposed Cu surfaces against moisture ingression after TCB. Therefore, to improve the scalability of D2W-TCB and to demonstrate throughput improvement, we have developed a parallelized bonding approach by separating the die tacking step from the die bonding step. By using the two-stage TCB approach, we show a ~3X throughput improvement without compromising mechanical or electrical reliability of the bonds. We also passivate the entire assembly with 10 nm Al_2O_3 (alumina) using atomic layer deposition (ALD) technique first discussed in [12] and demonstrate military-level reliability standards. We believe that the simplicity of TCB will allow for a superior capability to HB process at significantly lower cost and time to market.

II. TWO-STAGE TCB APPROACH

The reliance on a single long die bonding step (≥ 30 s/die in our earlier approach) for both mechanical strength and electrical connectivity is the prime challenge in conventional TCB. Unfortunately, this time is necessary for Cu grain growth to occur at the mating interfaces during the TCB process. To overcome this limitation, we propose a two-stage approach: die tacking to Si-IF or interposer substrate, followed by annealing of the assembly. During the die tacking stage, all the dies are aligned at relatively low temperature of 120 °C and placed within a total time of ≤ 10 seconds/die. This step does not ensure final bonding, but it does guarantee a firm-enough attach with a shear strength > 10 N. Once populated, the assembly is batch annealed (batch size depends on furnace capacity) in vacuum for 1 hour. This step ensures Cu grain growth across the mating surfaces needed for successful bonding. For a wafer consisting of 2500 dies, we can get an aggregate bonding time of ~11.5 seconds/die, thereby giving a throughput of > 300 units per hour (UPH). By batch annealing several wafers at a time, annealing time is no longer an issue for throughput, and further optimization is still possible at the tacking stage.

III. THROUGHPUT IMPROVEMENT STUDY

A Design of Experiment (DOE) using more than 200 dies of size 2×2 mm², bonded to multiple Si-IF substrates at 10 μm Cu bump/pillar pitch was performed to improve throughput. The goal of the DOE was to reduce the overall die tacking time to ≤ 10 seconds per die and achieve a die shear strength value greater than that specified by MIL-STD 883G (MIL-SPEC) [13] for the corresponding die size after annealing. For a 2x2 mm² die, shear

force must exceed 50 N. To start the experiment, all the parameters which affect the bond integrity were identified. These parameters and the values used for optimization are listed in Table I. The DOE was performed in three phases, with two parameters being optimized per phase. During each phase, the parameters which were not tested for were set to the largest value from Table I, unless already optimized.

TABLE I.   PARAMETERS FOR TWO-STAGE CU-CU TCB OPTIMIZATION

Phase	Parameter Name	Value	Unit
1	Die cleaning time	3, 5	s
1	Die placement time	3, 5	s
2	Chuck temperature	80, 100, 120	℃
2	Tacking pressure	50, 150, 300	MPa
3	Annealing temperature	200, 300, 400	℃
3	Annealing pressure	20, 50, 100	MPa

A. Phase 1: Die tacking time optimization

In phase 1, die tacking time was optimized. The total tacking time is the sum of die cleaning time, die to substrate alignment time and die placement time. The die to substrate alignment time is a constant of 2 seconds (s) which depends on the bonding tool used. Also, as discussed in our previous works [4] we use formic acid vapor (FA) to perform an in-situ cleaning of the die pads right before bonding. This represents the die cleaning time. As seen in Fig. 2, two cases were tested (first two columns), and both satisfied the 10 s/die tacking constraint, but the one with 3 s cleaning time and 5 s die placement time yielded better results.
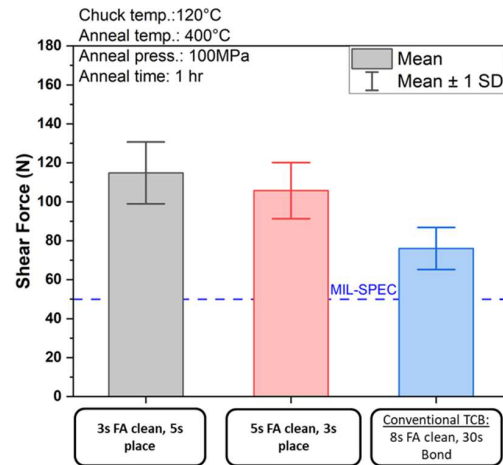

Fig. 2. Formic acid vapor (FA) cleaning and die placement optimization for phase 1 of two stage direct Cu-Cu TCB process. Two-stage TCB results are compared to conventional > 30s/die TCB and the former shows superior shear strength as compared to conventional TCB.

B. Phase 2: Die tacking condition optimization

Phase 2 addressed the substrate chuck temperature and tacking pressures optimizations. The results are summarized in

363
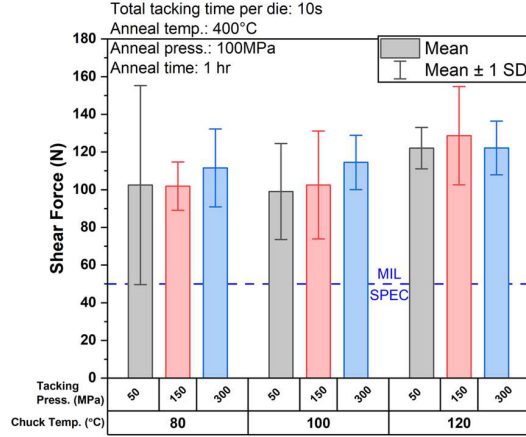
Fig. 3. Tacking pressure and chuck temperature optimization results from phase 2. Least variability in shear strength occurs at chuck temperature of 120 °C. Shear strength variability is high after annealing only if both chuck temperature and tacking pressure are low.

Fig. 3. Variability of shear strength decreased with increase of chuck temperature. This can be explained by considering the fact that in-situ formic acid (FA) vapor reaction is more effective at higher temperatures. Hence for a chuck temperature of 120 °C, the copper oxide reduction to Cu is most effective, and hence the bonding quality is higher. Increasing temperature further is not recommended as it affects alignment accuracy. The bonding process is found to be insensitive to tacking pressures, except at low chuck temperatures. At the end of the phase 2, a chuck temperature of 120 °C and tacking pressure of 50 MPa was finalized.

C. Phase 3: Annealing parameter optimizations

Annealing temperature and pressure were optimized in phase 3. Annealing temperature was found to have the most profound impact on bonding yields. As seen from Fig. 4, going from an anneal temperature of 400 °C to 300 °C, the bonding
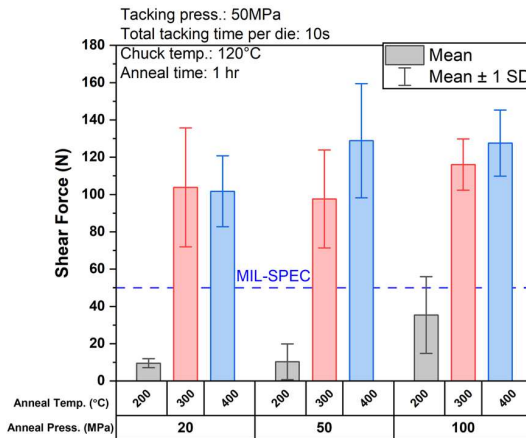
Fig. 4. Anneal temperature and anneal pressure optimization. Cluster direct bonding is found to fail at 200°C anneal temperatures which indicates that there is insufficient Cu diffusion across bonding interface.

quality generally degrades, whereas bonding catastrophically fails at 200 °C. Pressure has a smaller impact on shear strength for an anneal temperature > 200 °C.

IV. DOE RESULTS AND RELIABILITY EXPERIMENTS

### A. Optimized high throughput bonding process

Table II shows two possible variants for reliable two stage D2W TCB after optimization — a low temperature (LT) variant with anneal temperature of 300 °C but anneal pressure of 100 MPa, and a high temperature (HT) variant with anneal temperature of 400 °C and anneal pressure of 50 MPa. In either variant, the average die shear strength of 2x2 mm^2 dies is found to be > 110 N which is $\geq$ 2x the military specification (50 N). Depending on the requirements of the bonding process and types of dies used, one of the bonding variants can be chosen.

TABLE II.   OPTIMIZED TWO-STAGE DIRECT Cu-Cu TCB PROCESS

| Parameter Name	Unit	Low Temp. (LT) variant	High Temp. (HT) variant
Die cleaning time	s	3	3
Die placement time	s	5	5
Chuck temperature	ºC	120	120
Tacking pressure	MPa	50	50
Annealing temperature	ºC	300	400
Annealing pressure	MPa	100	50

B. Electrical and reliability results

Once the bonding parameters were optimized, dies with daisy chains were assembled using both LT and HT variants. A sample can be seen in Fig. 5, with cross-section in Fig. 6. No voids can be seen at the bonding interface in the cross-section micrograph, indicating a high-quality Cu-Cu bond. The cumulative distribution of resistance is presented in Fig. 7, measured across 40 daisy chains (DS) for each condition (before UHAST reliability test). The average Cu-Cu specific contact resistance was extracted to be 1.24×10^{-9} $\Omega \cdot cm^2$, comparable to the lowest reported values in HB case [14]. The samples were passivated using 10 nm Al_2O_3 deposition using ALD process as first discussed in [12]. To test the samples against extreme
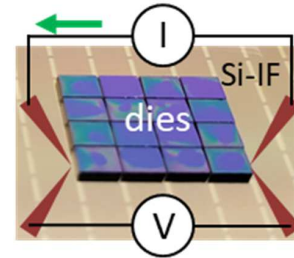

Fig. 5. D2W Cu-Cu two-stage TCB sample for electrical measurements. All resistance values are measured using four probe method. The average specific contact resistance of Cu-Cu contact after de-embedding the effect of Cu pad and wires is 1.24×10^{-9} $\Omega \cdot cm^2$ with a range from 1.2×10^{-9} $\Omega \cdot cm^2$ to 1.7×10^{-9} $\Omega \cdot cm^2$, measured in daisy chains across multiple dies. The variation in measured results occur due to misalignment during bonding process.

temperature and humidity conditions, we subjected passivated daisy chain samples to unbiased highly accelerated stress test (UHAST) at 130 °C and 85 % relative humidity for 96 hours (JEDEC JESD22-A118 test condition A). In both the bonding variants, the maximum resistance change was found to be < 10 %, with < 2 % change in case of HT bonding variant.
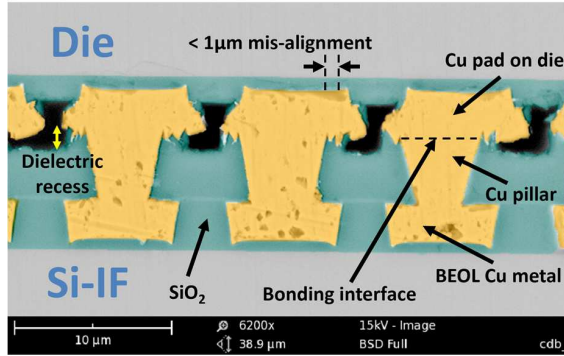

Fig. 6. scanning electron micrograph (SEM) of two-stage TCB cross-section. No voids are visible along the bonding interface. Dielectric recess enables direct metal-metal contact without dielectric involvement. Only pillar recess for bonding is also feasible. This sample is passivated with 10 nm Al_2O_3, but it cannot be resolved in this SEM.
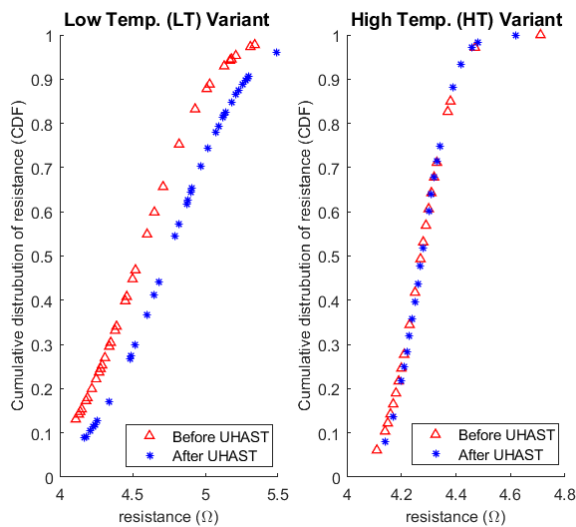
Fig. 7. The measured resistance values of 40 randomly chosen daisy chains each for both LT and HT bonding variants (some points may overlap due to measured resistance value of different daisy chains being similar) both before UHAST reliability measurement and after it. The average resistance change after UHAST is < 10% for both LT and HT variants.

V. TWO-STAGE TCB OF FUNCTIONAL DIES

Following passive daisy chain testing, functional dies were assembled using high throughput TCB process to observe the effect of bonding on the functionality of the dies. For this purpose, dies with active links, designed in 22 nm Global Foundry (22nm FDX) process, were bonded to corresponding Si-IF using the proposed two-stage bonding scheme. A micrograph of the 22 nm die before bonding is presented in Fig.
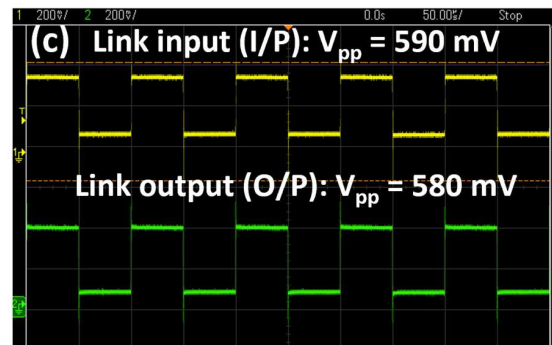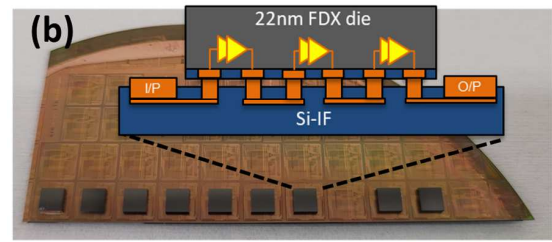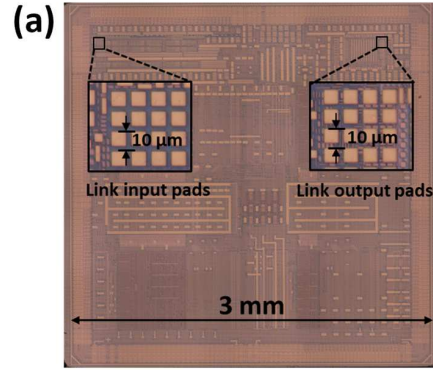
Fig. 8(a) 22nm Global Foundry (22nm FDX) dies showing 10μm pitch pads for active daisy chain measurements. The die micrograph was presented in [11]. **(b)** Assembly of 22 nm Global Foundry dies on Si-IF using two-stage Cu-Cu TCB using LT bonding variant. **(c)** Sample active link measurements after bonding. In active links, the input (in this case a 10 kHz square wave) is transmitted through buffers to the output. Functionality of the die is maintained after bonding.

8(a). The active links consisted of buffer drivers as shown in Fig. 8(b). The active links were verified to be working post bonding as shown in Fig. 8(c) where output follows the input. This test verifies that functionality of the dies remains intact after the bonding process.

VI. CONCLUSION

The proposed > 300 UPH novel two-stage direct TCB significantly improves the throughput of conventional D2W TCB. The throughput can be potentially increased to 1100 UPH with further optimizations to the tacking process. The two-stage TCB has greater tolerance to particulate control and does not require extensive optimization of chemical mechanical polishing to achieve reliable bonding. We have demonstrated that this bonding process surpasses MIL-SPEC requirements both in die-shear tests and electrical reliability tests. These

365

advantages, in our opinion, make two-stage TCB a reliable, simpler, and manufacturable bonding process with lower cost on process development when compared to D2W-HB. We expect two-stage high throughput TCB to provide a competitive and reliable low-cost alternative to D2W-HB on assemblies requiring bump pitches down to ~ 7 µm, with potential for faster time to market on 3D packaging products.

REFERENCES

[1] S. S. Iyer, "Heterogeneous Integration for Performance and Scaling," in IEEE Transactions on Components, Packaging and Manufacturing Technology, vol. 6, no. 7, pp. 973-982, July 2016, doi: 10.1109/TCPMT.2015.2511626.

[2] E. Beyne, D. Milojevic, G. Van der Plas and G. Beyer, "3D SoC integration, beyond 2.5D chiplets," 2021 IEEE International Electron Devices Meeting (IEDM), San Francisco, CA, USA, 2021, pp. 3.6.1-3.6.4, doi: 10.1109/IEDM19574.2021.9720614.

[3] A. Agrawal, S. Huang, G. Gao, L. Wang, J. DeLaCruz and L. Mirkarimi, "Thermal and Electrical Performance of Direct Bond Interconnect Technology for 2.5D and 3D Integrated Circuits," 2017 IEEE 67th Electronic Components and Technology Conference (ECTC), Orlando, FL, USA, 2017, pp. 989-998, doi: 10.1109/ECTC.2017.341.

[4] S. Jangam and S. S. Iyer, "Silicon-Interconnect Fabric for Fine-Pitch (≤10 μ m) Heterogeneous Integration," in IEEE Transactions on Components, Packaging and Manufacturing Technology, vol. 11, no. 5, pp. 727-738, May 2021, doi: 10.1109/TCPMT.2021.3075219.

[5] A. Elsherbini et al., "Enabling Hybrid Bonding on Intel Process," 2021 IEEE International Electron Devices Meeting (IEDM), San Francisco, CA, USA, 2021, pp. 34.3.1-34.3.4, doi: 10.1109/IEDM19574.2021.9720586.

[6] T. Kim et al., "Multi-Stack Wafer Bonding Demonstration utilizing Cu to Cu Hybrid Bonding and TSV enabling Diverse 3D Integration," 2021 IEEE 71st Electronic Components and Technology Conference (ECTC), San Diego, CA, USA, 2021, pp. 415-419, doi: 10.1109/ECTC32696.2021.00076.

[7] S. Pal et al., "I/O Architecture, Substrate Design, and Bonding Process for a Heterogeneous Dielet-Assembly based Waferscale Processor," 2021 IEEE 71st Electronic Components and Technology Conference (ECTC), San Diego, CA, USA, 2021, pp. 298-303, doi: 10.1109/ECTC32696.2021.00057.

[8] H. Ren, Y.-T. Yang, G. Ouyang, and S. S. Iyer, "Mechanism and Process Window Study for Die-to-Wafer (D2W) Hybrid Bonding," in ECS Journal of Solid State Science and Technology, vol. 10, no. 6, 2021. https://iopscience.iop.org/article/10.1149/2162-8777/ac0a52/meta

[9] K. Sahoo, S. Pal, N. Shakoorzadeh, Y. -T. Yang and S. S. Iyer, "Copper to gold thermal compression bonding in heterogenous wafer-scale systems," 2021 IEEE 71st Electronic Components and Technology Conference (ECTC), San Diego, CA, USA, 2021, pp. 487-493, doi: 10.1109/ECTC32696.2021.00088.

[10] S. S. Nagi, U. Rathore, K. Sahoo, T. Ling, S. S. Iyer and D. Marković, "A 16-nm 784-Core Digital Signal Processor Array, Assembled as a 2 × 2 Dielet With 10-µm Pitch Interdielet I/O for Runtime Multiprogram Reconfiguration," in IEEE Journal of Solid-State Circuits, vol. 58, no. 1, pp. 111-123, Jan. 2023, doi: 10.1109/JSSC.2022.3212685.

[11] K. Sahoo, U. Rathore, S. Chandra Jangam, T. Nguyen, D. Markovic and S. S. Iyer, "Functional Demonstration of < 0.4-pJ/bit, 9.8 µm Fine-Pitch Dielet-to-Dielet Links for Advanced Packaging using Silicon Interconnect Fabric," 2022 IEEE 72nd Electronic Components and Technology Conference (ECTC), San Diego, CA, USA, 2022, pp. 2104-2110, doi: 10.1109/ECTC51906.2022.00332.

[12] N. Shakoorzadeh Chase, K. Sahoo, Y. Tao Yang and S. S. Iyer, "Atomic Layer Deposited Al2O3 Encapsulation for the Silicon Interconnect Fabric," 2020 IEEE 70th Electronic Components and Technology Conference (ECTC), Orlando, FL, USA, 2020, pp. 1241-1246, doi: 10.1109/ECTC32862.2020.00198.

[13] MIL-STD-883G, available: http://everyspec.com/MIL-STD/MIL-STD-0800-0899/MIL-STD-883G_14587/

[14] Ong, Jia-Juen, Wei-Lan Chiu, Ou-Hsiang Lee, Chia-Wen Chiang, Hsiang-Hung Chang, Chin-Hung Wang, Kai-Cheng Shie, Shih-Chi Yang, Dinh-Phuc Tran, King-Ning Tu, and Chih Chen. 2022. "Low-Temperature Cu/SiO2 Hybrid Bonding with Low Contact Resistance Using (111)-Oriented Cu Surfaces" Materials 15, no. 5: 1888. https://doi.org/10.3390/ma15051888.