YAP+: Pad-Layout-Aware Yield Modeling and Simulation for Hybrid Bonding

Zhichao Chen, Puneet Gupta, *Fellow, IEEE*
zhichaochen@ucla.edu, puneetg@ucla.edu

***Abstract*—Three-dimensional (3D) integration continues to advance Moore's Law by facilitating dense interconnects and enabling multi-tier system architectures. Among the various integration approaches, Cu-Cu hybrid bonding has emerged as a leading solution for achieving high interconnect density in chiplet integration. In this work, we present YAP+, a yield modeling framework specifically tailored for wafer-to-wafer (W2W) and die-to-wafer (D2W) hybrid bonding processes. YAP+ incorporates a comprehensive set of yield-impacting failure mechanisms, including overlay misalignment, particle defects, Cu recess variations, surface roughness, and Cu pad density. Furthermore, YAP+ supports pad layout-aware yield analysis, considering critical, redundant, and dummy pads across arbitrary 2D physical layout patterns. To support practical evaluation, we developed an open-source yield simulator, demonstrating that our near-analytical model matches simulation accuracy while achieving over 1,000x speedup in runtime. This performance makes YAP+ a valuable tool for co-optimizing packaging technologies, assembly design rules, and system-level design strategies. Beyond W2W-D2W comparisons, we leverage YAP+ to investigate the impact of pad layout patterns, bonding pitch, and pad ratios across different pad types, and explore the benefits of strategically placing redundant pad replicas.**

***Index Terms*—yield modeling, hybrid bonding, wafer-to-wafer (W2W), die-to-wafer (D2W), critical area, dilation, chiplet, Cu dishing, particle defects, overlay, redundancy.**

I. Introduction

As the physical and economic boundaries of scaling traditional two-dimensional integrated circuits are increasingly challenged, Three-Dimensional Integrated Circuits (3D-ICs) have emerged as a compelling alternative to sustain the progression of Moore's Law. By vertically stacking multiple device layers, 3D-ICs offer notable advantages, including shorter interconnect paths, improved performance, reduced power consumption, and higher integration density. A key enabler of 3D-IC technology is the hybrid bonding (HB) process. Compared to earlier packaging technologies, HB offers a dramatic leap in interconnect density, reaching 10,000 to 1 million connections per mm^2, with sub-micron alignment accuracy down to 50 nm [1]–[3], and ultra-low energy consumption below 0.05 pJ/bit due to the lower resistance of direct Cu connections [4]. HB supports fine-pitch, high-reliability interconnects, making it ideal for applications such as high-bandwidth memory, logic-memory integration, and advanced sensing systems [5]–[9]. The two predominant HB approaches are wafer-to-wafer (W2W) and die-to-wafer (D2W). D2W provides enhanced flexibility by allowing verified top dies to be bonded onto known-good base dies, thereby improving overall yield [5]. In contrast, W2W bonding is more efficient for high-volume production and offers better alignment accuracy, but its yield is more vulnerable to defects in either wafer. These distinctions in process complexity and yield implications underscore the need for a thorough comparative analysis.

Accurate and predictive yield modeling is critical for advanced packaging technologies, as it enables early identification of potential failure mechanisms based on the design and process factors during the development cycle, facilitates system-technology co-optimization, and informs strategies for chiplet interconnect repair [10]. The overall yield in advanced integration schemes is influenced by several components, including the individual yields of chiplets, HB process, and through-silicon vias (TSVs). While system-level yield modeling has received considerable attention [11]–[15], existing models for HB yield are often overly simplified. For instance, [11], [12] propose yield models tailored to 3D stacked ICs, yet [11] omits the bonding process entirely, and [12] treats bonding yield as a fixed constant, which is a simplification also adopted by [14] and [15] in their chiplet system yield analyses. Such approaches fail to capture the physical failure mechanisms intrinsic to HB and therefore a detailed, process-aware yield model.

In addition to the presence of multiple failure mechanisms, the pad layout pattern, encompassing the ratio of different I/O pad types, their spatial distribution, and the redundancy scheme, can significantly influence the actual yield. Depending on the pitch and die size, a single die fabricated using the HB process may contain anywhere from hundreds of thousands to several million I/O pads, including signal pads, power/ground pads, and a large number of dummy pads. The ratio and spatial placement of these types of I/O pads are highly design-dependent. Signal pads are typically critical, as the failure of even a single signal pad can lead to die failure. [16] introduced a redundant TSV grouping technique to enhance the yield of 3D ICs; similarly, introducing redundancy for signal pads can be an effective strategy to mitigate yield loss in HB, particularly in the presence of large-scale clustering void defects. In terms of adding redundancy, it can be roughly categorized as *shared redundancy* and *dedicated redundancy*. A shared redundancy scheme is a fault-tolerance approach where a group of N components share a smaller pool of M spare components [17], [18]. Typically, the value of M is significantly less than N ($M < N$), allowing for a more efficient utilization of spare resources. A dedicated redundancy scheme is also referred to as 1:1 redundancy, where each main component is paired with its own exclusive, dedicated spare component. The interaction

Zhichao Chen and Puneet Gupta are with the Department of Electrical and Computer Engineering, University of California, Los Angeles (UCLA), 90095 Los Angeles, California, United States (e-mail: zhichaochen@ucla.edu; puneetg@ucla.edu).

among the redundancy scheme, the physical distance between main and redundant components, and defect clustering patterns can have various impacts on yield. Power and ground pads, on the other hand, are often replicated and redundantly distributed across the die to maintain power integrity. Dummy pads, which share the same pitch and size as functional I/O pads, play a key role in process optimization [20]–[22], especially for chemical mechanical planarization (CMP). These pads are strategically inserted to ensure uniform pattern density across the die, with surface coverage typically ranging from 40% to 90%, leading to improved surface planarity [23]. Dummy pad failures are non-critical to die functionality. Given the design-dependent impact of failures across different I/O pad types, it is crucial to incorporate pad-layout information into yield analysis.

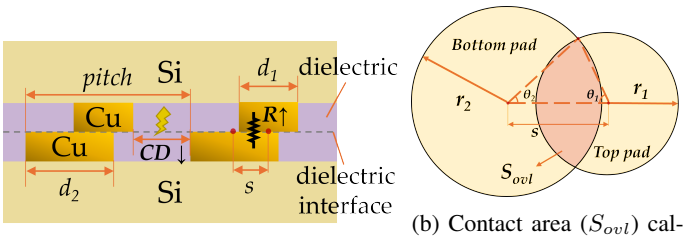
The complexity of the interaction between multiple failure mechanisms and the pad layout patterns in HB makes analytical modeling of yield challenging. In response to this challenge, this work extends our previous work [19], and proposes an enhanced version, YAP+. YAP+ is a physical mechanism-driven and pad-layout-aware near-analytical yield modeling framework. YAP+ introduces a detailed analysis and modeling framework capable of predicting bonding yield for arbitrary pad layouts, which is an important feature absent in YAP. The code of the yield model (YAP+) and simulator is available open-source at `https://github.com/nanocad-lab/YAP`. The key contributions of this work are outlined below:

- To the best of our knowledge, this is the first yield model designed for the HB process. It captures key failure mechanisms that contribute to yield loss, including overlay errors, particle defects, Cu recess variations, dielectric surface roughness, and excessive Cu pattern density.
- We propose a dilation-based method to adaptively calculate the critical area for arbitrary defect shapes and pad layouts with critical, redundant, and dummy I/O pads.
- We develop a yield simulator based on the statistical distributions of various failure mechanisms to validate the proposed yield model and evaluate its predictive accuracy.
- We conduct detailed case studies to analyze how process and design parameters affect yield, including comparisons between W2W and D2W HB. The results highlight the importance of process control for achieving high yield.

The remainder of the paper is organized as follows: Section II discusses the key failure mechanisms associated with HB processes. Section III introduces yield modeling methodologies for W2W HB and extends it to D2W HB. Section IV describes the experimental setup, details the Monte Carlo simulation for multiple failure mechanisms, and compares the simulation outcomes with the near-analytical model. Section V presents case studies analyzing the impact of design and process parameters on yield, including a comparative evaluation of W2W and D2W bonding approaches. Section VI concludes the paper and highlights potential directions for future research.

II. Overview of Failure Mechanisms of Hybrid Bonding

This section presents an overview of the key failure mechanisms inherent to HB processes, including overlay misalignment, Cu recess variation, and particle-induced void defects. Each of these factors can significantly cause yield loss if the bonding process parameters are not properly controlled. A thorough understanding of these mechanisms is essential for optimizing the HB process and improving the bonding yield.



(a) Misalignment leads to resistance increase and dielectric breakdown.

(b) Contact area (S_{ovl}) calculation given the misalignment (s).

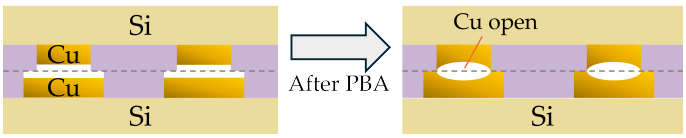
Fig. 1: Failure mechanism of Overlay errors.

A. Overlay Errors

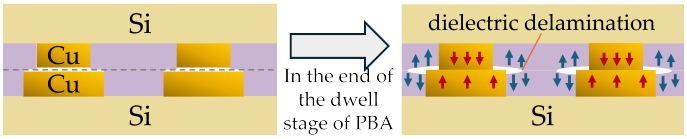
Ensuring the quality of the Cu connections formed during the HB process is essential for maintaining the electrical performance of the overall design. However, misalignment between Cu pillars of the top and bottom wafers is inevitable due to factors such as robot arm calibration errors and wafer warpage induced by thermal stress mismatches. As pad dimensions shrink and bonding pitch achieves the sub-micron scale, the impact of such misalignment on yield becomes increasingly significant [3], [24]. Fig. 1a and Fig. 1b show the front and top views of the bonding connection, respectively. As shown in the figures, the excessive misalignment (s) will decrease the contact area (S_{ovl}) of the Cu interface. This reduction leads to increased contact resistance and elevates the risk of electromigration-induced failures [25]. Additionally, the probability of dielectric breakdown increases as the critical distance (CD) between adjacent Cu pads decreases, resulting in a thinner insulating film between the upper pads and the lower pads of neighboring pillars [6]. Let the bonding pitch be denoted by p, and assume the pads are circular with diameters $d_1 = 2r_1$ for the top pad and $d_2 = 2r_2$ for the bottom pad. The critical distance between two perfectly aligned Cu pillars is defined as $CD = p - d_2$, representing the spacing between adjacent pads. In some designs, the top pad is intentionally made smaller than the bottom pad to enhance misalignment tolerance [26]. YAP+ supports modeling such asymmetric pad dimension configurations, enabling analysis across a wide range of real-world bonding scenarios. Overlay errors are generally categorized into pad-level random misalignment and systematic misalignment. To reduce the likelihood of Cu-to-Cu bonding failure, it is recommended that total misalignment remain within 50% of the bottom pad diameter [3], a constraint that becomes increasingly challenging in fine-pitch designs.

B. Cu Recess Variations

The CMP process will introduces Cu recess effects, often resulting in a concave surface profile on the Cu pad. As illustrated in Fig. 2a, excessive Cu recess can degrade the bonding quality or even incur Cu interconnect failure following post-bond annealing (PBA) [31], [32]. Conversely, Cu protrusion and insufficient Cu recess also negatively impact yield. High wafer surface roughness reduces the effective dielectric contact area during low-temperature bonding, which in turn lowers the density of covalent bonds formed after PBA, weakening both bonding strength and energy per unit area at the



(a) Excessive Cu recess reduces Cu contact area or even cause open.



(b) Insufficient Cu recess may cause dielectric delamination, depending on the dielectric bonding strength, Cu density, and post-bond annealing (PBA) temperature, etc.

Fig. 2: Failure mechanism of Cu recess variations.

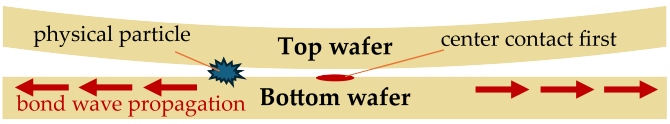
dielectric interface [32]–[34]. In fine-pitch designs, high Cu pattern density combined with insufficient Cu recess can lead to elevated peak peeling stress at the dielectric interface during the final stage of annealing [35]–[38]. As shown in Fig. 2b, if the dielectric interface bonding cannot withstand the peeling stress, dielectric delamination or cracking may occur, resulting in bonding failure [39]–[41]. Therefore, to achieve a high yield, especially for chiplets with a large number of Cu pads, a precise control of Cu recess variation within a range determined by Cu pattern density, surface roughness, etc., is necessary.

C. Particle Defects

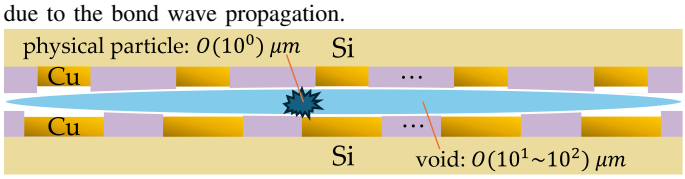
In the HB process, particles are generated during various steps such as wafer dicing, grinding, and polishing [42]. Additionally, any form of friction can produce particles, which is particularly problematic since hybrid bonding involves mechanically picking up dies and placing them onto other chips [4]. Achieving high yield requires stringent cleanliness standards to prevent the presence of physical particles, which can lead to void formation at the bonding interface [5]. Even a particle as small as 1 µm in thickness can cause a void with a diameter reaching hundreds of microns [4], [43]. In addition to physical particles, gas condensation during the bonding process can incur edge voids near the wafer bevel region. However, since the outer edge region is typically removed during the sawing process, dies located away from the wafer perimeter remain unaffected, and thus these voids do not impact overall yield [44]. Consequently, the proposed defect model focuses primarily on yield loss caused by particle-induced void formation. As shown in Fig. 3, during the W2W HB process, initial contact occurs at the center of the top wafer, which then propagates outward toward the edges. Due to bond wave propagation, the presence of a particle at the bonding interface can result in a *main void* accompanied by a trailing *void tail* extending radially [43]. In contrast, in the D2W case, void tail formation is uncommon, primarily due to the smaller die size relative to the wafer and differences in the bonding mechanism.

III. YAP+ Yield Model

In this section, we introduce our yield modeling methodology for W2W HB, which incorporates arbitrary pad layout configurations. The model is then extended to support D2W bonding scenarios. Specifically, since power and ground I/O pads typically have numerous replicas distributed widely across the die, it is highly unlikely that a die failure would occur solely



(a) The presence of a particle will form a main void and a void tail due to the bond wave propagation.



(b) Void formation can fail the dielectric and Cu bonding. A particle of a few microns can form a main void of hundreds of microns [43].

Fig. 3: Failure mechanism of particle defects.

due to the simultaneous failure of all such pads. Therefore, when deriving the yield model below, we do not consider bonding failures of power/ground I/O pads or dummy pads. To validate our derived model, we compare its predictions against simulation results across 300 distinct parameter sets. The simulation and validation workflow is illustrated in Fig. 4, with detailed experimental settings discussed in Section IV.

A. Overlay Model

Our proposed overlay model quantifies the yield loss resulting from the misalignment of Cu pads. We assume the bonding misalignment follows a normal distribution with zero mean and a process-dependent σ_1 [27]. Under this assumption, the possibility of survival (POS) of one single pad can be calculated as follows:

$$POS_{ovl,pad} = \frac{1}{\sigma_1\sqrt{2\pi}} \int_{-\delta}^{\delta} e^{-\frac{u^2}{2\sigma_1^2}} \mathrm{d}u \tag{1}$$

where u represents the random overlay error between the top and bottom pads, and δ denotes the maximum allowable misalignment to ensure the pad's survival. δ is determined based on the contact area constraint and the critical distance constraint. We define s as the systematic overlay error in a Cu connection, which arises from three primary distortion components: *translation*, *rotation*, and *magnification* [28]. Translation and rotation errors primarily stem from limitations in equipment precision, while magnification errors are mainly caused by wafer warpage/bow due to thermal expansion mismatches among different materials [35]. We define the translation errors in x, y directions as T_x, T_y, respectively, and denote the rotation error as α. Bonded wafer warpage typically ranges from a few micrometers to over 100 µm, but can be reduced to ∼10 µm through run-out compensation techniques [30]. Let B denote the warpage of the bonded wafer. Studies have shown that the magnification factor E is linearly correlated with B [29], [30]. Based on this observation, we construct a linear model to characterize E as follows:

$$E = k_{mag} \cdot B \tag{2}$$

where k_{mag} serves as a fitting parameter in the model and is influenced by factors such as the Cu pad depth, Cu pattern density, and the bonding process temperature, etc. [35]. We model the systematic misalignment $\Delta x, \Delta y$ in x, y directions respectively by

$$\begin{cases} \Delta x(x,y) = T_x - \alpha \cdot y + E \cdot x, \\ \Delta y(x,y) = T_y + \alpha \cdot x + E \cdot y. \end{cases} \tag{3}$$

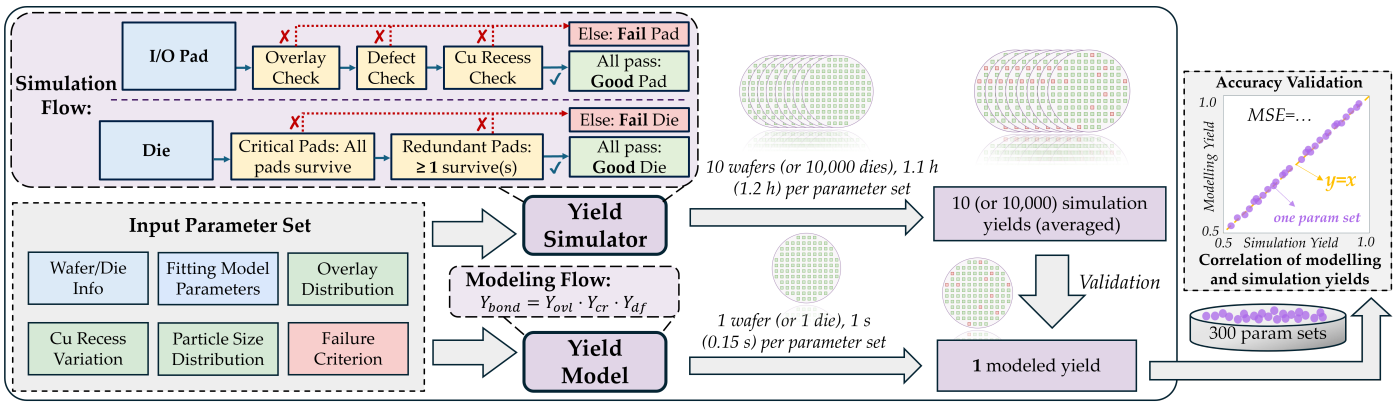


Fig. 4: Simulation workflow and the validation of modeling yield on various input parameter sets.

The systematic overlay error s at the location (x, y) is by

$$s(x,y) = \sqrt{[\Delta x(x,y)]^2 + [\Delta y(x,y)]^2} \tag{4}$$

As Fig. 1b shows, the contact area of two Cu pads can be calculated by

$$S_{ovl} = \begin{cases} \pi r_1^2, & s < r_2 - r_1 \\ \theta_1 r_1^2 + \theta_2 r_2^2 - s r_1 \sin\theta_1, & r_2 - r_1 \le s \le r_1 + r_2. \\ 0, & s > r_1 + r_2 \end{cases} \tag{5}$$

Assuming that pad survival requires the contact area to exceed a threshold defined by k_{ca} times the surface area of the top pad interface, i.e., $S_{ovl} > k_{ca}\pi r_1^2$, and that the critical distance CD must be greater than k_{cd} times the ideal critical distance, i.e., $CD > k_{cd}(p - d_2)$, then δ can be expressed as

$$\begin{aligned} \delta = \min \Bigg\{ & \frac{\theta_1 r_1^2 + \theta_2 r_2^2 - k_{ca}\pi r_1^2}{r_1 \sin\theta_1}, \\ & (1 - k_{cd})p - \frac{1}{2}d_1 + \left(k_{cd} - \frac{1}{2}\right) d_2 \Bigg\} \end{aligned} \tag{6}$$

In practice, the failure region associated with overlay errors usually spans a distance larger than that separating a redundant pad from its replica. Consequently, if misalignment causes one redundant pad to fail, its replica is very likely to fail as well. Therefore, the overall POS for a die is determined by the lowest POS among all interconnection pads, excluding dummy pads and power/ground I/O pads that have numerous, widely distributed replicas. Given above, say that a die has N_{cr} critical and N_{rd} redundant pads, its POS can be written as

$$POS_{ovl,die} = \frac{1}{\sigma_1\sqrt{2\pi}} \min_{i \in [1, N_{cr}+N_{rd}]} \left\{ \int_{-\delta - s_i}^{\delta - s_i} e^{-\frac{u^2}{2\sigma_1^2}} \mathrm{d}u \right\} \tag{7}$$

where s_i denotes the systematic overlay misalignment of the i-th interconnection pad (critical or redundant) on the die. Assuming one wafer has M dies, the overlay yield is by

$$Y_{ovl,W2W} = \frac{1}{M} \sum_{j=1}^{M} POS_{ovl,die,j} \tag{8}$$

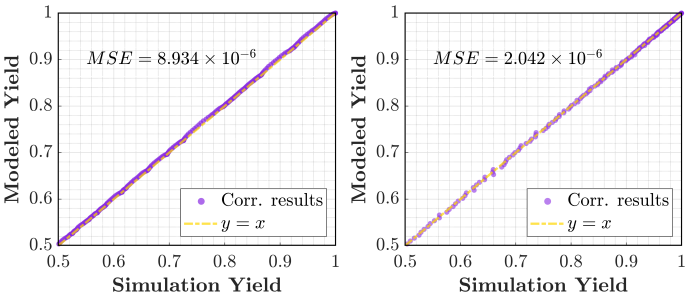
We vary input parameters for both the model and simulator, including translation error, rotation error, warpage, die size, pad layouts, and other relevant factors. The 300 comparison results (purple points) and the mean squared error (MSE) in Fig. 5a indicate that our model aligns closely with the simulation results, thereby validating its reliability and accuracy.

B. Cu Recess Model

We can assume the pad height after the CMP process follows a normal distribution according to [26], [45]. Taking the dielectric surface as the zero reference level, the pad height is considered negative for recessed pads and positive for protruded pads. It is evident that the combined height of the top and bottom pads also follows a normal distribution. Let h denote the sum of heights of two corresponding pads. The mean of this distribution is represented by μ_h, and the variance by σ_h^2. To prevent Cu bonding failure and dielectric delamination, the combined pad height h must be constrained within a safe range (ζ_-, ζ_+). Below we discuss the calculation of ζ_- and ζ_+.

a) Calculation of ζ_-: As observed in [45]–[47], the height variation resulting from Cu expansion during annealing exhibits a linear correlation with the annealing temperature. The *lower bound* ζ_- of the total Cu heights required to form a qualified Cu bonding area is determined by the cumulative Cu expansion after PBA. This ensures that any gap between the pads caused by recesses is sufficiently filled with Cu, thereby preventing bonding failure.

b) Calculation of ζ_+: The *upper bound* ζ_+ represents the critical condition for dielectric delamination, occurring when the combined pad height reaches this threshold. It is important to note that the surface roughness reduces the effective contact area of between the two bonding surfaces, which can exacerbate the risk of delamination. To calculate the normalized effective contact area $A_b^*(\sigma_z, R_z, E_d, w)$, we adopt the asperity-based roughness model proposed by [33], [48]. This model incorporates key roughness and bonding parameters, including the standard deviation of asperity height σ_z, asperity cap radius R_z, Young's modulus of contact surface material E_d,



(a) Overlay model. (b) Cu recess model.

Fig. 5: Correlation results of the overlay model and the Cu recess model with the simulation data for W2W HB.

and the bonding energy under full contact w [1]. The bonding energy of SiO_2 dielectric surface increase because the hydrogen bonds formed before annealing are converted into stronger covalent bonds through dehydration condensation after PBA [39]. The maximum tolerable peeling stress σ_{tol}, beyond which delamination may occur, can be expressed as

$$\sigma_{tol} = A_b^*(\sigma_z, R_z, E_d, w) \times \sqrt{\frac{2E_d w}{t_d}} \tag{9}$$

where t_d represents the thickness of the surface material [50]. As the ambient temperature fluctuates during PBA, the thermal expansion mismatch between the metal and dielectric materials induces various stresses at the bonding interface [41]. Among various interfaces in the bonding structure, the dielectric-dielectric (e.g. SiO_2-SiO_2) is more susceptible to delamination due to its relatively lower bonding strength and the elevated peeling stress observed at the end of the annealing dwell stage [36], [39]. For simplification, we employ a fitting model to evaluate the dependence of dielectric interface peeling stress, based on the asperity and bonding parameters.

$$\sigma_{peel} = k_{peel} \cdot D_{Cu} \cdot (h - h_0) \tag{10}$$

where D_{Cu} represents the Cu pattern density, h_0, k_{peel} are fitting parameters, and k_{peel} is influenced by factors such as annealing temperature, pad shape, pad arrangement, pad structure, etc. [36], [37], [41]. To avoid delamination, one should have

$$\sigma_{tol} \geq \sigma_{peel} \Rightarrow h \leq h_{peel} \tag{11}$$

Additionally, since the Cu protrusion after CMP can lead to delamination, the *upper bound* of the combined pad height is expressed by

$$\zeta_+ = \min\{0, h_{peel}\} \tag{12}$$

To summarize, the POS of this pad during PBA is given by

$$POS_{cr,pad} = \frac{1}{\sqrt{2\pi\sigma_h^2}} \int_{\zeta_-}^{\zeta_+} e^{-\frac{(h-\mu_h)^2}{2\sigma_h^2}} \mathrm{d}h \tag{13}$$

Assume that a die contains N_{cr} critical pads and N_r groups of non-power/ground redundant pads, with each group containing M_r replicas. The POS of the critical pads is given by $POS_{cr,pad}^{N_{cr}}$. The POS of the redundant pads is expressed as $[1-(1-POS_{cr,pad})^{M_r}]^{N_r}$. The die yield, regarding Cu recess variations, is given by the product of these two terms.

$$\begin{aligned} Y_{cr,W2W} &= POS_{cr,die} \\ &= POS_{cr,pad}^{N_{cr}} \cdot \left[1 - (1 - POS_{cr,pad})^{M_r}\right]^{N_r} \end{aligned} \tag{14}$$

We vary the input parameters of Cu recess, pitch, roughness, etc., to validate the Cu recess model. Fig. 5b presents the correlation between the model predictions and simulation results.

C. Defect Model

The relationship between a particle's properties and resulting void size is complex, involving factors such as particle thickness, Young's modulus, and wafer adhesion energy. Due to the significant discrepancies between theoretical predictions and experimental observations in [43], it is more practical to develop a fitting model that estimates the void size based on certain process information. Furthermore, as demonstrated in [7], for particles of specific size and material, both the main void size and the void tail length exhibit linear correlations with the particle's location and the square root of its thickness. To capture these relationships, we adopt simple linear models, with fitting parameters derived from the slope and intercept of the observed trends reported in [43].

[1]Modeling the interaction between two rough surfaces requires normalization of both the surface roughness σ_z and Young's modulus E_d [49].

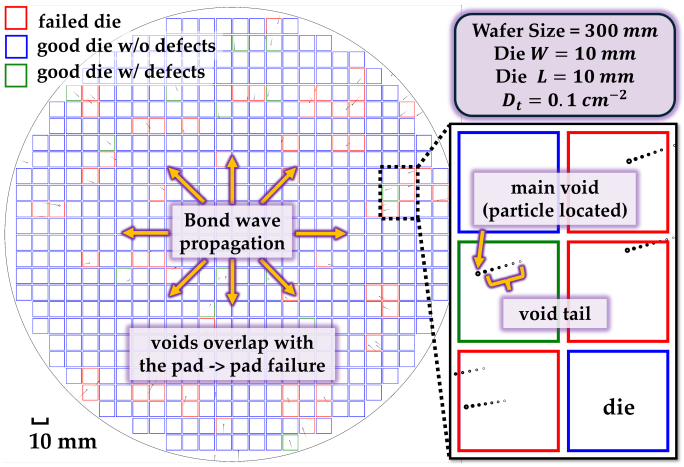


Fig. 6: Visualization of the void formation simulation.

1) Defect Shape Modeling: We model the size r_{mv} of the main void located at a distance L from the wafer center, where $0 \leq L < R$, using the following relationship:

$$r_{mv} = (k_r L + k_{r_0}) t^{1/2} \tag{15}$$

where t denotes the particle thickness. Similarly, the void tail length l can be modeled by

$$l = k_l L t^{1/2} \tag{16}$$

where k_r, k_{r_0}, k_l are fitting parameters. Fig. 6 visualizes the simulated void formation, which closely resembles the scanning acoustic microscopy images of voids reported in [43]. Since the average void tail length on the wafer can reach a few millimeters, more than 10 times the scale of the main void size (typically a few hundred μm) in W2W HB, the defect geometry can be reasonably simplified as a straight line characterized by its length l and outward orientation θ. Furthermore, a die is considered to have failed if the void tail overlaps the functional pad array area, since the void size is typically much larger than the HB pitch ($\leq 10\ \mu$m).

We assume the thickness distribution of particle defects as $D(t)$. A typical form of $D(t)$ can be modeled as [52]

$$D(t) = D_t \cdot \frac{(z-1) \cdot t_0^{z-1}}{t^z}, \qquad t > t_0 \tag{17}$$

where t_0 denotes the minimum particle thickness, and D_t represents the total particles count per unit area across all thicknesses. The parameter z controls the shape of the distribution and empirically ranges between 2 and 3 [53], [54]. The parameters in this distribution are obtained by fitting the model to the data of cleanroom concentration of particles presented in [4]. By Eq. 16, 17, the distribution of void tail length can be calculated by

$$f_l(l) = \begin{cases} \dfrac{2D_t(z-1)l}{z k_l^2 R^2 t_0}, & l \leq k_l R t_0^{1/2} \\ \dfrac{2D_t(z-1)(k_l^2 R^2 t_0)^{z-1}}{z l^{2z-1}}, & l > k_l R t_0^{1/2} \end{cases} \tag{18}$$

where R represents the wafer radius. The comparison of the

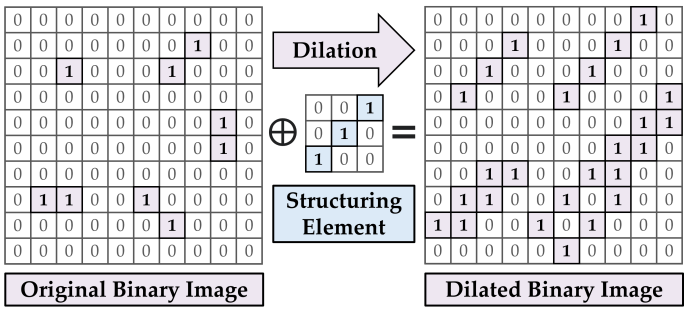


Fig. 7: An example of a dilation operation on a binary image using a structuring element. The structuring element defines the shape and extent of expansion applied to the original binary image. In this context, dilation simulates how a defect (void) affects nearby pads during hybrid bonding.

derived $f_l(l)$ and the simulated distribution is shown in Fig. 9a, confirming the precision of the derivation.

2) Dilation-based Critical Area Calculation: Die failure can result from the failure of a critical pad or from the simultaneous failure of a redundant pad and all its replicas. Based on this relationship, the critical area of the die is defined as the region where a defect would trigger either of these failure mechanisms. When modeling designs with specific pad layouts, including the location and number of different types of I/O pads, such as critical, redundant, and dummy pads, deriving such analytical expressions of critical area for random layouts is not straightforward. To cope with the challenge, we propose a *dilation-based* method for the critical area calculation. *Dilation*, a fundamental morphological operation, expands the boundaries of geometric features through a predefined structuring element. Intuitively, it can be understood as a process that 'grows' or 'thickens' objects in a binary image [56]. An example of a dilation operation on a binary image is shown in Fig. 7. This concept can be effectively applied to model the potential impact area of defects on a die layout, regardless of the shape of either the defect or the layout.

The process of dilation-based critical area calculation is illustrated in Fig. 8. To enable efficient computation, the die is divided into a grid of pad blocks, with each pad block containing only one type of I/O pad. Each pad block is treated as an atomic unit in the yield modeling, meaning that all pads within a block either survive or fail together. Using a finer gridding resolution enables a more accurate representation of the layout, but at the cost of increased computational time both in modeling and simulation, as it leads to larger bitmap dimensions and consequently longer dilation operations. To ensure acceptable accuracy, the gridding resolution must be carefully selected. This choice is typically guided by the size of the defects under consideration. The relationship between gridding resolution and defect dimensions will be discussed in detail in Section IV. Note that redundant pads and their corresponding replicas are not necessarily located within the same pad block. Based on this gridding, we generate a pad block bitmap for each type of I/O pad. Similarly, a void tail defect is also represented as a bitmap by using Bresenham's line algorithm [55], constructed according to its length and orientation. For critical pad blocks, the critical area bitmap $\mathbf{B}_{cr-dilate}$ is obtained by dilating the critical pad block bitmap $\mathbf{B}_{cr}$ using the defect bitmap $\mathbf{B}_{df}(l,\theta)$, as defined by

$$\mathbf{B}_{cr-dilate}(l,\theta) = \mathbf{B}_{cr} \oplus \mathbf{B}_{df}(l,\theta) \tag{19}$$

where $\oplus$ is the dilation operator, and $\mathbf{B}_{df}(l,\theta)$ is the approximated bit map of a void tail defect of length l and orientation θ. For redundant pad blocks, suppose a die contains N_{blk} groups of redundant pad blocks, within each group consists of one main pad block and its corresponding replicas. The HB process typically requires Class 1 / ISO 3 cleanrooms and equipment or better, which results in a relatively low particle defect density ($\sim$0.1/cm^2 [4]). The probability of multiple particle defects occurring in close proximity and independently disabling both redundant replicas is extremely low, particularly given that these replicas are typically placed near one another. As a result, the critical area for each group of redundant pads can be reasonably approximated as the region within which a single defect is sufficient to cause the failure of all replicas. Assuming that the i-th group contains M_r replicas, the critical area bitmap of this group $\mathbf{B}_{rd-dilate,i}$ can be approximated by the intersection of the dilated bitmaps of all redundant replica pad blocks $\mathbf{B}_{rd,i,1} \sim \mathbf{B}_{rd,i,M_r}$, where each is individually dilated using the defect bitmap. Formally, this is expressed as

$$\mathbf{B}_{rd-dilate,i}(l,\theta) = \bigcap_{j=1}^{M_r} (\mathbf{B}_{rd,i,j} \oplus \mathbf{B}_{df}(l,\theta)) \tag{20}$$

Finally, the critical area of the die, $A(l,\theta)$, regarding a void tail defect of length l and orientation θ, is defined as the area of the union of the two bitmaps $\mathbf{B}_{cr-dilate}$ and $\mathbf{B}_{rd-dilate}$. Here, $\mathbf{B}_{rd-dilate}$ represents the union of all group-wise redundant pad block bitmaps after dilation, i.e., $\mathbf{B}_{rd-dilate,1} \sim \mathbf{B}_{rd-dilate,N_{blk}}$. This can be expressed as

$$\begin{aligned} A(l,\theta) &\approx Area(\mathbf{B}_{cr-dilate}(l,\theta) \cup \mathbf{B}_{rd-dilate}(l,\theta)) \\ &= Area(\mathbf{B}_{cr-dilate}(l,\theta) \cup (\bigcup_{i=1}^{N_{blk}} \mathbf{B}_{rd-dilate,i}(l,\theta))) \end{aligned} \tag{21}$$

where $Area(\cdot)$ denotes the area of the region represented by the bitmap. Hence, the average number of particle-induced void tail defects that will cause a die to fail, Λ, can be expressed as

$$\Lambda = \int_0^{\infty} \int_0^{2\pi} A(l,\theta) f_l(l) \mathrm{d}\theta \mathrm{d}l \tag{22}$$

Using the Poisson yield model [18], the yield with respect to the particle-induced void formation is given by

$$Y_{df,W2W} = \exp(-\Lambda) \tag{23}$$

Compared to previous analytical computation, the dilation-based modeling method is more computationally expensive due to the repeated processing of binary images, although the use of the gridding strategy helps mitigate this overhead. However, for a given pad layout, the critical area only needs to be computed once. The result can be stored in a look-up table and retrieved as needed when varying the process parameters, thereby mitigating the drawback of increased computation time. To summarize, compared to traditional analytical models, the dilation-based approach provides improved geometric accuracy for arbitrary layouts and defect shapes, along with better scalability for large and complex designs, while its computational overhead has minimal impact during the application phase.

We vary the input parameters of particle defect density, die size, wafer size, etc., to validate the defect model. Fig. 9b demonstrates the correlation of the defect yield with the

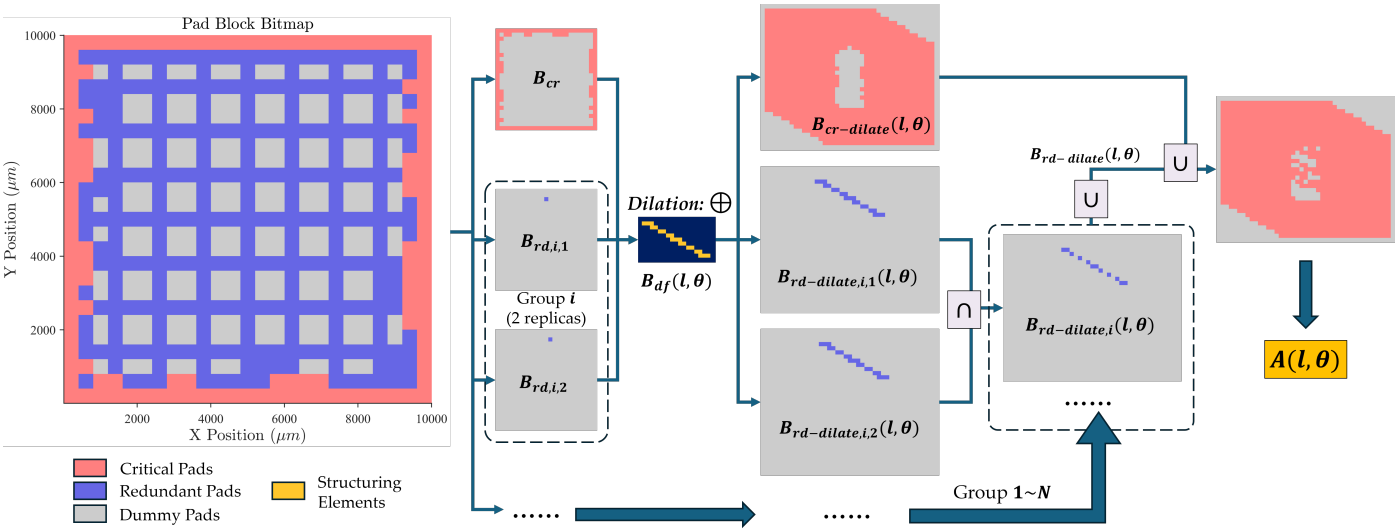


Fig. 8: Dilation-based critical area calculation diagram. In this example, the die size is 10 mm×10 mm, and the gridding resolution in modeling is 400 μm×400 μm. The die consists of 20% critical pads, 50% redundant pads, and 30% dummy pads. The structuring element is generated based on the length l and the orientation θ of a void tail and the gridding resolution.

simulation results.

D. Overall Bonding Yield Model

To develop the overall bonding yield model for W2W hybrid bonding, we assume that the overlay error, Cu recess variations, and particle defects affect die yield independently. By integrating the individual yield components, namely Eq. 8, 14, 23, the assembly yield is

$$Y_{W2W} = Y_{ovl,W2W} \cdot Y_{cr,W2W} \cdot Y_{df,W2W} \tag{24}$$

E. D2W Hybrid Bonding Yield Models

We extend the yield model to the D2W HB scenario. It is assumed that the Cu expansion behavior during PBA, as observed in W2W bonding, remains applicable to D2W HB. However, the yield components associated with overlay error and particle defects must be revised to reflect the distinct characteristics of the D2W bonding process.

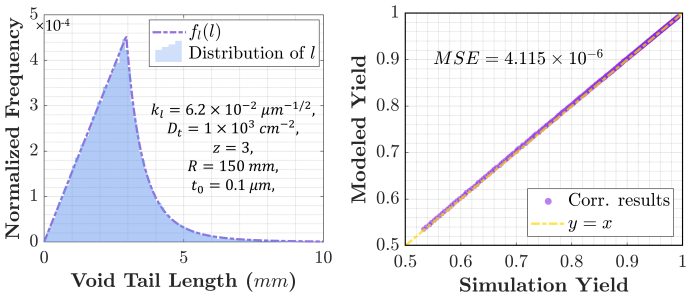
1) Overlay Model: In D2W hybrid bonding, systematic overlay errors occur independently for each die. Due to the smaller die size, an identical marker misalignment at the die edge leads to larger rotation α and magnification E errors compared to the W2W case. Similar to W2W HB, the overall POS for a die in D2W bonding is determined by the minimum POS among all interconnection pads, excluding dummy pads and power/ground I/O pads. For a die with N_{cr} critical and N_{rd} redundant pads, its overlay yield can be written as

$$Y_{ovl,D2W} = \frac{1}{\sigma_1\sqrt{2\pi}} \min_{i\in[1,N_{cr}+N_{rd}]} \left\{ \int_{-\delta-s_i}^{\delta-s_i} e^{-\frac{u^2}{2\sigma_1^2}} \mathrm{d}u \right\} \tag{25}$$

2) Defect Model: Given the smaller die scale in D2W HB, void tail formation is unlikely to occur. Thus, the D2W defect model considers only main void-induced failures. By combining Eq. 15, 17, the probability density function (PDF) of the main void size r_{mv} can be given by

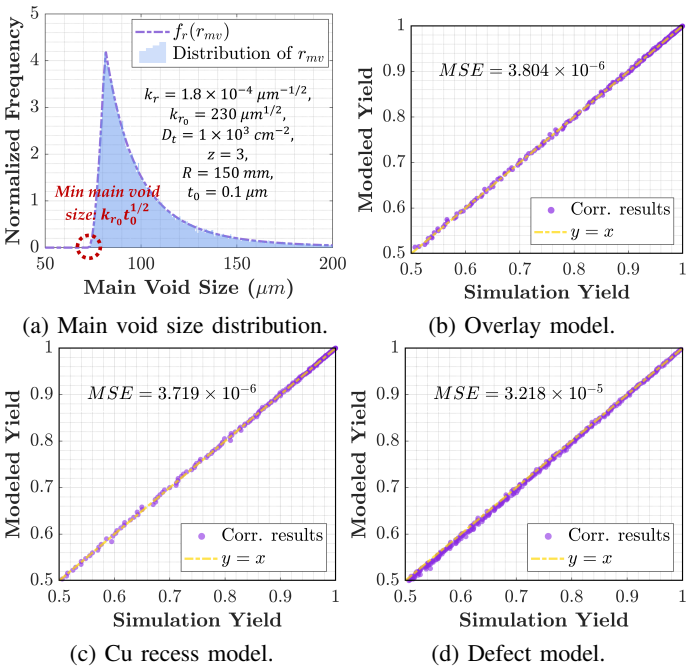
$$f_r(r_{mv}) = \begin{cases} \frac{D_t(z-1)t_0^{z-1}}{k_r^2R^2} \times [\frac{2r_{mv}}{zt_0^z} + \frac{2k_{r_0}^{2z}}{z(2z-1)r_{mv}^{2z-1}} \\ - \frac{2k_{r_0}}{(z-\frac{1}{2})t_0^{z-\frac{1}{2}}}], \quad k_{r_0}t_0^{1/2} < r < (k_rR+k_{r_0})t_0^{1/2}, \\ \frac{2D_t(z-1)t_0^{z-1}(k_rR+k_{r_0})^{2z-2}}{r_{mv}^{2z-1}} - \\ \frac{2D_t(z-1)^2t_0^{z-1}}{k_r^2R^2r_{mv}^{2z-1}} \times [\frac{(k_rR+k_{r_0})^{2z}-k_{r_0}^{2z}}{z} \\ - \frac{2k_{r_0}(k_rR+k_{r_0})^{2z-1}-2k_{r_0}^{2z}}{z-\frac{1}{2}} + \\ \frac{k_{r_0}^2(k_rR+k_{r_0})^{2z-2}-k_{r_0}^{2z}}{z-1}], \quad r \geq (k_rR+k_{r_0})t_0^{1/2}. \end{cases} \tag{26}$$

where R is the effective radius of the die, i.e., $R = (ab/\pi)^{1/2}$, aiming to remain the average number of particles on the die. The close alignment between $f_r(r_{mv})$ and the simulated distribution is shown in Fig. 10a. Similarly, the dilation-based method is applied for critical area calculation. The main void is pixelized, and the structuring element is constructed based on its size. The critical area of the die, $A(r_{mv})$, regarding a main void defect of radius r_{mv}, is defined as the area of the union of the two bitmaps $\mathbf{B}_{cr-dilate}$ and $\mathbf{B}_{rd-dilate}$. This can



(a) Void tail length distribution. (b) Defect model.

Fig. 9: Correlation results of the void tail length distribution and the defect model with the simulation data for W2W HB.



(a) Main void size distribution. (b) Overlay model.

(c) Cu recess model. (d) Defect model.

Fig. 10: Correlation results of main void size distribution and the yield model with the simulation data for D2W HB.

be expressed as

$$A(r_{mv}) \approx Area(\mathbf{B}_{cr-dilate}(r_{mv}) \cup \mathbf{B}_{rd-dilate}(r_{mv})) \quad (27)$$

By Eq. 26, 27, the average number Λ of particle-induced main void defects that will fail a die is given by

$$\Lambda = \int_{k_{r_0} t_0^{1/2}}^{\infty} A(r_{mv}) f_r(r_{mv}) \mathrm{d}r_{mv} \quad (28)$$

Using the Poisson yield model [18], the yield with respect to the particle-induced void formation for D2W HB is given by

$$Y_{df,D2W} = \exp(-\Lambda) \quad (29)$$

3) Overall Bonding Yield Model: Similarly, we assume the overlay error, Cu recess variations, and particle defects have independent impacts on the die yield for D2W HB. Fig. 10b, 10c, 10d show the correlation results of three yield terms, respectively. By combining Eq. 14, 25, 29, the bonding yield is by

$$Y_{D2W} = Y_{ovl,D2W} \cdot Y_{cr,D2W} \cdot Y_{df,D2W} \quad (30)$$

IV. YAP+ Simulator and Model Validation

To validate the derived model, a Monte Carlo simulator is developed for Cu-SiO_2 HB process. Key inputs include the wafer/die information, overlay distribution, Cu recess variations, particle size distribution, model fitting parameters, and failure criteria. The similarity of failure mechanisms reported in the Cu-SiCN hybrid bonding process further supports the general applicability of the YAP+ modeling framework [24], [26], [51]. With appropriate configuration of the input parameters, the model can be readily adapted to various bonding scenarios. Table I presents the baseline model parameters, with additional details available in our code. These are the parameter values used in our experiments unless otherwise stated. Fig. 4 outlines the simulation workflow.

TABLE I: Baseline Parameters in Yield Modeling and Simulation

Design Parameters	**Value**
Pad pitch [41]	1 µm
Bottom/Top pad size [41]	0.5 µm, 0.3 µm
Die size [4]	10 mm×10 mm
Wafer size	300 mm
Process Parameters	**Value**
Random misalignment [3]	0 nm (20 nm)*
System x, y translation [3]	0 nm (20 nm)*
System rotation [3]	0.05 µrad (0.01 µrad)*
System magnification [3]	0.05 ppm (0.01 ppm)*
Particle defect density [4]	0.1 cm^{-2}
Minimum particle thickness [4]	0.1 µm
Shaping factor z in Eq. 17 [53], [54]	3
Top/Bottom pad recess [26], [36]	10 nm (1 nm)*
Roughness σ_z [34], [58]	1 nm
Bonding energy w (SiO_2–SiO_2) [32], [39]	1.2 J/m^2
Young's modulus (SiO_2) [36], [41]	73 GPa
Dielectric thickness [57]	1.5 µm
Model Parameters	**Value**
Contact area constraint k_{ca} in Eq. 6 [3]	0.5
Critical distance constraint k_{cd} in Eq. 6 [3]	0.5
k_{mag} in Eq. 2 [30]	0.09 m^{-1}
k_{peel} in Eq. 10 [36]	6.55×10^{15} N · m^{-3}
h_0 in Eq. 10 [36]	75 nm
k_r in Eq. 15 [43]	1.8×10^{-4} $\mu m^{-1/2}$
k_{r0} in Eq. 15 [43]	230 $\mu m^{1/2}$
k_l in Eq. 16 [43]	6.2×10^{-2} $\mu m^{-1/2}$
k_n in Eq. 31 [43]	9×10^{-5} $\mu m^{-3/2}$
k_S in Eq. 32 [43]	2.7 $\mu m^{1/2}$
W2W gridding resolution†	400 µm×400 µm
D2W gridding resolution†	100 µm×100 µm

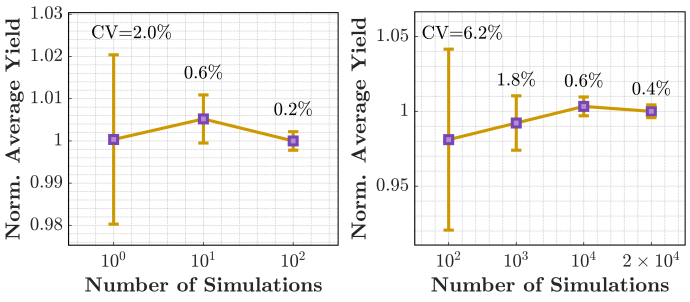
* Mean (Standard Deviation)
† Gridding resolution is used exclusively in the analytical modeling and does not affect the simulation results.

A. Overlay Check

Regarding overlay errors, we assume the random misalignment and three distortion components (translation, rotation, and magnification errors) follow respective normal distributions. In each input parameter set, multiple combinations of those values are sampled from the distributions and are used to calculate the overall misalignment. In W2W simulations, the parameters are sampled per wafer, whereas in D2W simulations, they are sampled per die. The Cu connection fails if the overall misalignment exceeds the maximum allowed overlay error δ.

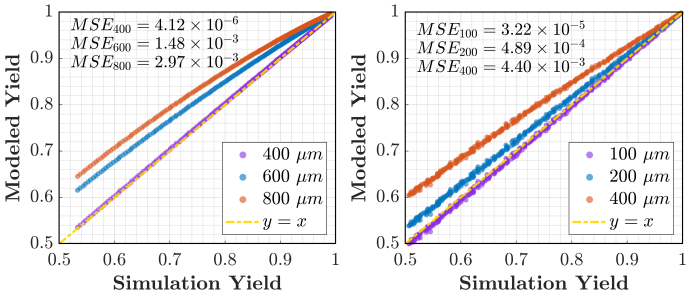
B. Defect Check

We initially assign thicknesses to the particles by randomly sampling the thickness distribution $D(t)$ given in Eq. 17. Although the number of particle defects may vary between individual wafers/dies, their occurrence across the wafer/die population follows the particle defect density D_t. Then, the particles are randomly and uniformly located across the wafer/die, and the void tails are generated based on the linear fitting model from [43], simulating the bond wave propagation (Fig. 6). The main void size and the void tail length are given by Eq. 15 and Eq. 16, respectively. To accurately simulate the defect



(a) W2W HB. (b) D2W HB.

Fig. 11: Simulation effort analysis for reliable yield results.



(a) Defect yield of W2W HB. (b) Defect yield of D2W HB.

Fig. 12: Correlation between 3 gridding-resolution-based Y_{df} modeling results and 1 reference simulation result. The gridding resolution in modeling does not affect the simulation results.

morphology, it is necessary to account for both the number and size distribution of voids in the void tail. The number of voids in the void tail is given by

$$n = k_n L t^{1/2} \tag{31}$$

The total area of the void tail is given by

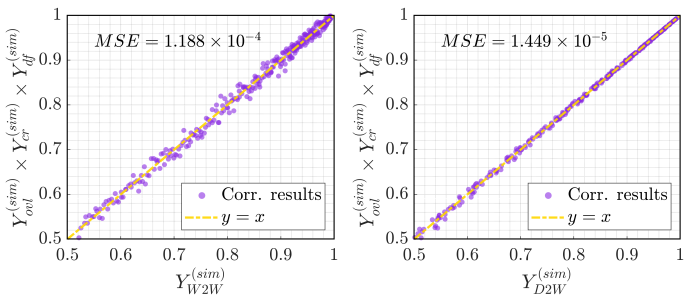
$$S = k_S L t^{1/2} \tag{32}$$

Within the void tail, the void size decreases linearly as its position shifts farther toward the wafer edge. The Cu connection fails if there is any void overlapping with the top pad.

C. Cu Recess Check

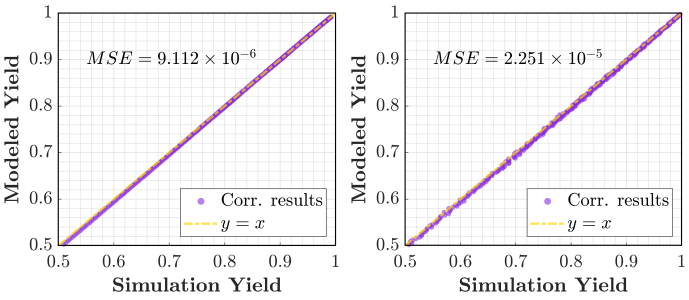
The Cu recess values of the top pad and bottom pad are sampled from their respective normal distributions. The peeling stress σ_{peel} during PBA and the gap between the two Cu pads after PBA are calculated. The Cu connection fails if: (1) the peeling stress is higher than the tolerance value, or (2) the gap still exists after PBA.

To summarize, a Cu connection survives only if it passes the *Overlay Check*, *Defect Check*, and *Cu Recess Check*. A die is considered to have survived if (1) all critical pads remain functional, and (2) in each group of redundant pads, at least one pad survives. The simulation results are closer to the actual conditions with less approximation compared to the model. However, to achieve accurate yield predictions across all failure mechanisms using baseline inputs, repeated simulations are necessary to obtain a statistically reliable mean yield. As illustrated in Fig. 11, we perform multiple simulations for a single input parameter set and compute the average yield. This process is repeated 10 times to observe the variability of the results. The coefficient of variation (CV), defined as the ratio of the standard deviation to the mean, is annotated



(a) Yield of W2W HB. (b) Yield of D2W HB.

Fig. 13: Yield impact of failure mechanism interactions.



(a) Yield of W2W HB. (b) Yield of D2W HB.

Fig. 14: Correlation of bonding yield with the simulation.

for each simulation count to quantify the relative dispersion. We determine the minimum number of simulations required to reduce the CV below 1% for both the W2W and D2W cases, and use this simulation setting to validate the YAP+ model. It requires 10 wafer samples (10,000 die samples) for W2W (D2W) HB simulation, taking 0.9 hours (2.0 hours) on a single CPU (AMD Ryzen 9 8945HS). Additionally, since dilation accounts for a significant portion of the model's runtime, our model validation examines the impact of a key factor in this process, the gridding resolution, on model accuracy. Gridding resolution plays a crucial role in determining the accuracy of defect yield estimation. As shown in Fig. 12, the acceptable gridding resolutions for W2W and D2W hybrid bonding are 400 µm×400 µm and 100 µm×100 µm, respectively. These settings are used as the baseline configurations in the subsequent experiments. A coarser grid reduces model accuracy by inadequately representing the pad layout and defect geometry, which compromises the resolution of the critical area, resulting in inaccurate defect yield estimation. Conversely, using an excessively fine grid offers only marginal accuracy improvement while significantly increasing the dilation runtime, as shown in Table II.

For tractability, we assume that different failure mechanisms occur independently. In practice, multi-physics coupling, such as reduced dielectric contact area from overlay error or void formation leading to stress-induced delamination, introduces some correlation, which is captured in our simulator. The interaction is captured in the simulator. To assess its impact at the yield level, we simulate each mechanism independently and quantify its yield as the fraction of surviving die samples. We compare the product of their individual yields with the overall bonding yield. As shown in Fig. 13, the deviation is minimal, indicating that the mechanisms can be treated as approximately independent, especially in high-yield cases. As yield decreases, these correlations become more pronounced,

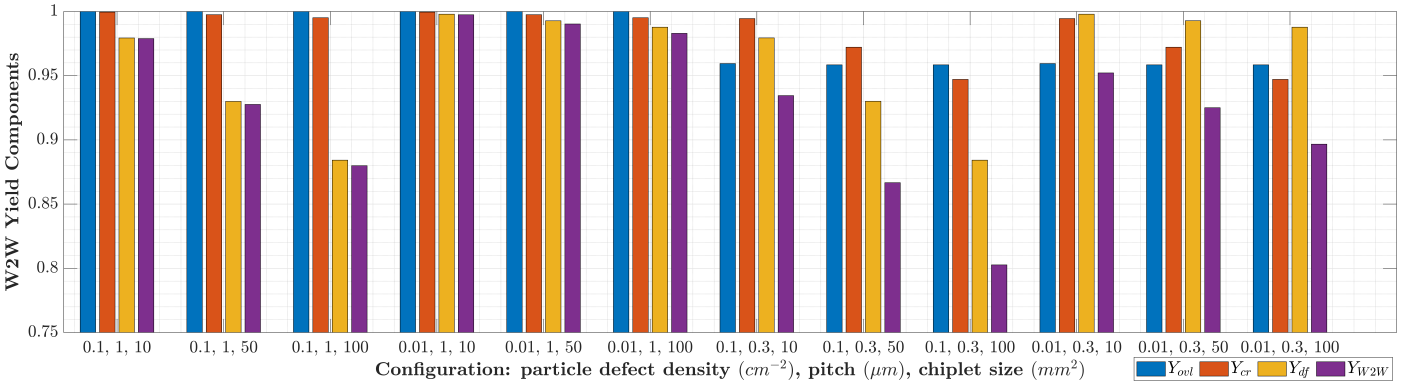


Fig. 15: W2W case studies for various configurations.

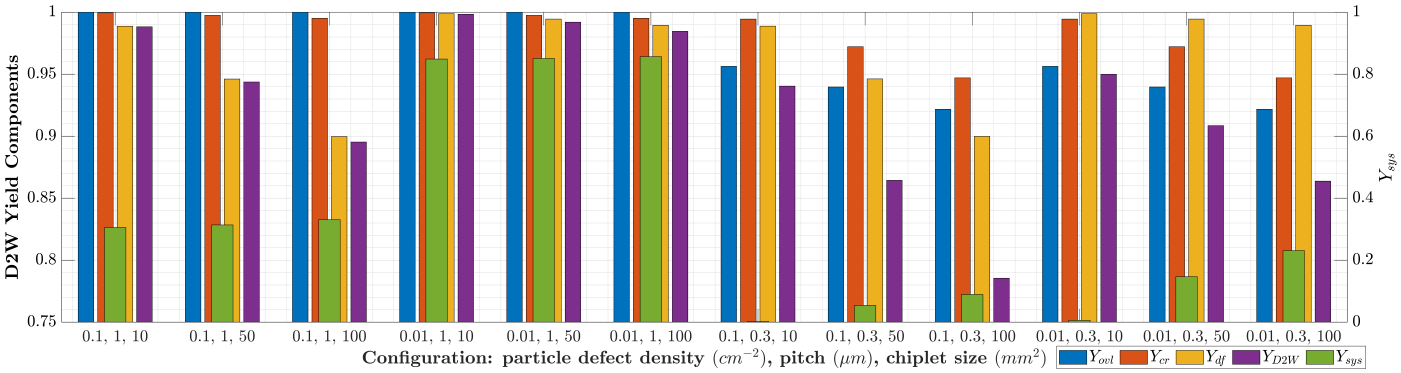


Fig. 16: D2W case studies for various configurations.

TABLE II: Runtime of the defect yield across different gridding resolutions (GR) for W2W and D2W HB

W2W		D2W	
GR (µm)	Y_{df} runtime (s)	GR (µm)	Y_{df} runtime (s)
800	0.10	400	0.25
600	0.11	200	1.17
400	0.14	100	7.76
200	0.38	50	65.97

making it increasingly necessary to incorporate stronger multi-physics interactions into both modeling and simulation to maintain accuracy.

Based on this observation, we next compare the overall modeling yield with the simulation yield. Under the baseline configuration in Table I, the yield model achieves virtually identical accuracy compared with simulation results, as shown in Fig. 14 in 3.7 s (10.2 s) for W2W (D2W) HB, offering over *870x (720x) runtime improvement*. If the computational overhead of the dilation-based critical area calculation in the defect model is amortized through the use of a look-up table, the runtime can be further reduced to 2.9 s (2.3 s) and can achieve over *1,100x (3,200x) runtime improvement* for W2W (D2W) HB. The high modeling efficiency enables the usage in yield optimization and pathfinding optimization loops.

V. Experimental Results

After validating the model with simulation results, we use YAP+ to conduct case studies that demonstrate the impact of various process and design factors on bonding yield, indicating its strengths in system-technology co-optimization. We vary particle defect density (0.01/cm^2, 0.1/cm^2), pitch (0.3 µm, 1 µm), and chiplet sizes (10 mm^2, 50 mm^2, 100 mm^2) in the modeling. A pad layout composed exclusively of critical pads is adopted in this experiment. The yield breakdown and the overall bonding yield are reported in Fig.15 (W2W setup) and Fig.16 (D2W setup).

A. The Impact of Particle Defect Density

The HB process requires the strict removal of particles at the bonding interface. Fig. 15, 16 show that under a relaxed bonding pitch (1 µm), bonding yield is notably affected by defect-related failures. W2W HB exhibits higher sensitivity to particle contamination due to void tail formation during bond wave propagation, resulting in a larger critical area per die. The results indicate that a 10x improvement in defect density (ISO 2) enables near-perfect (∼99 %) defect yield for both W2W and D2W across all chiplet sizes.

B. Impact of Bonding Pitch

In this case study, the bottom pad size is set to half the corresponding pitch. As shown in Fig.15 and Fig.16, reducing the pitch from 1 µm to 0.3 µm leads to a noticeable drop in yield across various chiplet sizes, with the effect more pronounced in D2W HB. Smaller pitches increase sensitivity to Cu pillar misalignment, demanding tighter overlay control. Currently, both W2W and D2W HB technologies are capable of achieving 50 nm overlay accuracy [2], [3]. However, in D2W HB, the overlay yield (Y_{ovl}) at a 0.3 µm pitch gradually degrades as the chiplet size increases. In contrast, W2W HB maintains a relatively stable overlay yield across all chiplet sizes. This distinction, given the comparable alignment accuracy, arises from how overlay error affects yield in these two bonding schemes. In D2W HB, if the alignment error at the chiplet edge

exceeds the failure threshold, the die is discarded. However, in W2W HB, chiplets located near the wafer center are more likely to survive even when edge alignment reaches the failure limit, thus resulting in higher Y_{ovl}.

Reducing the bonding pitch significantly increases the number of I/O pads, heightening sensitivity to Cu recess variations. As shown, the yield loss in W2W HB at smaller pitches is mainly due to reduced Y_{ovl} for smaller chiplets and reduced Y_{cr} for larger chiplets. A practical strategy to mitigate the impact of increasing I/O pad counts is to introduce redundancy to critical pads. The defect yield, on the other hand, remains largely unaffected, as the void sizes exceed the pitch, keeping the critical area roughly constant.

C. Analyzing Yield Limiters with Varying Chiplet Sizes

Bonding yield decreases with increasing chiplet size for both D2W and W2W bonding, primarily due to greater Cu recess variation (from more I/O pads per die) and heightened defect sensitivity.

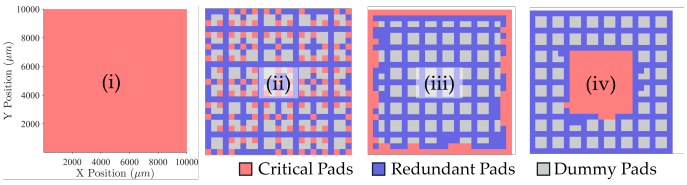
D2W hybrid bonding can be applied in 2.5D integration to assemble large chiplet systems. As such, evaluating yield based solely on a single chiplet can be misleading. Although full system-level yield modeling is beyond the scope of this work, we include a system yield (Y_{sys}, shown in Fig. 16), calculated as $Y_{D2W}^{\#chiplets}$, assuming no chiplet redundancy and a nominal system size of 1000 mm^2. Note that Y_{sys} is plotted against the right vertical axis. This approach reflects the cumulative probability of successful bonding across all chiplets in the system. Increasing chiplet size reduces the total number of chiplets required, helping to mitigate the compounding effect of Y_{D2W} degradation on the Y_{sys}. Interestingly, even though Y_{D2W} decreases with increasing chiplet size, the system yield Y_{sys} remains slightly higher. [2] Overall, building large chiplet-based systems using 2.5D integration necessitates tighter process control, particularly in overlay alignment and Cu recess variations.

D. The Impact of the I/O Pad Layout

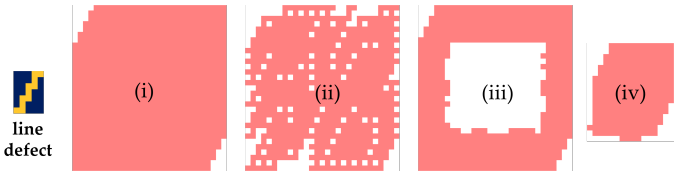
The layout of I/O pads significantly affects bonding yield in both W2W and D2W HB, as the failure of different types of I/O pads has different contribution mechanisms to the yield loss. To investigate this, four distinct I/O pad spatial distribution patterns are studied:

(i) *Full*: all pads are critical, evenly distributed across the die area;
(ii) *Sparse*: critical pads are spread sparsely with additional redundant pads and dummy pads;
(iii) *Peripheral*: critical pads are arranged around the edges with redundant and dummy pads inside.
(iv) *Centralized*: critical pads are concentrated at the center, surrounded by redundant and dummy pads.

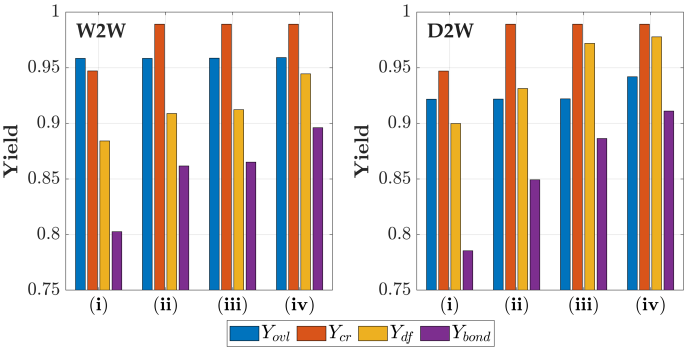
Fig. 17a visualizes these four patterns. These I/O pad layouts are configured at 0.3 µm pitch to highlight their impact on Y_{ovl} and Y_{cr}, as yield loss due to overlay errors and Cu recess variations is more pronounced at finer pitches. *Full* layout consists of 100% critical pads, while the remaining layouts



(a) I/O pad layouts: (i) *Full*; (ii) *Sparse*; (iii) *Peripheral*; (iv) *Centralized*.



(b) Comparison of critical areas in W2W HB across different pad layouts, evaluated with a void tail defect of 3 mm in length and oriented at 30°.



(c) Yield breakdown of W2W and D2W HB across I/O pad layouts.

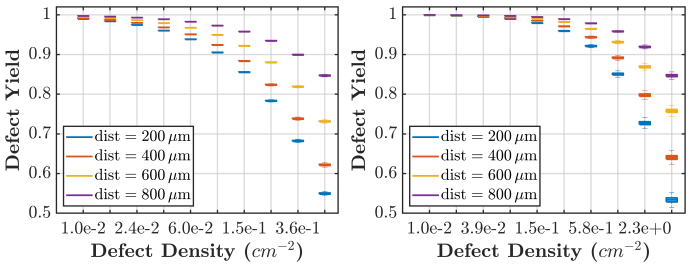
Fig. 17: Impact of I/O pad layouts on W2W and D2W bonding yield at 0.3 µm pitch.

each contain 20% critical pads, 50% redundant pads, and 30% dummy pads. *Peripheral* layout is widely adopted in 2.5D integration, where die-to-die interconnections require placing I/O pads at the die periphery to simplify routing. *Centralized* layout partially resembles that of 3D-stacked memory systems like HBM, where a horizontally distributed central TSV region delivers signals and power vertically and is surrounded by multiple memory channels [9]. For D2W HB, a pad block size of 100 µm×100 µm is adopted consistently across all layouts.

Fig. 17b depicts the critical areas of these pad layouts, evaluated with a void tail defect in W2W HB, under the same scale for direct comparison. Fig. 17c shows the yield breakdown across four layouts. Overall, compared to *Full* layout, the other three have higher Cu recess yield Y_{cr} and defect yield Y_{df}. This improvement arises from the reduced number of critical pads, which lowers the cumulative impact of individual pad failures due to Cu recess variations and slightly decreases the critical area vulnerable to particle-induced defects. Within three layouts with less critical pads, *Sparse* layout results in relatively low Y_{df} improvement, as its critical pads are more spatially isolated. As the pattern (ii) shown in Fig. 17b, this reduces the overlap among the critical areas of adjacent pad blocks, leading to a larger cumulative critical area. In contrast, *Peripheral* and *Centralized* layouts feature more clustered critical pads, where overlapping of their critical areas effectively 'offsets' the total exposed critical region, thus enhancing defect tolerance.

In the W2W case, *Centralized* layout achieves over 3% improvement in Y_{df} compared to *Peripheral* layout. Meanwhile,

[2] Note that this does not account for any worsened yield of a larger chiplet. A more complete system yield model can be found, for example, in [14], albeit with an oversimplified bonding yield model.



(a) Die size: 10 mm×10 mm. (b) Die size: 3.2 mm×3.2 mm.

Fig. 18: Impact of the main-replica spacing on the defect yield of W2W HB across 1,000 randomly generated pad layouts.

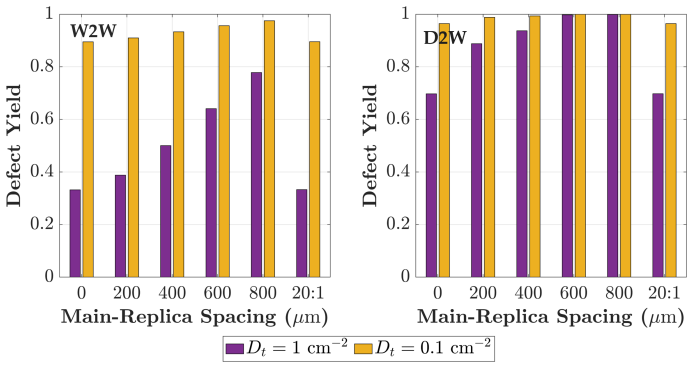


Fig. 19: Defect yield of W2W and D2W HB under various redundant replica configurations.

for D2W, the two layouts exhibit similar Y_{df} values, with a difference of less than 0.6%. This trend can be attributed to the differing defect morphologies in W2W and D2W bonding. In the W2W case, the dominant defect type is the void tail, which has a relatively large spatial footprint. As a result, the extended dimension of void tails occupies a significant portion of the critical area in the internal cavity in *Peripheral* layout, as the pattern (iii) in Fig. 17b, limiting the effectiveness of critical area reduction. In contrast, *Centralized* layout benefits from the tight clustering of critical pad blocks, which maximizes the overlap among their associated critical areas. This overlap acts as an effective compression or offset mechanism, substantially reducing the net defect-sensitive region and leading to higher Y_{df}. In the D2W case, the dominant defect is the main void, which is more localized and has a smaller spatial dimension. Under this condition, both *Peripheral* and *Centralized* layouts achieve comparable levels of critical area overlap, resulting in Y_{df} values. However, to mitigate the impact of debris contamination from edge chipping during die dicing, a *Centralized* layout is the recommended choice [42]. Besides offering a higher Y_{df}, the *Centralized* layout also contributes to improved Y_{ovl} in D2W hybrid bonding, particularly under ultra-fine pitch regimes, by mitigating maximum radial overlay errors such as those caused by rotation and magnification at the periphery of the critical pad array.

In summary, while any layout with fewer critical pads generally improves yield over a *Full* layout, the most effective strategy for boosting defect yield is to use a compact, clustered distribution of critical pads. This approach is particularly advantageous in processes characterized by high defect densities and large defects.

E. The Impact of the Redundant Replicas

In many IC chips, identical blocks of circuits are often replicated to enhance yield [18]. For fine-pitch, large-scale designs with a high number of I/O pads, numerous sparsely distributed dummy resources can be allocated for redundancy. This flexibility allows designers to adopt either shared or dedicated redundancy schemes. Adding redundancy provides a straightforward improvement in Cu recess yield Y_{cr}, with gains directly tied to the number of redundant pairs. However, the impact on defect-related yield, Y_{df}, depends not only on the redundancy strategy but also on the main-replica spacing, the physical separation between the primary pad and its replica, and the size of defects. In HB process, defect sizes, whether from void tails or main voids, are typically more than 100 times the pad pitch. In such cases, shared redundancy is generally ineffective because the small spacing between pads in the same group makes them prone to simultaneous failure. To achieve meaningful yield gains, dedicated redundancy with sufficiently large spacing between paired pads is necessary. The influence of main-replica spacing on Y_{df} is explored below.

In this experiment, we specify a pad block size of 200 µm×200 µm for both W2W and D2W HB. To eliminate the influence of critical pads, the die is composed entirely of 100% redundant pads. Among the redundant blocks, half are designated as main pad blocks and the other half as their corresponding replica pad blocks, where the replica pads are located. The redundancy scheme adopts a 1:1 mapping between main pads and replica pads. In the pad block assignment, the main pad blocks are randomly distributed across the die. For each main pad block, the algorithm searches for candidate blocks located at a specified Euclidean distance and assigns one as the replica pad block. This process continues until all replicas are paired.

We first investigate the impact of the physical locations of replicas on the defect-related yield, Y_{df}. For both a 10 mm×10 mm die and a 3.2 mm×3.2 mm die, 1,000 pad layouts are randomly generated with a fixed main-replica spacing, and yield is evaluated under varying defect densities to assess the benefit of adding redundancy. Taking W2W HB as an example, the Y_{df} results shown in Fig. 18 demonstrate that defect yields across these 1,000 layouts are tightly clustered. This indicates that the main-replica spacing significantly impacts yield, while the exact placement of main and replica pads does not. Furthermore, the benefit of larger spacing becomes more pronounced at higher defect densities. Next, we examine the yield improvement achieved with different main-replica spacings. As shown in Fig. 19, we report the Y_{df} for both W2W and D2W HB across various redundancy configurations and two defect densities. Here, a spacing of 0 indicates no redundancy, while "20:1" denotes a shared redundancy strategy where 20 main pads share one replica. In W2W HB, redundancy consistently improves yield at both low and high defect densities, as this process is more sensitive to particle-induced defects compared to D2W HB. As the main-replica spacing increases from 200 µm to 800 µm, the yield improvement for W2W becomes more significant. Conversely, in D2W HB, the improvement diminishes with increasing spacing. This behavior correlates with the characteristics of dominant defects. In W2W HB, the primary issue is void tail defects, which can span several millimeters, as shown in Fig. 9a. Substantial yield improvement only occurs when the main-replica spacing exceeds a certain threshold (400 µm in this case), allowing one

pad in the pair to escape the defect region. In contrast, D2W HB is mainly affected by main void defects, which are typically smaller than 200 µm (see Fig. 10a). Thus, a spacing of 200 µm is generally sufficient to ensure redundancy effectiveness, and increasing it further brings diminishing returns. Moreover, the results confirm that shared redundancy does not improve Y_{df}, likely due to the increased risk of simultaneous failures within closely packed groups.

In summary, the defect yield is primarily influenced by main-replica spacing, not the exact locations of redundant I/O pads. A dedicated redundancy strategy with a spacing tuned to the dominant defect size in the bonding process can significantly enhance yield. However, since longer spacings may introduce routing delays, designers must carefully balance redundancy effectiveness and performance trade-offs during the early design stages.

VI. Conclusion

This work presents YAP+, a pad-layout-aware yield modeling framework for W2W and D2W hybrid bonding, which is proposed as an enhanced version of YAP. YAP+ models overlay errors, particle-induced void defects, Cu recess variations, and analyze the interaction between the failure of critical and redundant pads and the die failure. YAP+ is validated against a physics-inspired yield simulator. The proposed YAP+ yield model accurately predicts bonding yield across various chiplet sizes, pitches, pad layout configurations, and process parameters, and achieves a 1,000x runtime speedup over direct simulations while maintaining negligibly small mean square error. Case studies using YAP+ underscore critical tradeoffs between pad layouts, bonding approaches, and redundancy schemes, and reveal distinct yield-limiting mechanisms in W2W versus D2W bonding, offering concrete guidance for process control and chiplet architecture design.

Looking ahead, we aim to: (1) extend YAP+ into system-level assembly yield modeling that integrates chiplet, TSV, and interconnect yield; (2) generalize the framework to alternative bonding technologies, such as thermal-compression bonding; (3) explore yield enhancement techniques, including adaptive pad redundancy and fault-tolerant design strategies, informed by YAP+'s insights; (4) incorporate more multi-physics considerations into both modeling and simulation to capture the interactions among multiple failure mechanisms.

Acknowledgement

This work was supported in part by CHIMES, one of the seven centers in JUMP 2.0, a Semiconductor Research Corporation (SRC) program.

References

[1] Mitsuishi, H., Mori, H., Maeda, H., Ushijima, M., Kamashita, A., Okada, M., Aramata, M., Shiomi, T., Sakamoto, S., Takahata, K. & Others 50 nm overlay accuracy for wafer-to-wafer bonding by high-precision alignment technologies. *2023 IEEE 73rd Electronic Components And Technology Conference (ECTC)*. pp. 1664-1671 (2023)

[2] Sano, I., Yamagishi, M., Takyu, S., Kirihata, T., Kitazawa, R., Fukushima, T. & Kurita, Y. Direct Transfer Bonding Technology Enabling 50-nm Scale Accuracy for Die-to-Wafer 3D/Heterogeneous Integration. *2025 IEEE 75th Electronic Components And Technology Conference (ECTC)*. pp. 308-312 (2025)

[3] Ryan, K., Sreenivasan, R., Ramamoorthy, S., Patlolla, R., Appell, J., Hebding, J., Krishnan, S., Chudzik, M., Theile, B., Probst, G. & Theile, B. Integration, Materials and Equipment Innovations to Enable 100 nm Pitch W2W Bonding for Memory-to-Logic and Logic-to-Logic 3D Stacking. *2025 IEEE 75th Electronic Components And Technology Conference (ECTC)*. pp. 542-546 (2025)

[4] Patel, D., Xie, M., & Koch, J. Hybrid Bonding Process Flow – Advanced Packaging Part 5. https://semianalysis.com/2024/02/09/hybrid-bonding-process-flow-advanced/ (2024)

[5] Elsherbini, A., Jun, K., Vreeland, R., Brezinski, W., Niazi, H., Shi, Y., Yu, Q., Qian, Z., Xu, J., Liff, S. & Others Enabling hybrid bonding on Intel process. *2021 IEEE International Electron Devices Meeting (IEDM)*. pp. 34-3 (2021)

[6] Ikegami, Y., Onodera, T., Chiyozono, M., Sakamoto, A., Shimizu, K., Kagawa, Y. & Iwamoto, H. Study of Ultra-Fine 0.4 μm Pitch Wafer-to-Wafer Hybrid Bonding and Impact of Bonding Misalignment. *2024 IEEE 74th Electronic Components And Technology Conference (ECTC)*. pp. 299-304 (2024)

[7] Lau, J. Recent Advances and Trends in Cu–Cu Hybrid Bonding. *IEEE Transactions On Components, Packaging And Manufacturing Technology*. **13**, 399-425 (2023)

[8] Lee, S., Jee, Y., Park, S., Lee, S., Hwang, B., Jo, G., Lee, C., Park, J., Jang, A., Jung, H. & Others A study on memory stack process by hybrid copper bonding (HCB) technology. *2022 IEEE 72nd Electronic Components And Technology Conference (ECTC)*. pp. 1085-1089 (2022)

[9] Park, M., Cho, H., et al. A 192-Gb 12-High 896-GB/s HBM3 DRAM with a TSV Auto-Calibration Scheme and Machine-Learning-Based Layout Optimization. *2022 IEEE International Solid-State Circuits Conference (ISSCC)*. **65** pp. 444-446 (2022)

[10] Marinissen, E., Pancholi, V., Chuang, P. & Keim, M. IEEE Std P3405: New Standard-under-Development for Chiplet Interconnect Test and Repair. *2024 IEEE 42nd VLSI Test Symposium (VTS)*. pp. 1-11 (2024)

[11] Singh, E. Analytical modeling of 3D stacked IC yield from wafer to wafer stacking with radial defect clustering. *2014 27th International Conference On VLSI Design*. pp. 26-31 (2014)

[12] Xu, Q., Jiang, L., Li, H. & Eklow, B. Yield enhancement for 3D-stacked ICs: Recent advances and challenges. *17th Asia And South Pacific Design Automation Conference*. pp. 731-737 (2012)

[13] Campbell, D. Yield modeling of 3D integrated wafer scale assemblies. *2010 Proceedings 60th Electronic Components And Technology Conference (ECTC)*. pp. 1935-1938 (2010)

[14] Graening, A., Pal, S. & Gupta, P. Chiplets: How small is too small?. *2023 60th ACM/IEEE Design Automation Conference (DAC)*. pp. 1-6 (2023)

[15] Agnesina, A., Brunion, M., Kim, J., Garcia-Ortiz, A., Milojevic, D., Catthoor, F., Mirabelli, G., Komalan, M. & Lim, S. Power, Performance, Area, and Cost Analysis of Face-to-Face-Bonded 3-D ICs. *IEEE Transactions On Components, Packaging And Manufacturing Technology*. **13**, 300-314 (2023)

[16] Zhao, Y., Khursheed, S. & Al-Hashimi, B. Cost-Effective TSV Grouping for Yield Improvement of 3D-ICs. *2011 Asian Test Symposium*. pp. 201-206 (2011)

[17] Singh, A. Interstitial redundancy: an area efficient fault tolerance scheme for large area VLSI processor arrays. *IEEE Transactions On Computers*. **37**, 1398-1410 (1988)

[18] Koren, I. & Koren, Z. Defect tolerance in VLSI circuits: techniques and yield analysis. *Proceedings Of The IEEE*. **86**, 1819-1838 (1998)

[19] Chen, Z., & Gupta, P. YAP: Yield Modeling and Simulation for Advanced Packaging. *2025 62nd ACM/IEEE Design Automation Conference (DAC)*. pp. 1-7 (2025)

[20] Lau, J. Current Advances and Outlooks in Hybrid Bonding. *IEEE Transactions On Components, Packaging And Manufacturing Technology*. **15**, 651-681 (2025)

[21] Mariappan, M., Hashimoto, H., Mihara, K., Hare, T., Fukushima, T., Inoue, F. & Uedono, A. Impact of Cu Pad Density on Cu-CMP and Bonding Yield for Chip-to-Wafer Hybrid Bonding. *2024 International 3D Systems Integration Conference (3DIC)*. pp. 1-5 (2024)

[22] Kim, J., Seo, S., Kim, H., Kim, Y., Jo, C. & Kim, D. A study on bonding pad structure and layout for Fine pitch hybrid bonding. *2022 IEEE 72nd Electronic Components And Technology Conference (ECTC)*. pp. 712-715 (2022)

[23] Chen, J., Ho, C., Chuang, C., Chen, S., Chou, S., Shen, H., Yaung, D., Wang, C., Hung, F. & Ting, S. 3DIC structure and method for hybrid bonding semiconductor wafers. (2017,5), https://patents.google.com/patent/US9666566B1/en

[24] Zhang, B., Chew, S., Stucchi, M., Dewilde, S., Iacovo, S., Witters, L., Webers, T., Van Sever, K., De Vos, J., Miller, A., Beyer, G. & Beyne, E. Scaling Cu/SiCN Wafer-to-Wafer Hybrid Bonding down to 400 nm interconnect pitch. *2024 IEEE 74th Electronic Components And Technology Conference (ECTC)*. pp. 312-318 (2024)

[25] Moreau, S., Bouchu, D., Jourdon, J., Ayoub, B., Lhostis, S., Frémont, H. & Lamontagne, P. Recent Advances on Electromigration in Cu/SiO_2 to Cu/SiO_2 Hybrid Bonds for 3D Integrated Circuits. *2023 IEEE International Reliability Physics Symposium (IRPS)*. pp. 1-7 (2023)

[26] Kim, S., Fodor, F., Heylen, N., Iacovo, S., De Vos, J., Miller, A., Beyer, G. & Beyne, E. Novel Cu/SiCN surface topography control for 1 μm pitch hybrid wafer-to-wafer bonding. *2020 IEEE 70th Electronic Components And Technology Conference (ECTC)*. pp. 216-222 (2020)

[27] Ghaida, R., Gupta, M. & Gupta, P. Framework for exploring the interaction between design rules and overlay control. *Journal Of Micro/Nanolithography, MEMS, And MOEMS*. **12**, 033014-033014 (2013)

[28] Armitage Jr, J. & Kirk, J. Analysis of overlay distortion patterns. *Integrated Circuit Metrology, Inspection, And Process Control II*. **921** pp. 207-223 (1988)

[29] Okudur, O., Iacovo, S., Kang, S., Gonzalez, M. & Beyne, E. Simulations of Wafer-to-Wafer Bonding Dynamics and Deformation Mechanisms. *2024 IEEE 10th ESTC*. pp. 1-5 (2024)

[30] Kang, S., Iacovo, S., D'havé, K., Van Huylenbroeck, S., Okudur, O., Alexeev, A., Plach, T., Probst, G., Ding, T., Wimplinger, M. & Others. Investigation of Distortion in Wafer-to-wafer Bonding with Highly Bowed Wafers. *2024 IEEE 74th Electronic Components And Technology Conference (ECTC)*. pp. 386-393 (2024)

[31] Ren, H. Process Development and Process Window Investigation of Copper-Silicon Dioxide Die-to-Wafer (D2W) Hybrid Bonding. (University of California, Los Angeles, 2021)

[32] Chidambaram, V., Leong, Y. & Ren, Q. Wafer Level Fine-Pitch Hybrid Bonding: Challenges and Remedies. *2020 IEEE 22nd Electronics Packaging Technology Conference (EPTC)*. pp. 459-463 (2020)

[33] Gui, C., Elwenspoek, M., Tas, N. & Gardeniers, J. The effect of surface roughness on direct wafer bonding. *Journal Of Applied Physics*. **85**, 7448-7454 (1999)

[34] Dubey, V., Wünsch, D., Gottfried, K., Fischer, T., Helke, C., Hasse, M., Hanisch, A., Hofmann, L., Reuter, D., Wiemer, M. & Others Impact of Dielectric Types on Surface Topography for Wafer-Level Hybrid Bonding. *2024 IEEE 10th Electronics System-Integration Technology Conference (ESTC)*. pp. 1-5 (2024)

[35] Ji, L., Che, F., Ji, H., Li, H. & Kawano, M. Modelling and characterization on wafer to wafer hybrid bonding technology for 3D IC packaging. *2019 IEEE 21st EPTC*. pp. 87-94 (2019)

[36] Ji, L., Che, F., Ji, H., Li, H. & Kawano, M. Wafer-to-wafer hybrid bonding development by advanced finite element modeling for 3-D IC packages. *IEEE Transactions On Components, Packaging And Manufacturing Technology*. **10**, 2106-2117 (2020)

[37] Wang, H., Chen, H., Xiang, J. & Yang, X. Research on simulation of Cu/SiO_2 hybrid bonding process and interface failure mechanism by Finite Element Analysis. *2023 24th International Conference On Electronic Packaging Technology (ICEPT)*. pp. 1-7 (2023)

[38] Beilliard, Y., Estevez, R., Parry, G., McGarry, P., Di Cioccio, L. & Coudrain, P. Thermomechanical finite element modeling of Cu–SiO_2 direct hybrid bonding with a dishing effect on Cu surfaces. *International Journal Of Solids And Structures*. **117** pp. 208-220 (2017)

[39] Fujii, N., Furuse, S., Yoshioka, H., Ogawa, N., Yamada, T., Hirano, T., Saito, S., Hagimoto, Y. & Iwamoto, H. Bonding Strength of Cu-Cu Hybrid Bonding for 3D Integration Process. *ECS Transactions*. **112**, 3 (2023)

[40] Le, X. & Choa, S. Assessment of the Risk of Crack Formation at a Hybrid Bonding Interface Using Numerical Analysis. *Micromachines*. **15**, 1332 (2024)

[41] Zhao, G., Zeng, Y. & Zhao, Y. Simulation and Experimental Analysis of Thermomechanical Stress Around Interconnects for W2W Hybrid Bonding. *2024 IEEE 10th ESTC*. pp. 1-6 (2024)

[42] Xie, L., Shoji, Y., Jukei, M., Nomura, K., Fujiwara, T., Araki, H., Tanigaki, Y., Chong, S., Sekhar, V., Kumar, M., Tupaen, H. & Rao, V. Hybrid Bonding With Particle Accommodation Using Polymer Dielectric: Design, Process and Yield Study. *2025 IEEE 75th Electronic Components And Technology Conference (ECTC)*. pp. 92-98 (2025)

[43] Nagano, F., Iacovo, S., Phommahaxay, A., Inoue, F., Chancerel, F., Naser, H., Beyer, G., Beyne, E. & Gendt, S. Void formation mechanism related to particles during wafer-to-wafer direct bonding. *ECS Journal Of Solid State Science And Technology*. **11**, 063012 (2022)

[44] Kim, Y., Nguyen, T. & Choa, S. Enhancement of the bond strength and reduction of wafer edge voids in hybrid bonding. *Micromachines*. **13**, 537 (2022)

[45] De Messemaeker, J., Witters, L., Zhang, B., Tsau, Y., Fodor, F., De Vos, J., Beyer, G., Croes, K. & Beyne, E. New Cu "Bulge-Out" Mechanism Supporting SubMicron Scaling of Hybrid Wafer-to-Wafer Bonding. *2023 IEEE 73rd Electronic Components And Technology Conference (ECTC)*. pp. 109-113 (2023)

[46] Lin, H., Tran, D., Chiu, W., Chang, H. & Chen, C. In-situ measurement of thermal expansion in Cu/SiO_2 hybrid structures using atomic force microscopy at elevated temperatures. *Applied Surface Science*. **662** pp. 160103 (2024)

[47] Lin, H., Chiu, W., Chang, H. & Chen, C. Observation of Thermal Expansion Behavior of Nanotwinned-Cu/SiO_2 & Regular-Cu/SiO_2 Hybrid Structure via In-Situ Heating AFM. *2024 IEEE 74th Electronic Components And Technology Conference (ECTC)*. pp. 816-820 (2024)

[48] Maugis, D. On the contact and adhesion of rough surfaces. *Journal Of Adhesion Science And Technology*. **10**, 161-175 (1996)

[49] Rieutord, F., Moriceau, H., Beneyton, R., Capello, L., Morales, C. & Charvet, A. Rough surface adhesion mechanisms for wafer bonding. *ECS Transactions*. **3**, 205 (2006)

[50] Hutchinson, J. & Suo, Z. Mixed mode cracking in layered materials. *Advances In Applied Mechanics*. **29** pp. 143-145 (1991)

[51] Nagano, F., Inoue, F., Phommahaxay, A., Peng, L., Chancerel, F., Naser, H., Beyer, G., Uedono, A., Beyne, E., De Gendt, S. & Others Origin of Voids at the SiO_2/SiO_2 and SiCN/SiCN Bonding Interface Using Positron Annihilation Spectroscopy and Electron Spin Resonance. *Ecs Journal Of Solid State Science And Technology*. **12** (2023)

[52] Glang, R. Defect size distribution in VLSI chips. *IEEE Transactions On Semiconductor Manufacturing*. **4**, 265-269 (1991)

[53] Stapper, C. Modeling of defects in integrated circuit photolithographic patterns. *IBM Journal Of Research And Development*. **28**, 461-475 (1984)

[54] Bruls, E. & Others Characterization of defects in integrated circuits: resources, models and applications. (1992)

[55] Bresenham, J. Algorithm for computer control of a digital plotter. *IBM Systems Journal*. **4**, 25-30 (1965)

[56] Gonzalez, R. Digital image processing. (Pearson Education India, 2009)

[57] Chidambaram, V., Lianto, P., Wang, X., See, G., Wiswell, N. & Kawano, M. Dielectric materials characterization for hybrid bonding. *2021 IEEE 71st ECTC*. pp. 426-431 (2021)

[58] Dubey, V., Wünsch, D., Gottfried, K., Wiemer, M., Fischer, T., Schermer, S., Dittmar, N., Helke, C., Haase, M., Ghosal, S., Hanisch, A., Bonitz, J., Luo-Hofmann, J., Hofmann, L., Lykova, M., Stoll, F., Vogel, K. & Schulz, S. Impact of Dielectric and Copper Via Design on Wafer-to-Wafer Hybrid Bonding. *2023 IEEE 73rd ECTC*. pp. 795-799 (2023)

Zhichao Chen received the B.Eng. degree in VLSI design and system integration from the Electronic Science and Engineering Department at Nanjing University, Nanjing, China, in 2023. He is currently pursuing a Ph.D. degree in the Electrical and Computer Engineering Department at the University of California, Los Angeles (UCLA), CA, USA, under the supervision of Prof. Puneet Gupta. His current research interests include advanced packaging, computer-aided design, and design-technology co-optimization.

Puneet Gupta (Fellow, IEEE) received the B.Tech. degree in electrical engineering from the Indian Institute of Technology Delhi, New Delhi, India, in 2000, and the Ph.D. degree from the University of California at San Diego, San Diego, CA, USA, in 2007. He is currently a Faculty Member with the ECE Department, University of California at Los Angeles, and has authored over 200 papers, 18 U.S. patents, a book, and a book chapter in the areas of system-technology co-optimization as well as variability/reliability aware architectures.