

Design-for-Test Solutions for 3D Integrated Circuits*

Shao-Chun Hung*, Partho Bhoumik†, Arjun Chaudhuri‡, Sammitra Banerjee†, and Krishnendu Chakrabarty†

*Department of Electrical and Computer Engineering, Duke University

†School of Electrical, Computer and Energy Engineering, Arizona State University

‡NVIDIA Corporation

Abstract—As Moore’s Law approaches its limits, 3D integrated circuits (ICs) have emerged as promising alternatives to conventional scaling methodologies. However, the benefits of 3D integration in terms of lower power consumption, higher performance, and reduced area are accompanied by testing challenges. The unique vertical stacking of components in 3D ICs introduces concerns related to the robustness of bonding surfaces. Moreover, immature manufacturing processes during 3D fabrication can lead to high defect rates in different tiers. Therefore, there is a need for design-for-test solutions to ensure the reliability and performance of 3D-integrated architectures. In this paper, we provide a comprehensive survey of existing testing strategies for 3D ICs. We describe recent advances, including research efforts and industry practice, that address concerns related to bonding defects, elevated power supply noise, fault diagnosis, and fault localization specific to the unique characteristics of 3D ICs.

I. INTRODUCTION

Three-dimensional (3D) integration is a promising technology for achieving enhanced performance, reduced power consumption, and a smaller footprint compared to conventional 2D integrated circuits (ICs) [1]. Heterogeneous integration, involving the stacking of component layers with diverse fabrication processes, technology nodes, or sourced from different vendors, provides significant flexibility in producing cost-effective devices. According to the Heterogeneous Integration Roadmap (HIR) [2], 3D ICs and heterogeneous integration are important to advance the next-generation Internet of Things (IoT), mobile devices, data centers, and intelligent automotive technologies.

In today’s 3D integration technologies, die/wafer bonding with through-silicon vias (TSVs) is widely used because of its compatibility with conventional fabrication processes. In a TSV-based 3D IC, individual dies are produced on separate wafers using existing 2D manufacturing techniques. Through significant advancements in wafer thinning and TSV filling, these candidate dies can be precisely aligned and bonded together. Various schemes have been explored for TSV fabrication during manufacturing processes. In the via-first approach, TSVs are formed prior to complementary

metal–oxide–semiconductor (CMOS) processes, while the via-middle flow involves fabricating TSVs after the completion of front-end-of-line (FEOL) steps but before back-end-of-line (BEOL) processes. The via-last scheme manufactures TSVs at the end of the BEOL processing on the wafer [3].

TSV-based architectures have found widespread applications in commercial products. SK Hynix’s High Bandwidth Memory (HBM) places a logic die at the bottom and stacks cores of components for dynamic random-access memory (DRAM) to achieve compact footprint and low power consumption [4]. Intel Foveros enables logic-on-logic die stacking, incorporating a central processing unit (CPU), graphics, and artificial intelligence (AI) processors [5]. Sony’s CMOS image sensors leverage TSVs to stack pixels, DRAM, and logic components [6]. AMD has introduced 3D V-Cache to stack memory devices on a bottom compute die [7]. These ongoing explorations of 3D-integrated commercial products demonstrate the promise of 3D and heterogeneous integration technologies for the advancement of next-generation IC designs.

Monolithic 3D (M3D) integration is another promising technology that can achieve greater performance and lower power consumption than conventional 2D ICs. M3D leverages fine-grained monolithic inter-tier vias (MIVs) to achieve precise alignment and exceptionally thin device tiers [8]. Compared to TSVs, MIVs are associated with sizes that are one to two orders of magnitude smaller, and the induced capacitance is negligible [9]. These benefits motivate the extensive utilization of MIVs in M3D designs, significantly reducing the total wirelength.

M3D integration has the potential to facilitate a diverse range of applications. M3D NAND flash memory has entered commercial production in recent years, offering superior performance and cost-effectiveness compared to traditional 2D planar NAND Flash [10]. In [11], heterogeneously-integrated M3D IC is achieved by leveraging a CO₂ far-infrared laser annealing (CO₂-FIR-LA) technology to stack logic tiers, memory devices, and analog circuitry. In addition, M3D has received increased attention as a solution to address memory bandwidth limitations in AI accelerators. [12] introduces an M3D nonvolatile RAM interface to alleviate memory-related challenges during the training of convolutional neural networks (CNNs). In [13], a CNN accelerator leveraging heterogeneous M3D architecture with multiple technology nodes is proposed to achieve scalable peripheral logic, while memory cells remain at legacy nodes.

Despite the rapid advancements in 3D integration technolo-

*This research was supported in part by the National Science Foundation under grants CCF-1908045 and CCF-2309822, by the Semiconductor Research Corporation (SRC) under Contract 2470, and Intel Corporation. It was also supported in part by the DARPA ERI 3DSOC program under Award HR001118C0096, and CHIMES, one of the seven centers in JUMP 2.0, an SRC program sponsored by DARPA. The work of Arjun Chaudhuri at NVIDIA Corporation is unrelated to the contents of this paper.

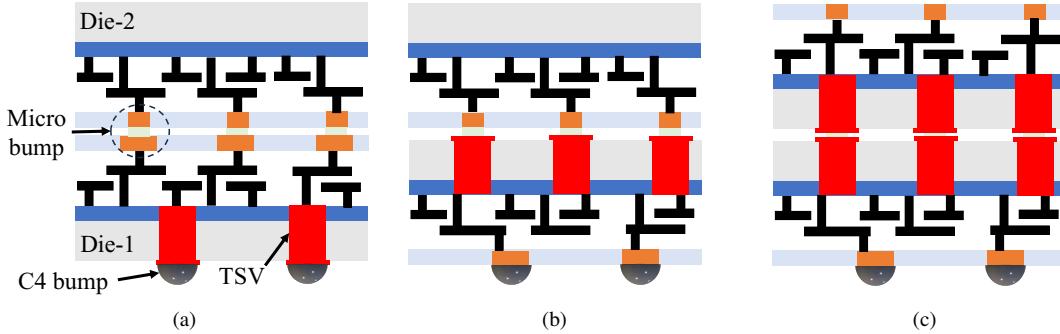


Fig. 1: Stacking methods for TSV-based 3D ICs: (a) face-to-face stacking; (b) face-to-back stacking; (c) back-to-back stacking [14].

gies and their applications, current 3D chip designs often overlook test cost considerations and the potential impact on testability. Reliability testing and analysis are required to be improved both at the board level and at the system level to eliminate potential risks throughout the production lifetime [15]. The lack of enhancement in 3D IC testing creates a gap between the predicted and actual benefits of 3D integration. This paper provides an overview of research efforts to develop test strategies for both TSV-based and M3D ICs. Our goal is to identify emerging test challenges and enhance the understanding of effective design-for-test methodologies.

The rest of the paper is organized as follows. Section II provides a background on TSV-based 3D and M3D integration technologies. Section III describes proposed test solutions for TSV-based 3D ICs. Section IV provides test solutions dealing with challenges unique to M3D integration. Industrial practices are presented in Section V. Finally, Section VI outlines further challenges and opportunities of 3D IC testing.

II. BACKGROUND

A. TSV-based 3D Integration

Breakthroughs in die/wafer bonding and TSV fabrication have paved the way for the realization of TSV-based 3D ICs. These advanced technologies enable the separate fabrication of different dies using existing 2D manufacturing processes, leading to minimal changes to current fabrication flows. Two bonding solutions have been proposed to stack different dies in a TSV-based 3D, including die-to-wafer bonding and wafer-to-wafer bonding. Die-to-wafer bonding does not require uniform sizes for dies and wafers, while wafer-to-wafer bonding is particularly well-suited for high-yield wafers of identical sizes, as the yield tends to decrease with an increase in the number of stacked wafers [16].

Fig. 1 illustrates the stacking strategies employed in TSV-based 3D ICs. In the face-to-face (F2F) stacking approach, two dies are precisely aligned and bonded with their active surfaces facing each other. The active surfaces are connected using microbumps; TSVs are utilized for bumping and package connections. For face-to-back (F2B) stacking, front-side connections of the first die are used to connect to the package, while TSVs and microbumps are used for inter-die connections. In back-to-back (B2B) stacking, dies in the TSV-based 3D IC face away

from each other, leading to the maximum separation between the front-end circuits [14].

TSV formation is another key development to realize TSV-based 3D ICs. Typically, this process commences with wafer thinning to reduce TSV thickness. Next, photolithography is employed to precisely define TSV locations on the wafer, paving the way for the subsequent etching process. The Bosch process [17], a solution based on deep reactive ion etching (DRIE), is widely adopted due to its advantage of creating deep TSVs. Following the etching phase, liner deposition is executed to enhance the electrical isolation of TSVs. To prevent metal diffusion into silicon and facilitate the selective deposition of conductive materials within the TSVs, barrier and seed layer deposition is carried out subsequently. Finally, electroplating fills the TSVs with a conductive material, followed by the chemical mechanical polishing step to planarize the wafer surface and eliminate excess materials. This comprehensive series of steps forms the foundation for TSV formation in TSV-based 3D ICs.

Commercial electronic design automation (EDA) tools have been developed to support TSV-based 3D integration throughout the fabrication processes. The Cadence Integrity 3D-IC Platform offers a comprehensive design and analysis system for 3D ICs [18]. The Synopsys 3DIC Compiler presents a unified environment for efficient 3D IC development [19]. The Siemens EDA 3D IC Design Flow provides a set of tools for 3D IC design flow, ranging from design architect to signal integrity [20]. Moreover, 3D IC integration platforms have been developed in industry to produce reliable 3D interconnects [21]. The evolution of these commercial tools highlights the significance of 3D ICs as a promising frontier for next-generation IC designs. With the rapid progress of 3D ICs, the exploration of testing solutions for these advanced technologies becomes extremely important to ensure the robustness and reliability of 3D-integrated designs.

B. M3D Integration

M3D integration is a promising technology to continue performance improvements beyond the constraints of Moore's Law. In an M3D design, all device tiers are sequentially fabricated on a single wafer, as illustrated in Fig. 2. Conventional fabrication processes for 2D design are applicable to the bottom-tier FEOL and BEOL. However, for top-tier

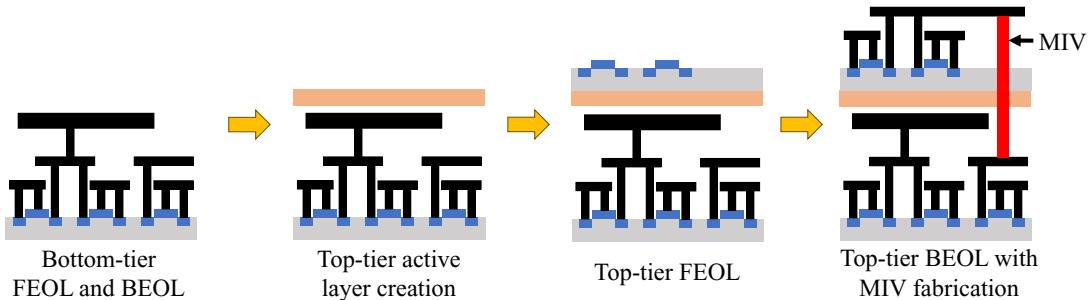


Fig. 2: Illustration of M3D fabrication processes [22].

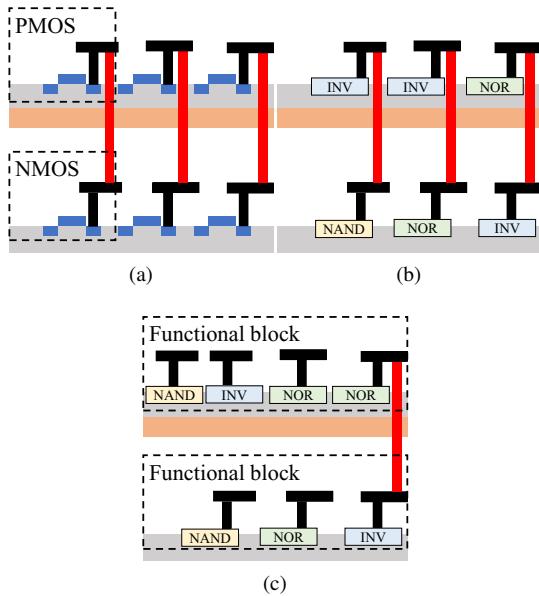


Fig. 3: Partitioning styles of M3D ICs: (a) transistor-level partitioning; (b) gate-level partitioning; (c) block-level partitioning [23].

fabrication, the adoption of low-temperature processes becomes essential to prevent interconnects and components in the bottom tier from damage. The realization of this approach has been made possible by significant breakthroughs in low-temperature manufacturing processes [24]. Recent research [16] has demonstrated that processing steps under the 550°C thermal budget yield reliable devices without adversely affecting the overall circuit performance.

M3D leverages MIVs to serve as interconnects between tiers. Compared to TSVs, MIVs can be adapted from conventional vertical plug technology, with their size being scaled down as regular 2D vias [22]. Moreover, the induced tensile stress during MIV fabrication is relatively minimal. These advantages make it feasible to incorporate a large number of MIVs in an M3D design, leading to a significant reduction in total wirelength. In addition, the effective use of MIVs as interconnects between memory devices and logic circuits has been demonstrated to address the communication bottleneck in M3D-integrated memory architectures [12].

Various partitioning approaches, including transistor-level, gate-level, and block-level partitioning, have been demonstrated in [25]. Figure 3 illustrates three distinct partitioning styles in M3D design. In transistor-level M3D ICs, P-channel and

N-channel transistors are distributed across different tiers; gate-level M3D involves the arrangement of standard cells within each tier. In block-level M3D ICs, functional blocks are strategically partitioned across multiple tiers. Research efforts have been devoted to the development of effective 3D partitioning algorithms. In [26], the RC parasitics are derived by scaling down dimensions of standard cells by $1/\sqrt{2}$ during placement and routing. Next, the entire design is projected onto tiers with half of the original footprint to convert 2D designs into an M3D manner. [27] introduces a graph neural network (GNN)-based algorithm to achieve significant improvements in power, performance, and area. However, current design methods primarily focus on optimizing functional operations. Exploration of testing algorithms for M3D remains limited. Therefore, there is a need for additional research efforts to optimize the testing flow for M3D to help ensure the robustness and reliability of M3D designs. This paper provides an overview of proposed M3D testing strategies and identifies challenges for further exploration

C. Silicon Photonics

The rapid emergence of big data from mobile, Internet of Things (IoT), and edge devices and the continuous growth in computing power have enabled deep neural networks (DNNs) to perform various complex tasks, such as natural-language processing, action recognition, game-playing, and image classification [28]. However, with Moore's law approaching its end, electronic artificial intelligence (AI) accelerators are facing fundamental bottlenecks due to the slowdown in the improvement of processing capabilities over the past several decades. In particular, moving data electronically in metallic wires in these accelerators is a major bandwidth and energy bottleneck [29].

Communication in the optical domain has enabled low-cost and high-bandwidth data transfer at a low power overhead [30]. However, earlier efforts to perform efficient photonic processing by employing bulk optics hit roadblocks due to the inefficiency and lack of scalability of the resulting systems [31]. However, in recent years, it has been shown that silicon photonic (SiPh) devices, where light is guided through the silicon medium on a CMOS chip, can perform ultra-fast computations in the optical domain. For example, custom SiPh hardware accelerators have shown significant promise by providing a monolithic, miniaturized, and scalable platform for carrying out resource-intensive computations in DNNs, even surpassing conventional

CMOS computing. In particular, SiPh components can be used to lower the computational complexity of N -dimensional matrix multiplication from $O(N^2)$ to $O(1)$ by exploiting the natural parallelism of optics [32]. Moreover, owing to the rapid advancement in foundry technology, integrated SiPh circuits can be fabricated with CMOS-compatible manufacturing techniques [33]. In summary, SiPh circuits enable high-speed parallelized matrix multiplication and high-bandwidth data transmission using optical interconnects – therefore, they are promising post-Moore alternatives to electronic AI accelerators.

Integration and packaging between SiPh and silicon CMOS components can exploit 2D, 2.5D, or 3D as well as even monolithic integration technologies as SiPh components utilize form factors compatible with silicon electronics [34]. 3D photonic integration on an SiPh platform offers low power consumption, high-density functionalities, high-yield manufacturing on a CMOS platform, and its natural convergence with the current 2.5D/3D packaging techniques on silicon interposers.

Traditionally, multilayer integration in SiPh ICs have been demonstrated using evanescent couplers that can reduce the overall footprint by a factor of the layer number [35]. However, such couplers are typically a few hundred microns long and, therefore, prohibit the scaling towards future dense integration. To enable photonic integration at the sub-10 μm pitch, through-silicon-optical-vias (TSOVs), which are analogous to TSVs in electronic ICs, have been proposed [36].

Despite the aforementioned advances, there exist several roadblocks to the further advancement of SiPh ICs. Fabrication process variations and thermal variations as well as device aging can adversely impact the robustness of SiPh ICs by introducing undesirable crosstalk noise, optical phase shifts, frequency drifts, power penalties, and photodetection current mismatches. Additionally, their performance is sensitive to optical losses accumulating when optical devices are serially connected. Recent work has shown that the inferencing accuracy of a multi-layer perceptron-based neural network implemented on SiPh ICs can drop by up to 70% due to very minute nanometer-scale changes in the critical dimensions (e.g., waveguide width and thickness) of the photonic devices [37] [38]. The accuracy of optical computation has also been shown to be sensitive to quantization errors [39], optical loss [40], and crosstalk [40]. The impact of such variations in SiPh ICs is also very different from those in electronic ICs. For example, while the concepts of better and worse are often clear in electronic devices under variations, many critical parameters in SiPh ICs (e.g., effective index of a waveguide) do not have intrinsically good or bad values. Instead, it is vital to understand and model process-variation distributions and study how variations are correlated across and within devices in SiPh ICs.

Existing efforts on analyzing such variations and aging effects in photonic communication networks cannot be applied to SiPh processors due to the fundamental differences in the functionalities of devices in ICs where, for example, there is a need to analyze the integrity of feedforward signals under variations. As a result, optimization techniques, such as circuit compaction and power-gating redundant components, that work

for electronic ICs, are ineffective in the case of photonic circuits [41], [42]. These uncertainties in SiPh ICs, coupled with the challenges of 3D integration, therefore, necessitate deeper exploration before SiPh 3D ICs can be reliably manufactured in high volume.

array

III. TSV-BASED 3D IC TESTING

Numerous types of defects can manifest in TSVs, each potentially impacting their efficacy. These defects include the presence of voids or cracks, instances of incomplete fillings within the TSVs, and the formation of pinholes in the oxide layer. Moreover, TSVs are susceptible to imperfections such as insufficient thickness of their insulating layers and misalignment of bonding pads during the stacking process. Moreover, interfacial delamination, die warpage, and cracks in micro-bump and TSV can also emerge due to thermo-mechanical stresses exerted during the manufacturing process [43] [44]. These defects occur both before and after the bonding process. Although we need to make sure the 3D stacked IC is functional after the bonding stage, we can increase the overall yield by ensuring more Known Good Dies (KGD) by testing the TSVs prior to bonding.

A. Pre-bond Testing

Testing TSVs during the pre-bond stage presents a significant challenge because one end of the TSV is embedded within the substrate while the other end is accessible for testing. There are two primary methods for testing the TSV through its exposed terminal: one involves direct physical contact using a probing mechanism, and the other employs an indirect approach utilizing Design-for-Test (DfT) architectures. The former method carries the risk of potential TSV damage during probing, while the latter necessitates area overhead while being susceptible to variations in PVT [45].

1) Probing Method: The main difficulty of probing a TSV directly is the limitation of probing technology due to the lower pitch of the TSV network and also the inadequate planarity of the microbumps [46] [47]. To mitigate these issues, [46] proposed a novel approach where multiple TSVs are accessed with a single probe needle forming a TSV network. A driver inside the needle formed a charge sharing circuit. A capacitor on that circuit is charged through the TSV(s) and the charging time is compared to a fault free TSV to detect the defective TSV(s). [48] further sped up this test by terminating the process after identifying certain number of faulty TSVs that cannot be repaired by existing redundant TSVs inside the chip. In [49], an Integer Linear Programming method to generate optimal test sessions as a mean of reducing testing time. The authors of [47] generated test sessions that sampled a TSV network from untested TSVs and when a faulty session is found, it is bi-partitioned repeatedly until the location of the TSV is found. They reduced the test time further by 50% by taking the clustered defect distribution of TSVs into consideration in [50].

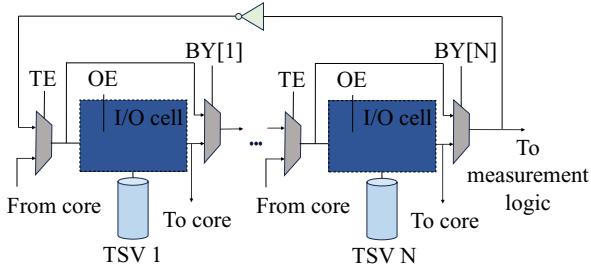


Fig. 4: Ring oscillator circuit where TSVs are integrated as load elements, with TE (Test Enable) and BY (Bypass) selection pins to toggle between testing and functional modes. Recreated from [51].

2) DfT Architectures: In this approach, a built-in DfT architecture is utilized to test the pre-bond TSV. The added circuit typically has different designs depending on the proposed algorithm. [51] proposed a ring oscillator-based TSV testing scheme where the exposed end of a TSV is connected to a ring oscillator as load. Deviations in the resistive and capacitive parameters of the TSV impact the overall delay of the oscillator circuit Fig. 4. Evaluating the oscillation frequency, we can detect a fault inside the TSVs. This paper also showed that by using multiple voltage levels, the detection of both resistive-open faults and leakage can be enhanced. In [52], the authors used the same model and showed that local process variations can also significantly impact the frequency of oscillation, thereby compromising the test. They proposed to use a relative frequency variation ($\Delta F/F$) for robustness against these PVT variations. The effectiveness of this method is verified by [53].

Several other approaches have also been considered. A pulse shrinkage based methodology is discussed in [54] and [55]. Here, the testing is carried out by evaluating the rise and fall delays of an input pulse. Although this approach offers a better resolution at the picoseconds level, it requires additional hardware near each TSV. In charge pump and pulse counting-based testing, the TSV is modeled as capacitors and the number of input pulses needed to charge them is counted. Comparing the number of counts with that for a fault-free TSV, a faulty TSV is detected [56] [57]. This proposed method in [45] utilizes a test structure with a weak current source to digitally quantify the charge/discharge rate of the TSVs' capacitance, allowing the detection of weaker faults that conventional techniques might overlook. This method requires more switches and contributes to area overhead. Another proposed BIST methodology compares the behavior of a TSV under test against a reference TSV, using a circuit composed of two cross-coupled inverters. By observing the final stable state of the well-known butterfly curve, it is possible to determine whether the TSV under test has a fault [58].

B. Post-bond Testing

After pre-bond testing is carried out, with the KGD and fault-free functional TSVs at hand, the dies are stacked. During the stacking process, defects can be caused by TSV misalignment, high temperature, and high pressure [59]. Several things can physically go wrong in a TSV; at the circuit level, they can be modeled as either a resistive open fault or a leakage fault.

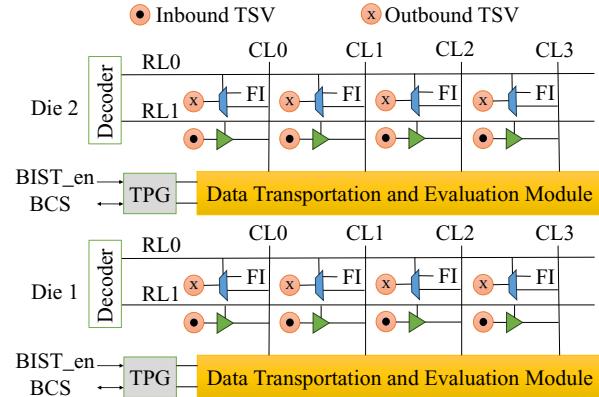


Fig. 5: Illustration of testing TSVs in post-bond stage. Circuit simplified from [60].

The post-bond test is important to detect these faults because one of these faulty vertical interconnects a stack of dies can be rendered faulty, and the stack might have to be discarded after packaging. As TSV-based 3D IC technology advances, the number of TSVs in a stack correspondingly increases. Today's main challenge lies in testing all the TSVs efficiently within a short time frame while simultaneously reducing hardware overhead.

In [60], a novel BIST scheme is proposed for testing TSVs during the post-bond stage. The TSVs are organized in a row-column configuration, akin to memory arrays Fig. 5. The testing process involves isolating and examining one row at a time for faults. If a single faulty TSV is detected within a row, the entire row is deemed faulty, prompting a detailed diagnostic process where each TSV in that row is individually tested to localize the fault. [61] considered a parallel testing and diagnosis process. In the proposed diagnostic scheme for a 4x4 TSV array, test patterns such as 1010 or 0101 are applied to each TSV row, with a bit shift after each test cycle to alternate the sequence. A Test Data Register (TDR) compares the responses from the TSVs with their previous state using an XOR operation to tell the difference between fault-free (output 1) TSVs and defective (output 0) TSVs. This is done because the test patterns dictate that for a fault-free TSV, the input to the XOR gate should be 0 and 1. A faster testing scheme was proposed in [62]. It discusses a method for simultaneous testing of multiple TSVs by assigning separate test circuits to each TSV. This method uses two sets of identical flip-flops to send test patterns across the TSVs and to a comparator circuit at the same time to identify defective TSVs. This strategy significantly reduces test duration but incurs a higher area overhead, which is compensated by an integrated mechanism.

The test architecture in [63] is adept at identifying resistive shorts in the oxide pre-bond and assessing resistance fluctuations within the TSV pathway post-bond. It features an innovative test circuit that can transform into a signal recovery circuit to rectify moderate defects identified before or after bonding, thus enhancing signal integrity. The test structure lets signals flow freely from the sender to the receiver for TSVs that pass inspection. After bonding, it checks each TSV again

to see if the resistance has changed. TSVs that initially pass the pre-bond tests but later exhibit significant resistance changes leading to signal degradation will switch to signal recovery mode after bonding. [64] used the same model and tried to find the weak open faults by calculating the delay of a pulse propagating through the TSV using a voltage divider framework. It is unable to detect any bridging faults if two of the TSVs were shorted together. [65] addressed this issue using a serial architecture, but this approach leads to an increase in test time. [66] proposed a test architecture that arranges TSVs in a 10x10 virtual matrix for efficient testing of 100 TSVs. It employs a partial parallel test method, testing TSVs initially line-by-line horizontally and then vertically, significantly reducing test time. This method identifies bridge defects between TSVs that might be missed in horizontal testing during the vertical test phase.

The methods presented in [64] and [65] employ an analog circuit for comparison, which is susceptible to process variations. In contrast, [67] introduces a ring oscillator-based testing strategy that leverages digital circuitry. This approach connects two TSVs to a ring oscillator, with their outputs directed to an XOR circuit. In contrast to the pre-bond ring oscillator solution, this method identifies faults by comparing the delay times within the oscillator across the two TSVs. For parallel testing, the outputs from all XOR gates are combined into an OR tree. However, it is important to note that an increase in the number of TSVs leads to a proportional increase in the area required for this testing setup. Leveraging the same model, [68] offered a new solution by feeding a periodic signal through the TSVs. The received signal is processed through two unbalanced inverters. By comparing the duty cycle of the output of the inverters with a fault-free case, faulty TSVs are detected. Another test approach was described in [69]. The main idea is that defects change all the electrical parameters of a TSV including resistance, conductance, and capacitance. This approach measures all these parameters and is able to detect open faults, leakage faults, high-resistance faults, and delay faults associated with a TSV.

C. TSV Repair

Faulty TSVs can significantly impact the overall chip yield. It is therefore important to tackle not only the testing of TSVs, but it is also important to develop repair mechanisms that can ensure recovery from TSV defects. Research shows that in most cases faulty TSVs are clustered together due to fabrication defects [70], [71]. If one TSV is defective there is a high probability that the neighboring TSVs are also defective. In order to address this problem, redundant TSVs (RTSVs) can be used along with the rerouting of the signal from a faulty TSV (FTSV) through the nearest RTSV. Several such designs have been proposed to perform this task focusing on different optimization metrics. We discuss some of these methods and compare the solutions based on repairability, area overhead, and delay.

1. Router-Based Architecture: [71] proposed a method where all the TSVs are bundled in a grid and RTSVs are placed in one row and one column in the outermost region of the grid. Each

TSV is equipped with multiple 3-to-1 MUXs to enable signal shifting and rerouting. For the repair of FTSVs, the strategy involves leveraging both neighboring TSVs and RTSVs situated far away. In the event of clustered defects, the FTSV closest to the RTSVs is directly connected to one of the RTSVs for rerouting. To repair the remaining FTSVs, each of their signals is bypassed to a neighboring good TSV (STSV). Since an FTSV signal occupies this STSV, the signal carried by this STSV is bypassed to another neighboring STSV. The bypassing chain is continued until the final STSV of this chain is connected to an RTSV.

2. Ring-Based Architecture: In this mechanism [70], TSVs are organized in a mesh pattern and divided into several rings, and the diameter of each ring increases from the inner ring to the outermost ring. RTSVs are placed on the outermost ring. Signals in each TSV can be shifted to the TSVs of the same ring or its outer ring. Similar to the router-based design, in the case of a fault, the signal can move from the inner ring to its adjacent outer rings one by one until an RTSV is reached.

3. Group-Based Architecture: In this approach, a TSV block is divided into multiple smaller groups of TSVs and each group contains one RTSV [72]. The RTSV in each group is equipped with a MUX consisting of two input components. The first includes all the signals from the same group and the second contains one signal from each of the remaining groups. Since the number of RTSVs and interconnects directly depends on the number of groups, balancing the grouping of TSV blocks is critical. Having too few groups (RTSVs) can lead to inefficient hardware usage by necessitating the incorporation of larger MUXs, whereas having too many groups (RTSVs) can overly complicate the system by requiring complex signal rerouting. The repair strategy is similar to that for router-based solution.

4. Cobweb-Based Architecture: In the cobweb-inspired design, TSVs are organized as a series of concentric layers, termed cobwebs [73]. The RTSVs are placed at the vertices of the outermost cobweb. Each inner cobweb can shift the signal to its neighboring TSV in the same cobweb and to the adjacent TSVs in the next outer cobweb. Signals are directed clockwise in the odd-numbered cobweb layers and anticlockwise in the even-numbered layers. However, the outermost cobweb has the flexibility to relay signals in both directions. Each TSV in this design is equipped with different sizes of MUXs, determined by the need to route signals from their nearby TSVs. The repair mechanism for this architecture is similar to the methods discussed so far. This approach offers a higher repair rate than the ring-based design and lower hardware overhead compared to the router-based solution.

5. Honeycomb-based Architecture: The authors proposed a design in [74] where TSVs are bundled in a honeycomb-based structure with each layer consisting of at least one honeycomb. TSV sharing is also utilized between each adjacent pair of honeycombs. RTSVs are placed at the center of each honeycomb in the outermost layer. Signal shifting can be achieved either in a clockwise or anticlockwise fashion. MUXs of various sizes are employed within the TSVs, where the MUX size is contingent upon the location of the TSVs in the

TABLE I: Qualitative comparison between different solutions for TSV repair

Solution	Repairability	Area Overhead	Delay
Router	Extremely High	Extremely High	Extremely High
Ring	Very Low	Moderate	High
Group	High	High	Very High
Cobweb	High	Low	Not Specified
Honeycomb	Very High	High	Very Low

design. During testing, if an FTSV is detected, the proposed repair algorithm identifies a repair path, primarily focusing on minimizing both wire and MUX delays to maintain signal integrity and system performance.

Each repair solution presented so far can achieve a considerable TSV repair rate; however, there remains ample scope for improvement in reducing hardware overhead and repair path delays to enable the production of smaller and faster 3D ICs. Table II presents a comparative analysis as a reference for 3D IC designers to select the most suitable repair strategy based on their specific requirements.

D. IEEE Std P1838

To allow testability and controllability in 3D ICs throughout all the stages of the fabrication process, adequate access is needed to the dies within a stack. In conventional 2D devices, accessing the dies is relatively easy through the core-level IEEE Std. 1149.1 [75] & IEEE Std. 1500 [76]. However, for 3D systems, only the first die can be accessed from one side of the package. In addition, all the dies in a stack may not be from the same provider, and testing them is a challenge [77]. To mitigate these issues, IEEE Std. 1838 was created to facilitate testability on a per-die basis, and the resulting DFT architecture can test both inter and intra-die logic paths. It allows test access and the freedom to deploy different testing strategies. It consists of three principal hardware components.

1. **Serial Control Mechanism:** The purpose of this mechanism is twofold: first, to establish communication between the interface and the stacked dies, facilitating the transfer of test stimuli and responses; second, to configure each die to its respective test modes. Each die has its own primary and secondary interfaces. When a new die needs to be stacked, its primary interface is connected to the secondary interface of the bonded die. Both the interfaces are equipped with five terminals in opposite order: TDI (Test Data Input), TCK (Test Clock), TMS (Test Mode Select), and TRSTN (Test Reset Not), and the output terminal TDO (Test Data Output) [77]. There are some specific registers carrying out unique tasks: an instruction register to set up the die in INTEST or EXTEST mode, a TAP configuration register to select a particular die and configure its TDI TDO path, and a bypass register to save time.

2. **Die Wrapper Register (DWR):** The DWR enables the controllability and observability of the input and output terminals of a 3D chip or chiplet. The goal of this is to test the interconnects between the dies in a stack Fig 6. The structure of DWR offers considerable versatility in its construction. It

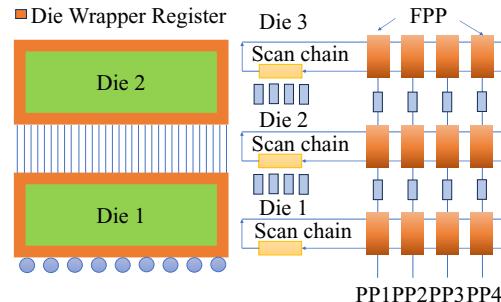


Fig. 6: IEEE Std 1838 components. DWR (left) and FPP (right). This figure is adapted from [77] and redrawn.

can be put together with different parts, such as core wrapper register segments, specialized wrapper cells, and boundary scan cells that are compliant with IEEE Std. 1149.1 [78]. In this setup, the scan chain's input is linked to the test input of the Die Wrapper Register (DWR), and the DWR's test output is connected to the scan chain's output. This arrangement of TAPs forms a connection with the DWR's inherent test inputs and outputs. [77].

3. **Flexible Parallel Ports (FPP):** In practical chip designs with multiple stacked dies, testing the last dies in the stack requires the shifting of test patterns through numerous flip-flops. For faster access to all the dies and high-bandwidth testing, there is an optional flexible parallel port mandated by IEEE Std. 1838 that allows the transport of multiple stimuli up and down the stack Fig 6. To reduce area overhead, on-chip scan chains are accessed with multiplexers for this port. In this configuration, there are two types of lanes: registered and non-registered. Registered lanes are utilized for transferring and receiving test data, while non-registered lanes provide essential support functions like clocking, power supply, and enabling signals. This distinction helps in efficiently managing data that requires storage and data that does not [77].

E. Testing SiPh ICs

Real-time health monitoring using an in-package run-time test methodology is necessary to operate the feedback control operations that counter the impact of run-time uncertainties in SiPh ICs. However, one of the main barriers to in-package testing in photonic ICs is the lack of efficient non-invasive monitoring tools to inspect the optical signal propagation without altering its operating point. Current solutions for in-package optical monitoring in photonic ICs [79] lack scalability and cannot be applied to SiPh ICs, where thousands of photonic components need to be probed [80], which would impose significant optical attenuations to the circuits they monitor. For example, traditional integrated light sensors, such as germanium photodiodes rely on extracting a portion of the light from the main propagation path to measure it. Such approaches are not feasible for high-density circuits, as extensive monitoring would cause an unacceptable power penalty at the output [81].

A promising approach that used only capacitive access with no optical attenuation to monitor optical signal quality was proposed in [82]. In this work, the authors propose a ContactLess Integrated Photonic Probe (CLIPP) that exploits

the change in the waveguide conductance induced by native interaction of photons in the Si-SiO₂ interface. CLIPP utilizes a capacitive access to the waveguide, thereby avoiding direct contact with the waveguide core, and no specific treatments at the waveguide surface need to be done.

Although, as discussed above, there has been recent efforts in developing optical probes that can be scaled to high-density designs, test access and control mechanisms for SiPh ICs (analogous to boundary scan, scan chains etc. in electronic ICs) have largely remained unexplored. Moreover, test scheduling for SiPh ICs and in-field run-time tests also need to be explored. The search for efficient test solutions for 3D ICs, should, therefore, not only be focused on Si-CMOS-based circuits, but should also consider post-Moore emerging technologies such as SiPh.

IV. M3D IC TESTING

M3D integration introduces new challenges in testing and diagnosis. Power consumption is one of the major concerns for M3D. Recent work has demonstrated that M3D designs suffer more from power supply noise (PSN) compared to their 2D counterparts [9]. The problem of PSN-induced voltage droop is exacerbated during testing due to excessive switching activities, leading to increased circuit delay and failure output responses. In addition, unlike TSV-based 3D ICs, M3D ICs feature all device tiers fabricated *in situ* on the same wafer. Therefore, pre-bond testing algorithms for TSV-based 3D designs are inapplicable to M3D. Moreover, MIVs suffer from a high defect rate as they need to penetrate through the inter-tier dielectric. The surface roughness of the inter-tier dielectric can create voids within and between MIVs, leading to MIV open and short defects. However, testing strategies for TSVs cannot be applied to M3D designs due to substantial area overhead when addressing a large number of MIVs. Therefore, there is a need for innovative test solutions to mitigate these challenges before M3D can become ready for commercial exploitation.

A. Power-safe Testing

PSN-induced voltage droop is severe in M3D designs because current is necessary to traverse upper-tier power delivery networks (PDNs) to supply bottom-tier devices. In addition, the reduced footprint decreases the number of power supplies (i.e. C4 bumps), leading to an increase in current density. Research effort has been devoted to designing a reliable PDN for M3D designs [9], [83]–[85]. However, existing algorithms primarily concentrate on mitigating PSN during functional operations, leaving PSN-induced voltage droop as a major concern during M3D testing.

In [23] [86], innovative pattern reshaping algorithms are introduced to address yield loss caused by PSN. The approaches extend the established power and rail analysis flow in commercial tools for 2D designs to the context of M3D, which can appropriately identify the worst-case voltage droop during testing. Next, the critical delay is scaled based on the obtained voltage droop values to extract test patterns that are susceptible to yield loss. Such patterns are regenerated with don't-care

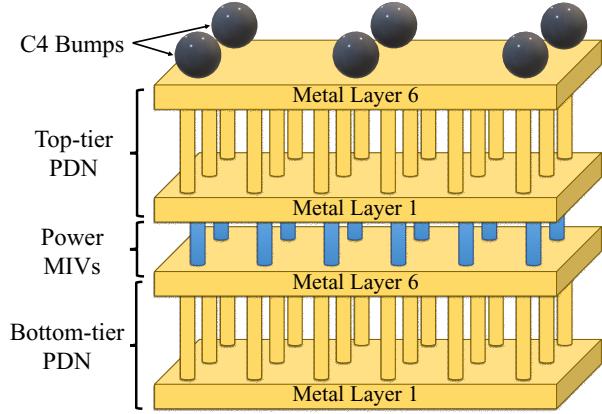


Fig. 7: Illustration of an M3D PDN [85].

bits (X-bits) unfilled, followed by the proposed X-bit filling algorithms. Two X-bit filling algorithms, including an integer linear programming-based solution and a simulated annealing-based solution, have been proposed to mitigate the PSN-induced voltage droop. Experimental results have demonstrated the efficacy of these algorithms in eliminating PSN-induced yield loss issues for both M3D designs with and without test compression.

DfT strategies have been explored to ensure power-safe testing for M3D designs [88] [89]. These strategies highlight the necessity of considering the impacts of vertically stacked M3D PDNs to accurately assess the PSN-induced voltage droop. Fig. 7 illustrates an M3D PDN design, where the top-tier PDN can be extended to include PCB, package, and C4 bumps [83] [85]. To supply the bottom-tier PDN, current needs to flow through top-tier PDN metal layers and power MIVs. Such a long detour of the conduction path from power supplies to local receivers can cause additional IR-drop. Neglecting this impact may lead to an underestimation of the PSN-induced voltage droop in the bottom-tier devices by up to 250 mV [89]. In addition, the solutions identify that simultaneous switching activities of gates in both tiers which share the same power source are the primary contributors to high PSN-induced voltage droop. Validation through partitioning post-routed M3D designs into tiles, based on C4 bump locations, demonstrates a strong correlation between tile-based switching activity and worst-case voltage droop, with the Pearson Correlation Coefficient of 0.87. Therefore, minimizing the tile-based switching activity during testing is an important factor to ensure power-safe testing.

To mitigate tile-based switching activities, [88] introduces a reinforcement learning (RL)-based framework for test point (TP) insertion. In each iteration, RL models determine the optimal tile for TP insertion and the types of TPs to be inserted. The training process utilizes a proposed reward function, combining metrics to balance switching activity reduction and test coverage improvement. The comparison with baseline cases using commercial TP insertion tools highlights the effectiveness of the proposed solution in reducing test power without compromising test coverage. In addition, [89] presents a scan segmentation framework to guide pattern

TABLE II: Diagnosis signatures obtained from Equation (1) with different fault origins [87], where R_o is the equivalent resistance of an MIV open.

Fault origin	Size of defect	Fault model	Output diagnosis signatures from Equation (1)				
			r_{MIV}	$r_{\Delta L}$	r_{HRS}	r_{ref0}	r_{HRS}
MIV open	$5.8 \text{ k}\Omega \geq R_o > 300 \text{ }\Omega$	Stuck-at-1 fault	0	0	1	1	1
	$21 \text{ k}\Omega \geq R_o > 5.8 \text{ k}\Omega$	Undefined-state fault	0	0	1	1	1
	$R_o > 21 \text{ k}\Omega$	Stuck-at-0 fault	0	0	1(0)	0	1(0)
Variation in oxide thickness	Increase in oxide thickness $\geq 20\%$	Stuck-at-1 fault	1	0	1	1	1
Variation in length	Increase in length $\geq 6.2\%$	Stuck-at-0 fault	0	0	0	0	1
	$6.2\% > \text{Increase in length} \geq 3.8\%$	Undefined-state fault	0	0	1	0	1
	Decrease in length $> 2.2\%$	Stuck-at-1 fault	1	1	1	1	1

generation with minimal tile-based switching activities. A two-step optimization flow includes Tile-level SDFF segmentation, which leverages the customized K-means clustering algorithm to partition scan d flip-flops (SDFFs) into segments in each tile, and the RL-based Top-level enable signal assignment, which assigns optimal control signals to the SDFF segments that help in mitigating redundant tile-based switching activities during testing. Compared to existing scan segmentation algorithms, the proposed solution achieves up to a 3x reduction in tile-based switching activities without impacting test coverage. As control units of SDFF segments are equally distributed in the existing scan chains, the test time and area overhead can be easily amortized.

B. Diagnosis of Memory-on-logic Stacked M3D

Memory-on-logic stacking stands out as a primary application of M3D integration. Among the next-generation non-volatile memory architectures, resistive random-access memory (RRAM) has demonstrated its compatibility with M3D to achieve high memory density and low cost per bit [90] [91]. Despite these promising aspects, both M3D and RRAM are relatively new technologies, which are susceptible to high defect rates. Therefore, there is a crucial need to test and diagnose the root causes within an M3D-integrated RRAM device to facilitate effective yield learning and enhance the reliability of these emerging technologies.

In [87], a novel diagnosis sequence is proposed to differentiate between MIV defects and RRAM process variations. The characterization results of defective M3D-integrated RRAM devices, partitioned with the method in [91], demonstrate that both MIV opens and RRAM process variations can be manifested as various faults, including stuck-at-1, stuck-at-0, slow-to-fall, undefined-state, and overforming faults. While existing test algorithms can identify these faults, distinguishing fault origins that lead to the same behaviors remains challenging due to identical output signatures during testing. To improve the diagnostic resolution, [87] develops a diagnosis sequence as follows:

$$\{\uparrow\downarrow (w0, r_{MIV}, r_{\Delta L}, r_{HRS}, r_{ref0}); \uparrow\downarrow (w1, r_{HRS})\} \quad (1)$$

where $w0$ ($w1$) is the write-0 (write-1) operation, r_{MIV} , $r_{\Delta L}$, r_{HRS} , and r_{ref0} are read operations with their corresponding reference resistances. The reference resistances for r_{HRS} and r_{ref0} represent inherent RRAM technology properties,

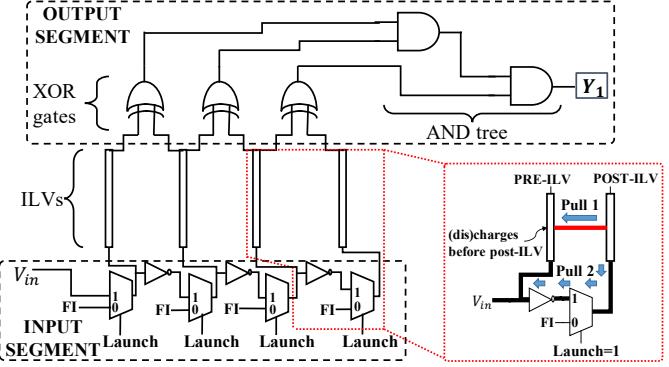


Fig. 8: Illustration of the XOR-BIST architecture.

specifically the nominal high-resistance-state and the upper bound of undefined states, respectively. For r_{MIV} and $r_{\Delta L}$, reference resistances can be derived from SPICE simulations or calculated using the proposed generalized solution.

Table II provides the details of output signatures using the RRAM model from [92] and the Nangate 45nm Standard Cell Library. Note that slow-to-fall faults and overforming faults are excluded from the list because only one fault origin can produce such behaviors. Therefore, test outputs and existing algorithms are sufficient to identify the root-cause fault origins. For the same fault model, distinct fault origins generate different output signatures. By leveraging these signatures along with test output responses, the proposed solution achieves a diagnostic resolution of 100%. This solution does not necessitate additional DfT structures, making it applicable to large-scale M3D-integrated RRAM systems.

C. MIV Built-in Self-test

In [93], [94], a built-in self-test (BIST) framework has been proposed that can effectively detect single and multiple SAFs, shorts, and opens in MIVs. The proposed BIST methodology achieves nearly 100% fault coverage (both single and multiple faults) of the MIVs with only two test patterns.

1) XOR-BIST Architecture for Fault Detection

The BIST architecture to test for faults in MIVs is illustrated in Figure 8. On the output side of the MIVs, 2-input XOR gates are inserted between neighboring MIVs. For a set of N MIVs placed in a 1D array-like manner where every MIV has at most two nearest neighbors, $(N - 1)$ XOR gates are inserted. The XOR outputs are fed as inputs to a space compactor which

is an optimally balanced AND tree with $(N - 1)$ inputs and a 1-bit output signature Y_1 . By observing Y_1 , it can be determined whether a fault is present in the MIVs under consideration. Test patterns are fed to the inputs of the MIVs from an input source V_{in} ; V_{in} feeds an inverter chain that generates complementary signals to adjacent MIVs in the test mode. A 2:1 multiplexer is present at the input of every MIV to switch between test mode and mission mode (functional input—*FI*) based on the *Launch* signal. The MIVs are tested in two clock cycles by switching V_{in} . The test patterns to the MIVs are “010...” ($V_{in} = 0$) and “101...” ($V_{in} = 1$) in the first and second cycles, respectively. It can be proven that a group of MIVs does not contain a hard fault if and only if Y_1 is 1 in both clock cycles. Aliasing occurs only when all MIVs are alternately stuck at 0 and 1, leading to masking of the MIV faults. However, the probability of occurrence of such a scenario is $\frac{2}{3^n}$, where n is the number of MIVs under test.

The inverter chain-based method of driving the MIVs in the test mode leads to a deterministic hard-short behavior. If a short is present between two MIVs, the MIV appearing first (pre-MIV) in the path of the incoming test signal from V_{in} , via the inverter chain, will drive the other MIV (post-MIV); this is illustrated in the inset of Figure 8. It is because the short provides a path of lower resistance (“pull 1”) compared to the path through the multiplexer and inverter (“pull 2”).

2) Dual-BIST Architecture

The BIST design described above may be affected by SAFs, which in turn can potentially mask MIV fault(s). To reduce the likelihood of masking, a second propagation path is added from the MIV outputs to a 1-bit signature Y_2 . The topology of this path to Y_2 (BIST-B) is identical to that of the path from the MIV outputs to Y_1 (BIST-A). The XOR and AND gates in BIST-A are substituted with the corresponding logical dual gates (XNOR and OR, respectively) in BIST-B. The MIVs under test, along with the “dual-BIST” engine, are considered to be fault-free if and only if $Y_1 = 1$ and $Y_2 = 0$ for both test patterns. With the “dual-BIST” architecture, it can be proven that a single fault in the dual-BIST engine cannot mask MIV fault(s). Furthermore, the probability of masking due to multiple faults in the dual-BIST engine is negligible.

D. Fault Localization

For the emerging M3D technology, different tiers are susceptible to diverse defects due to immature manufacturing processes. Levering low-temperature processes is necessary during top-tier fabrication to prevent damages from bottom-tier devices; however, the adoption of these emerging low-temperature processes can lead to performance mismatches of up to 20% between tiers [95]. In addition, the MIV layer is prone to defects as it traverses the inter-tier dielectric. The presence of surface roughness on the dielectric can induce voids within MIVs during the etching process [96]. These defects are manifested as delay faults situated in different tiers. Fault localization is therefore important to provide valuable feedback to the foundry and facilitate yield learning.

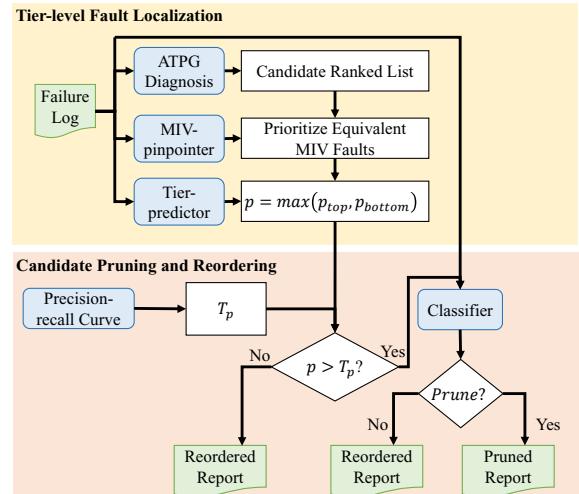


Fig. 9: Flowchart of the GNN-based localization framework [97].

In [98] [97], a diagnosis framework is introduced based on graph neural networks (GNN) to pinpoint delay faults at the tier level, as illustrated in the flowchart presented in Fig. 9. Given a failure log file from the tester, the diagnosis flow in a commercial automatic test pattern generation (ATPG) tool and the inference of two GNN models, namely MIV-pinpointer and Tier-predictor, are conducted simultaneously. The ATPG diagnosis generates a ranked list including all potential defect locations at the gate level, while MIV-pinpointer and Tier-predictor predict faulty MIVs and the faulty tier, respectively. These predictions offer early feedback to the foundry for manufacturing process review, without the need for subsequent physical failure analysis (PFA).

Another application of GNN predictions is to enhance the quality of the diagnosis report. Candidates corresponding to the predicted faulty MIVs are prioritized in the ranked list from the ATPG diagnosis as MIVs are prone to defects. Next, the probability of the predicted faulty tier from Tier-predictor, denoted as p , is compared to a threshold T_p obtained from a precision-recall curve during training. If $p > T_p$, the prediction of Tier-predictor has high confidence. Therefore, the third GNN model, Classifier, is employed to decide whether to prune or reorder the diagnosis report. This step aims at enhancing diagnostic resolution and the first-hit index. Otherwise, a reordered report is generated to maintain diagnosis accuracy. The improvement in the quality of the diagnosis report helps in significantly reducing the runtime for PFA. Moreover, the proposed framework is transferable to designs with different configurations (e.g., clock frequency, 3D partitioning results, TP insertion) without requiring retraining.

Due to the presence of a large number of MIVs in an M3D design, many fault-effects in a tier propagate to scan flops and observation points (OPs) in the other tier; if those fault-effects are not captured within the same tier where they originate, tier-level fault localization is hindered. Tier-level fault localization is crucial in identifying M3D tiers that are systematically prone to process variations, M3D-specific defects such as defects in MIVs and interfacial bonding, and tier-to-tier mismatch in the

transistor's threshold voltage. Tier-level fault localization is enabled by ensuring the capture of maximum fault-effects of the logic gates in an M3D tier by a set of dedicated OPs on the tier's POs (outgoing MIVs). As the POs of a circuit have high observability, they are not suitable candidates for OPI targeted at enhancement of circuit observability. Therefore, high testability and high fault localization are two orthogonal objectives for inserting OPs with the intent of tier-level fault localization.

In [99], the authors present a methodology for inserting OPs on the input-ends of specific outgoing MIVs so that the inserted OPs can detect most of the faults in the tier in which they are located. However, a higher OP count leads to insertion of more scan elements, thereby increasing the area and adversely affecting PPA. Typically, a large number (100K or more) of MIVs are likely in complex M3D designs. Thus, there are many candidate MIVs where OPs can potentially be inserted. The location of the OPs must be carefully determined to maximize fault localization and minimize the impact on PPA. The authors present the NodeRank algorithm that analyzes the topology of the M3D netlist and the MIV locations in the different M3D tiers, and uses a heuristic node-weight split-and-gather algorithm to determine the MIVs with the higher accumulated node weights which are indicators of effective OPs. A metric called *degree of fault localization* is used to evaluate the effectiveness of NodeRank-guided OP insertion towards tier-level fault localization.

The degree of fault localization DFL_i for tier T_i ($1 \leq i \leq n$) in an n -tier M3D design is given by:

$$DFL_i = \frac{X_i^{P_i}/F_i}{(X_i^{P_i}/F_i) + \sum_{\substack{j=1 \\ j \neq i}}^n ((1 - f_j^{P_i}) \cdot (X_j^{P_i}/F_j))}$$

Here, F_i is the total number of faults in T_i , P_i is the pattern-set to test for faults in T_i , $X_i^{P_i}$ ($X_j^{P_i}$) is the number of faults in T_i (T_j) that are activated by at least one pattern in P_i , with the errors being captured by scan flops and OPs in T_i . The fraction of $X_j^{P_i}$ faults that are P_j -localized is denoted by $f_j^{P_i}$; a fault F is said to be P_j -localized if F is present in T_j , activated by P_j , and the error due to F is captured by scan flops and OPs in T_j . Higher value of DFL_i indicates higher effectiveness of OPI for tier-level fault localization.

E. Development Trends

The comprehensive comparisons among existing M3D testing and diagnosis solutions are demonstrated in Table III. Powersafe testing remains a significant hurdle in M3D. Existing solutions have been shown to successfully reduce switching activities during testing. However, they necessitate extra test patterns to maintain test coverage, leading to increased test time. Therefore, there is a continuous demand for the evolution of novel techniques that concurrently optimize power consumption, test coverage, and test time.

The diagnosis of M3D-integrated RRAM is crucial for enhancing yield learning. A significant application of M3D-integrated RRAM lies in its capacity to store multiple

bits within a single cell, known as multi-level cell (MLC) RRAM [91]. However, MLC RRAM is more vulnerable to manufacturing defects and process variations than single-bit RRAM due to the closer distance between neighboring levels. It is important to extend current solutions to address defects in MLC RRAM effectively to ensure the full realization of the advantages offered by M3D integration.

In addition, ensuring the reliability of MIVs is crucial for guaranteeing the functionality of M3D designs. In heterogeneous M3D integration, MIVs are used to connect tiers with different processing nodes. Detecting and diagnosing faults within MIVs and bonding interfaces are essential to uphold design functionality. Therefore, enhancing MIV BIST and diagnosis solutions is required to effectively address defects occurring between tiers with varying processing nodes.

V. INDUSTRY PRACTICES

In recent years, companies have shown increased interest in 3D IC testing. Numerous developments have occurred with respect to commercial tools and products, which highlight the crucial role of testing in ensuring the reliability of 3D IC designs [100]. In this section, we provide an overview of industrial practices in 3D IC testing.

Synopsys has been actively engaged in advancing 3D IC testing. In [101], pre-bond and post-bond testing scenarios are proposed for an illustrative three-die stacked device. In the pre-bond testing scenario, integrating KGD testing into the 3D IC test flow promises to reduce test costs. However, a challenge arises in the absence of probe pads for KGD testing on non-bottom dies, given that all I/Os are accessible solely through TSVs. One possible solution involves incorporating sacrificial probe pads specifically for KGD testing. By leveraging test features in test compression environments, the number of such probe pads and area overhead can be minimized. Power-aware testing and 3D DFT implementation are also major challenges in 3D IC testing. In [101], existing tools demonstrate the potential to overcome these challenges by extending current flows designed for 2D designs in a 3D context.

Siemens EDA offers insights into the challenges and opportunities associated with 3D IC testing [20]. To address the testing of individual dies and interconnects between dies, Siemens EDA proposes Tesson Multi-die to support the IEEE P1838 standard (see Section III-D) and extend the IEEE 1149.1 and the IEEE 1687 for multi-die devices to generate compatible test access ports (TAPs) and the IJTAG network [102]. Emphasizing that test access is crucial for successful 3D IC testing, Siemens EDA addresses challenges related to communication between dies, establishing compatibility criteria for dies from different processes, and the testing of 3D IP cores. The Siemens 3D IC Design Flow covers aspects from physical design to post-silicon monitoring, which facilitates thoughtful planning interactions at the die and package levels for physical design, power analysis, packaging, and test hardware.

Advantest has incorporated 3D IC testing solutions into their automated test equipment (ATE) [103]. Their SmarTest Program Manager enables collaborative software program development

TABLE III: Comparisons among M3D IC testing and diagnosis solutions.

Methodology	Target	Proposed solutions	Accomplishments	Costs	Applications
[23], [86]	Power-safe testing	ILP-based/SA-based pattern reshaping	PSN-induced yield loss elimination	Test time increase / loss of fault coverage	Logic-on-logic design
[88]		RL-based TPI	Power reduction & test coverage co-optimization	Area overhead	
[89]		Scan segmentation using RL	ATPG compatibility / no post-processing	Area and test time overhead	
[87]	Memory diagnosis	March diagnosis sequence	Fault-origin identification	Additional write & read operations	M3D-integrated RRAM memory-on-logic design
[93], [94]	MIV BIST	XOR- and Dual-BIST architecture	MIV fault detection	Area overhead	Logic-on-logic design
[97], [98]	Fault diagnosis	GNN-based fault localization	Tier-level resolution / diagnosis quality improvement	Loss of diagnosis accuracy	
[99]	Fault detection & localization	NodeRank OP insertion	Tier-level resolution	Area overhead	

for their products. The ATE, featuring concurrent test and multiport capabilities, allows the simultaneous testing of multiple dies, significantly reducing overall test time. Moreover, Advantest's ATE accommodates up to 128 asynchronous clock domains. This is a crucial advantage for future 3D IC designs with stacked dies hosting multiple IP cores, each requiring distinct clock frequencies and domains. The investigation of TSV testing and the utilization of microelectro mechanical system (MEMS) probes are important for reliable 3D interconnects and the thinned wafer's resilience against force and pressure during the probing process, respectively.

TSMC's 3DFabric, a platform of 3D silicon stacking and advanced packaging technologies, is an outcome of efforts by a consortium of companies, including Advantest, Cadence, Keysight, proteanTecs, Siemens EDA, Synopsys, and Teradyne [104]. The collective effort aims to address 3D testing challenges, mitigate yield losses, and enhance power delivery efficiency. These collaborations play a crucial role in expediting the development of 3D testing, which is essential for the volume production of 3D systems. However, there are additional challenges that must be addressed before 3D IC is ripe for commercial exploitation. Therefore, there is a need for increased research efforts and industrial practices to advance solutions for 3D IC testing.

VI. ADDITIONAL CHALLENGES IN 3D IC TESTING

Power management is a major concern for 3D IC testing. Due to the reduced footprint, a small number of power bumps need to supply numerous devices in different dies and tiers, leading to an increase in current density. This issue is more severe in the test mode than in the functional mode because of excessive switching activity. In addition, dissipating heat in 3D ICs becomes challenging as non-bottom dies can only release heat through 3D interconnects (i.e., TSVs and MIVs). While effective solutions have been proposed to target heat dissipation and power reduction for 3D ICs, they mainly focus on optimizations during functional operations. There is a lack of exploration into dedicated solutions specifically for 3D IC testing.

Test access is another challenge for 3D ICs. For non-bottom dies and tiers, test data can only be accessed by TSVs and MIVs. While DfT solutions have been explored to enhance the testability of these dies, the associated area overhead increases proportionally with the number of 3D interconnects. With advanced integration technologies, a significant number of 3D interconnects are anticipated to fully leverage the advantages of 3D integration. Moreover, in gate-level and transistor-level partitioned designs, standard cells in a functional block are distributed among different tiers. DfT algorithms relying on wrapper cells, enabling separate testing of each tier, become ineffective due to the dual usage of TSVs and MIVs for both I/Os between functional blocks and for signal propagation within functional blocks. Testing each tier separately is insufficient to comprehensively assess the functionality of 3D ICs. These challenges demonstrate the inadequacy of existing solutions in ensuring the reliability of future 3D designs. More advanced DfT solutions are needed to enhance the 3D testability in a cost-effective manner.

In addition to devices within dies and tiers, ensuring the reliability of 3D interconnects is a crucial aspect of 3D ICs. Unlike conventional BEOL vias, TSVs and MIVs need to penetrate through silicon substrates and inter-tier dielectrics. Therefore, they are susceptible to defects induced by voids and contaminations within the substrates and dielectrics. The proximity of TSV pairs and MIV pairs can lead to coupling effects, leading to the degradation in circuit performance or even malfunctions. Specific testing solutions for 3D interconnects are essential to ensure their reliability. Inserting BIST structures is an effective way to test 3D interconnects; however, the area overhead caused by BIST logic becomes severe with an increasing number of TSVs and MIVs. Developing an efficient approach to enhance the testability of 3D interconnects remains a challenging task.

Heterogeneous integration has the potential to minimize manufacturing costs by enabling the integration of dies from various vendors. However, the testing complexity significantly increases design after stacking and bonding because each die has its

own DfT structures. Different dies may be fabricated using diverse technology nodes to optimize the overall performance of the 3D design. Detecting faults induced by process variations among dies becomes a critical challenge. To fully harness the advantages of 3D integration in the post-Moore era, there is a crucial need to develop a DfT standard that incorporates all DfT designs across all the dies. Such a standard is significantly important for the creation of a high-performance and highly reliable 3D IC.

REFERENCES

- [1] W.-W. Shen and K.-N. Chen. Three-dimensional integrated circuit (3D IC) key technology: Through-silicon via (TSV). *Nanoscale research letters*, 12:1–9, 2017.
- [2] W. Chen and B. Bottoms. Heterogeneous integration roadmap: Driving force and enabling technology for systems of the future. In *Symposium on VLSI Technology*, pages T50–T51, 2019.
- [3] J. P. Gambino, S. A. Adderly, and J. U. Knickerbocker. An overview of through-silicon-via technology and manufacturing challenges. *Microelectronic Engineering*, 135:73–106, 2015.
- [4] J. Kim and Y. Kim. HBM: Memory solution for bandwidth-hungry processors. In *IEEE Hot Chips Symposium (HCS)*, pages 1–24, 2014.
- [5] D. B. Ingerly, S. Amin, L. Aryasomayajula, A. Balankutty, D. Borst, A. Chandra, K. Cheemalapati, C. S. Cook, R. Criss, K. Enamul, W. Gomes, D. Jones, K. C. Kolluru, A. Kandas, G.-S. Kim, H. Ma, D. Pantuso, C.F. Petersburg, M. Phen-givoni, A. M. Pillai, A. Sairam, P. Shekhar, P. Sinha, P. Stover, A. Telang, and Z. Zell. Foveros: 3D integration and the use of face-to-face chip stacking for logic devices. In *IEEE International Electron Devices Meeting*, pages 19.6.1–19.6.4, 2019.
- [6] H. Tsugawa, H. Takahashi, R. Nakamura, T. Umebayashi, T. Ogita, H. Okano, K. Iwase, H. Kawashima, T. Yamasaki, D. Yoneyama, J. Hashizume, T. Nakajima, K. Murata, Y. Kanaishi, K. Ikeda, K. Tatani, T. Nagano, H. Nakayama, T. Haruta, and T. Nomoto. Pixel/DRAM/logic 3-layer stacked CMOS image sensor technology. In *IEEE International Electron Devices Meeting*, pages 3.2.1–3.2.4, 2017.
- [7] R. Agarwal, P. Cheng, P. Shah, B. Wilkerson, R. Swaminathan, J. Wu, and C. Mandalapu. 3D packaging for heterogeneous integration. In *IEEE Electronic Components and Technology Conference*, pages 1103–1107, 2022.
- [8] P. Batude, T. Ernst, J. Arcamone, G. Arndt, P. Coudrain, and P. Gaillardon. 3D sequential integration: A key enabling technology for heterogeneous co-integration of new function with CMOS. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, 2(4):714–722, 2012.
- [9] S. K. Samal, K. Samadi, P. Kamal, Y. Du, and S. K. Lim. Full chip impact study of power delivery network designs in monolithic 3D ICs. In *IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, pages 565–572, 2014.
- [10] M. Joodaki. Uprising nano memories: Latest advances in monolithic three dimensional (3D) integrated Flash memories. *Microelectronic Engineering*, 164:75–87, 2016.
- [11] T.-T. Wu, C.-H. Shen, J.-M. Shieh, W.-H. Huang, H.-H. Wang, F.-K. Hsueh, H.-C. Chen, C.-C. Yang, T.-Y. Hsieh, B.-Y. Chen, Y.-S. Shiao, C.-S. Yang, G.-W. Huang, K.-S. Li, T.-J. Hsueh, C.-F. Chen, W.-H. Chen, F.-L. Yang, M.-F. Chang, and W.-K. Yeh. Low-cost and TSV-free monolithic 3D-IC with heterogeneous integration of logic, memory and sensor analogy circuitry for Internet of Things. In *IEEE International Electron Devices Meeting (IEDM)*, pages 25.4.1–25.4.4, 2015.
- [12] Y. Yu and N. K. Jha. SPRING: A sparsity-aware reduced-precision monolithic 3D CNN accelerator architecture for training and inference. *IEEE Transactions on Emerging Topics in Computing*, pages 1–1, 2020.
- [13] G. Murali, X. Sun, S. Yu, and S. K. Lim. Heterogeneous mixed-signal monolithic 3-D in-memory computing using resistive RAM. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 29(2):386–396, 2021.
- [14] E. Beyne. The 3-D interconnect technology landscape. *IEEE Design Test*, 33(3):8–20, 2016.
- [15] R. Lu, Y.-C. Chuang, J.-L. Wu, and J. He. Reliability challenges from 2.5D to 3DIC in advanced package development. In *IEEE International Reliability Physics Symposium (IRPS)*, pages 1–4, 2023.
- [16] T. Lu, C. Serafy, Z. Yang, S. K. Samal, S. K. Lim, and A. Srivastava. TSV-based 3-D ICs: Design methods and tools. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 36(10):1593–1619, 2017.
- [17] F. Laermer and A. Schilp. Plasma polymerizing temporary etch stop, 1996. US Patent US5,501,893.
- [18] Integrity 3D-IC platform. Web Link. [Online] Available: https://www.cadence.com/content/dam/cadence-www/global/en_US/documents/tools/digital-design-signoff/Integrity-3D-IC-DS-v1.pdf
- [19] 3D IC: opportunities, challenges, and solutions. Web Link. [Online] Available: <https://semiengineering.com/3d-ic-opportunities-challenges-and-solutions/>
- [20] Shifting left for earlier testing in 2.5D and 3D IC design. Web Link. [Online] Available: <https://blogs.sw.siemens.com/semiconductor-packaging/2023/03/02/shifting-left-for-earlier-testing-in-3d-ic-design/>
- [21] C.-T. Ko and K.-N. Chen. Reliability of key technologies in 3D integration. *Microelectronics Reliability*, 53(1):7–16, 2013.
- [22] P. Batude, L. Brunet, C. Fenouillet-Beranger, F. Andrieu, J.-P. Colinge, D. Lattard, E. Vianello, S. Thuries, O. Billoint, P. Vivet, C. Santos, B. Mathieu, B. Sklenard, C.-M. V. Lu, J. Micout, F. Deprat, E. Avelar Mercado, F. Ponthenier, N. Rambal, M.-P. Samson, M. Cassé, S. Hentz, J. Arcamone, G. Sicard, L. Hutin, L. Pasini, A. Ayres, O. Rozeau, R. Berthelon, F. Nemouchi, P. Rodriguez, J.-B. Pin, D. Larmagnac, A. Duboust, V. Ripoche, S. Barraud, N. Allouti, S. Barnola, C. Vizioz, J.-M. Hartmann, S. Kerdiles, P. Acosta Alba, S. Beaurepaire, V. Beugin, F. Fournel, P. Besson, V. Loup, R. Gassilloud, F. Martin, X. Garros, F. Mazen, B. Previtali, C. Euvrard-Colnat, V. Balan, C. Combroure, M. Zussy, Mazzocchi, O. Faynot, and M. Vinet. 3D sequential integration: Application-driven technological achievements and guidelines. In *IEEE International Electron Devices Meeting*, pages 3.1.1–3.1.4, 2017.
- [23] S.-C. Hung, Y.-C. Lu, S. K. Lim, and K. Chakrabarty. Power supply noise-aware at-speed delay fault testing of monolithic 3-D ICs. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 29(11):1875–1888, 2021.
- [24] L. Brunet, C. Fenouillet-Beranger, P. Batude, S. Beaurepaire, F. Ponthenier, N. Rambal, V. Mazzocchi, J.-B. Pin, P. Acosta-Alba, S. Kerdiles, P. Besson, H. Fontaine, T. Lardin, F. Fournel, V. Larrey, F. Mazen, V. Balan, C. Morales, C. Guerin, V. Joussemae, X. Federspiel, D. Ney, X. Garros, A. Roman, D. Scevola, P. Perreau, F. Kouemeni-Tchouake, L. Arnaud, C. Scibetta, S. Chevalliez, F. Aussenac, J. Aubin, S. Reboh, F. Andrieu, S. Maitrejean, and M. Vinet. Breakthroughs in 3D sequential technology. In *IEEE International Electron Devices Meeting (IEDM)*, pages 7.2.1–7.2.4, 2018.
- [25] S. Panth, K. Samadi, Y. Du, and S. K. Lim. Design and cad methodologies for low power gate-level monolithic 3D ICs. In *IEEE/ACM International Symposium on Low Power Electronics and Design (ISLPED)*, pages 171–176, 2014.
- [26] B. W. Ku, K. Chang, and S. K. Lim. Compact-2D: A physical design methodology to build commercial-quality face-to-face-bonded 3D ICs. In *Proceedings of the International Symposium on Physical Design*, pages 90–97, 2018.
- [27] Y.-C. Lu, S. Pentapati, L. Zhu, K. Samadi, and S. K. Lim. TP-GNN: a graph neural network framework for tier partitioning in monolithic 3D ICs. In *ACM/IEEE Design Automation Conference (DAC)*, pages 1–6, 2020.
- [28] M. Raghu and E. Schmidt. A survey of deep learning for scientific discovery. *arXiv preprint arXiv:2003.11755*, 2020.
- [29] M. M. Waldrop. The chips are down for moore's law. *Nature News*, 530(7589):144, 2016.
- [30] D. Ivanovich, C. Zhao, X. Zhang, R. D. Chamberlain, A. Deliwala, and V. Gruev. Chip-to-chip optical data communications using polarization division multiplexing. In *2020 IEEE High Performance Extreme Computing Conference (HPEC)*, pages 1–8. IEEE, 2020.
- [31] P. Ambs. Optical computing: A 60-year adventure. *Advances in Optical Technologies*, 2010.
- [32] R. A. Athale and W. C. Collins. Optical matrix–matrix multiplier based on outer product decomposition. *Applied optics*, 21(12):2089–2090, 1982.
- [33] L. Chrostowski and M. Hochberg. *Silicon photonics design: from devices to systems*. Cambridge University Press, 2015.
- [34] Y. Zhang, A. Samanta, K. Shang, and S. J. B. Yoo. Scalable 3d silicon photonic electronic integrated circuits and their applications. *IEEE Journal of Selected Topics in Quantum Electronics*, 26(2), 2020.

[35] K. Shang, S. Pathak, B. Guan, G. Liu, and S. J. B. Yoo. Low-loss compact multilayer silicon nitride platform for 3d photonic integrated circuits. *Optics Express*, 23(16):21334–21342, 2015.

[36] Y. Zhang, Y.-C. Ling, Y. Zhang, K. Shang, and S. J. B. Yoo. High-density wafer-scale 3-d silicon-photonic integrated circuits. *IEEE Journal of Selected Topics in Quantum Electronics*, 24(6):1–10, 2018.

[37] S. Banerjee, M. Nikdast, and K. Chakrabarty. Modeling silicon-photonic neural networks under uncertainties. In *2021 Design, Automation & Test in Europe Conference & Exhibition (DATE)*. IEEE, 2021.

[38] S. Banerjee, M. Nikdast, and K. Chakrabarty. On the impact of uncertainties in silicon-photonic neural networks. *IEEE design test of computers*, 2022.

[39] S. Banerjee, M. Nikdast, and K. Chakrabarty. Characterizing coherent integrated photonic neural networks under imperfections. *Journal of Lightwave Technology*, 41(5):1464–1479, 2022.

[40] A. Shafiee, S. Banerjee, K. Chakrabarty, S. Pasricha, and M. Nikdast. Loci: An analysis of the impact of optical loss and crosstalk noise in integrated silicon-photonic neural networks. In *Proceedings of the Great Lakes Symposium on VLSI 2022*, pages 351–355, 2022.

[41] S. Banerjee, M. Nikdast, S. Pasricha, and K. Chakrabarty. Champ: Coherent hardware-aware magnitude pruning of integrated photonic neural networks. In *2022 Optical Fiber Communications Conference and Exhibition (OFC)*. IEEE, 2022.

[42] S. Banerjee, M. Nikdast, S. Pasricha, and K. Chakrabarty. Pruning coherent integrated photonic neural networks using the lottery ticket hypothesis. In *2022 IEEE Computer Society Annual Symposium on VLSI (ISVLSI)*, pages 128–133. IEEE, 2022.

[43] E. J. Marinissen. Testing TSV-based three-dimensional stacked ICs. In *Design, Automation Test in Europe Conference Exhibition (DATE)*, pages 1689–1694, 2010.

[44] K. Chakrabarty, S. Deutsch, H. Thapliyal, and F. Ye. TSV defects and TSV-induced circuit failures: The third dimension in test and design-for-test. In *IEEE International Reliability Physics Symposium (IRPS)*, pages 5F.1.1–5F.1.12, 2012.

[45] K. Xu, Y. Yu, and X. Fang. The detection of open and leakage faults for prebond TSV test based on weak current source. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 41(9):2768–2779, 2022.

[46] B. Noia and K. Chakrabarty. Pre-bond probing of through-silicon vias in 3-D stacked ICs. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 32(4):547–558, 2013.

[47] D. K. Maity, S. K. Roy, and C. Giri. Identification of faulty TSVs in 3D IC during pre-bond testing. In *31st International Conference on VLSI Design and 17th International Conference on Embedded Systems (VLSID)*, pages 109–114, 2018.

[48] B. Zhang and V. D. Agrawal. An optimal probing method of pre-bond TSV fault identification in 3D stacked ICs. In *SOI-3D-Subthreshold Microelectronics Technology Unified Conference (S3S)*, pages 1–3, 2014.

[49] B. Zhang and V. D. Agrawal. Diagnostic tests for pre-bond TSV defects. In *28th International Conference on VLSI Design*, pages 387–392, 2015.

[50] D. K. Maity, S. K. Roy, and C. Giri. Identification of random/clustered TSV defects in 3D IC during pre-bond testing. *Journal of Electronic Testing*, 35:741–759, 2019.

[51] S. Deutsch and K. Chakrabarty. Non-invasive pre-bond TSV test using ring oscillators and multiple voltage levels. In *Design, Automation Test in Europe Conference Exhibition (DATE)*, pages 1065–1070, 2013.

[52] Y. Fkhi, P. Vivet, B. Rouzeyre, M.-L. Flottes, and G. Di Natale. A 3D IC BIST for pre-bond test of TSVs using ring oscillators. In *IEEE 11th International New Circuits and Systems Conference (NEWCAS)*, pages 1–4. IEEE, 2013.

[53] N. Georgoulopoulos and A. Hatzopoulos. Effectiveness evaluation of the TSV fault detection method using ring oscillators. In *6th International Conference on Modern Circuits and Systems Technologies (MOCAST)*, pages 1–4, 2017.

[54] Chang H. and Liang H. Pulse shrinkage based pre-bond through silicon vias test in 3D ic. In *IEEE 33rd VLSI Test Symposium (VTS)*.

[55] M. Yi, J. Bian, T. Ni, C. Jiang, H. Chang, and H. Liang. A pulse shrinking-based test solution for prebond through silicon via in 3-D ICs. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 38(4):755–766, 2019.

[56] S. M. Menon and R. R. Roeder. No-touch stress testing of memory I/O interfaces, December 30 2014. US Patent 8,924,786.

[57] S. Das, F. Su, and S. Chakrabarty. A PVT-resilient no-touch DFT methodology for prebond tsv testing. In *IEEE International Test Conference (ITC)*, pages 1–10, 2018.

[58] D. Arumí, R. Rodríguez-Montaños, and J. Figueras. Prebond testing of weak defects in TSVs. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 24(4):1503–1514, 2016.

[59] E. J. Marinissen. Challenges in testing TSV-based 3D stacked ICs: test flows, test contents, and test access. In *IEEE Asia Pacific Conference on Circuits and Systems*, pages 544–547. IEEE, 2010.

[60] Y.-J. Huang, J.-F. Li, J.-J. Chen, D.-M. Kwai, Y.-F. Chou, and C.-W. Wu. A built-in self-test scheme for the post-bond test of TSVs in 3D ICs. In *29th VLSI test symposium*, pages 20–25. IEEE, 2011.

[61] D. Maity, S. Roy, C. Giri, and H. Rahaman. Identification of faulty TSV with a built-in self-test mechanism. In *IEEE 27th Asian Test Symposium (ATS)*, pages 1–6. IEEE, 2018.

[62] D. K. Maity, S. K. Roy, and C. Giri. Built-in self-repair for manufacturing and runtime TSV defects in 3D ICs. In *IEEE International Test Conference India*, pages 1–6. IEEE, 2020.

[63] M. Cho, C. Liu, D. H. Kim, S. K. Lim, and S. Mukhopadhyay. Pre-bond and post-bond test and signal recovery structure to characterize and repair TSV defect induced signal degradation in 3-D system. *IEEE Transactions on Components, Packaging and Manufacturing Technology*, 1(11):1718–1727, 2011.

[64] F. Ye and K. Chakrabarty. TSV open defects in 3D integrated circuits: Characterization, test, and optimal spare allocation. In *Proceedings of the 49th Annual Design Automation Conference*, pages 1024–1030, 2012.

[65] H. Sung, K. Cho, K. Yoon, and S. Kang. A delay test architecture for TSV with resistive open defects in 3-D stacked memories. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 22(11):2380–2387, 2013.

[66] Y.-W. Lee, H. Lim, and S. Kang. Grouping-based TSV test architecture for resistive open and bridge defects in 3-D-ICs. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 36(10):1759–1763, 2016.

[67] S.-G. Papadopoulos, V. Gerakis, Y. Tsiatouhas, and A. Hatzopoulos. Oscillation-based technique for post-bond parallel testing and diagnosis of multiple tsvs. In *27th International Symposium on Power and Timing Modeling, Optimization and Simulation (PATMOS)*, pages 1–6, 2017.

[68] R. Rodríguez-Montaños, D. Arumí, and J. Figueras. Postbond test of through-silicon vias with resistive open defects. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 27(11):2596–2607, 2019.

[69] Y. Yu, X. Fang, and X. Peng. A post-bond TSV test method based on RGC parameters measurement. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 39(2):506–519, 2018.

[70] W.-H. Lo, K. Chi, and T. Hwang. Architecture of ring-based redundant tsv for clustered faults. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 24(12):3437–3449, 2016.

[71] L. Jiang, Q. Xu, and B. Eklow. On effective through-silicon via repair for 3-d-stacked ics. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 32(4):559–571, 2013.

[72] I. Lee, M. Cheong, and S. Kang. Highly reliable redundant tsv architecture for clustered faults. *IEEE Transactions on Reliability*, 68(1):237–247, 2018.

[73] T. Ni, D. Liu, Q. Xu, Z. Huang, H. Liang, and A. Yan. Architecture of cobweb-based redundant tsv for clustered faults. *IEEE transactions on very large scale integration (VLSI) systems*, 28(7):1736–1739, 2020.

[74] T. Ni, Y. Yao, H. Chang, L. Lu, H. Liang, A. Yan, Z. Huang, and X. Wen. Lchr-tsv: Novel low cost and highly repairable honeycomb-based tsv redundancy architecture for clustered faults. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 39(10):2938–2951, 2019.

[75] IEEE Computer Society, New York, NY, USA. *IEEE Std 1149.1TM-2013, IEEE Standard for Test Access Port and Boundary-Scan Architecture*, 2013.

[76] Institute of Electrical and Electronics Engineers. *IEEE Standard Testability Method for Embedded Core-based Integrated Circuits*, 8 2005. [Online]. Available: <https://doi.org/10.1109/IEEESTD.2005.96290>.

[77] E. J. Marinissen, T. McLaurin, and H. Jiao. IEEE Std P1838: Dft standard-under-development for 2.5 D-, 3D-, and 5.5 D-SICs. In *21th IEEE european test symposium (ETS)*, pages 1–10. IEEE, 2016.

[78] A. Cron and E. J. Marinissen. IEEE standard 1838 is on the move. *Computer*, 54(11):88–94, 2021.

[79] J. Michel, J. Liu, and L. C Kimerling. High-performance Ge-on-Si photodetectors. *Nature photonics*, 4(8):527–534, 2010.

[80] J. De Coster, P. De Heyn, M. Pantouvaki, B. Snyder, H. Chen, E. J. Marinissen, P. Absil, J. Van Campenhout, and B. Bolt. Test-station for flexible semi-automatic wafer-level silicon photonics testing. In *IEEE European Test Symposium (ETS)*. IEEE, 2016.

[81] V. Grimaldi, F. Zanetto, F. Toso, C. De Vita, and G. Ferrari. Non-invasive light sensor with enhanced sensitivity for photonic integrated circuits. In *2022 17th Conference on Ph. D Research in Microelectronics and Electronics (PRIME)*, pages 285–288. IEEE, 2022.

[82] F. Morichetti, S. Grillanda, M. Carminati, G. Ferrari, M. Sampietro, M. J. Strain, M. Sorel, and A. Melloni. Non-invasive on-chip light observation by contactless waveguide conductivity monitoring. *IEEE Journal of Selected Topics in Quantum Electronics*, 20(4):292–301, 2014.

[83] K. Chang, S. Das, S. Sinha, B. Cline, G. Yeric, and S. K. Lim. System-level power delivery network analysis and optimization for monolithic 3D ICs. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 27(4):888–898, 2019.

[84] A. Koneru, A. Todri-Sanial, and K. Chakrabarty. Reliable power delivery and analysis of power-supply noise during testing in monolithic 3D ICs. In *IEEE VLSI Test Symposium*, pages 1–6, 2019.

[85] S.-C. Hung and K. Chakrabarty. Design of a reliable power delivery network for monolithic 3D ICs. In *Design, Automation Test in Europe Conference Exhibition (DATE)*, pages 1746–1751, 2020.

[86] S.-C. Hung, Y.-C. Lu, S. K. Lim, and K. Chakrabarty. Power supply noise-aware scan test pattern reshaping for at-speed delay fault testing of monolithic 3D ICs. In *IEEE Asian Test Symposium*, pages 1–6, 2020.

[87] S.-C. Hung, A. Chaudhuri, S. Banerjee, and K. Chakrabarty. Fault diagnosis for resistive random-access memory and monolithic inter-tier vias in monolithic 3D integration. In *IEEE International Test Conference*, pages 118–127, 2022.

[88] S.-C. Hung, A. Chaudhuri, and K. Chakrabarty. Test-point insertion for power-safe testing of monolithic 3D ICs using reinforcement learning. In *IEEE European Test Symposium*, pages 1–6, 2023.

[89] S.-C. Hung, A. Chaudhuri, Banerjee S., and K. Chakrabarty. Scan cell segmentation based on reinforcement learning for power-safe testing of monolithic 3D ICs. In *IEEE International Test Conference*, pages 1–10, 2023.

[90] C. Xu, D. Niu, Y. Zheng, S. Yu, and Y. Xie. Impact of cell failure on reliable cross-point resistive memory design. *ACM Transactions on Design Automation of Electronic Systems*, 20(4), 2015.

[91] E. Esmanhotto, L. Brunet, N. Castellani, D. Bonnet, T. Dalgatay, L. Grenouillet, D. R. B. Ly, C. Cagli, C. Vizioz, N. Allouti, F. Laulagnet, O. Gully, N. Bernard-Henriques, M. Bocquet, G. Molas, P. Vivet, D. Querlioz, JM. Portal, S. Mitra, F. Andrieu, C. Fenouillet-Beranger, E. Nowak, and E. Vianello. High-density 3D monolithically integrated multiple 1T1R multi-level-cell for neural networks. In *IEEE International Electron Devices Meeting*, pages 36.5.1–36.5.4, 2020.

[92] Z. Jiang, S. Yu, Y. Wu, J. H. Engel, X. Guan, and H.-S. P. Wong. Verilog-A compact model for oxide-based resistive random access memory (RRAM). In *International Conference on Simulation of Semiconductor Processes and Devices*, pages 41–44, 2014.

[93] A. Chaudhuri, S. Banerjee, J. Kim, S. Lim, and K. Chakrabarty. Built-in self-test of high-density and realistic ILV layouts in monolithic 3-D ICs. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 31(03):296–309, 2023.

[94] A. Chaudhuri, S. Banerjee, H. Park, B. W. Ku, K. Chakrabarty, and S.-K. Lim. Built-in self-test for inter-layer vias in monolithic 3D ICs. In *IEEE European Test Symposium (ETS)*, pages 1–6, 2019.

[95] A. Mallik, A. Vandooren, L. Witters, A. Walke, J. Franco, Y. Sherazi, P. Weckx, D. Yakimets, M. Bardon, B. Parvais, P. Debacker, B. W. Ku, S. K. Lim, A. Mocuta, D. Mocuta, J. Ryckaert, N. Collaert, and P. Raghavan. The impact of sequential-3D integration on semiconductor scaling roadmap. In *IEEE International Electron Devices Meeting*, pages 32.1.1–31.1.4, 2017.

[96] A. Koneru, S. Kannan, and K. Chakrabarty. Impact of electrostatic coupling and wafer-bonding defects on delay testing of monolithic 3D integrated circuits. *ACM Journal on Emerging Technologies in Computing Systems*, 13(4), 2017.

[97] S.-C. Hung, S. Banerjee, A. Chaudhuri, J. Kim, S. K. Lim, and K. Chakrabarty. Transferable graph neural network-based delay-fault localization for monolithic 3-D ICs. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 42(11):4296–4309, 2023.

[98] S.-C. Hung, S. Banerjee, A. Chaudhuri, and K. Chakrabarty. Graph neural network-based delay-fault localization for monolithic 3D ICs. In *Design, Automation & Test in Europe Conference & Exhibition*, pages 448–453, 2022.

[99] A. Chaudhuri, S. Banerjee, and K. Chakrabarty. NodeRank: Observation-point insertion for fault localization in monolithic 3D ICs. In *IEEE Asian Test Symposium (ATS)*, pages 1–6, 2020.

[100] J. H. Lau. 3D IC integration and 3D IC packaging. In *Semiconductor Advanced Packaging*, pages 343–378. Springer, 2021.

[101] Test automation of 3D integrated systems. Web Link. [Online] Available: <https://api.semanticscholar.org/CorpusID:259654823>

[102] Tessent multi-die. Web Link. [Online] Available: <https://static.swcdn.siemens.com/siemens-disw-assets/public/5VeJ6OifkulPXkZlfNFMK/en-US/Siemens-SW-Tessent-multi-die-FS-84857-C2.pdf>

[103] ATE solutions to 3D-IC test challenges. Web Link. [Online] Available: <https://www3.advantest.com/documents/11348/6f2b1d79-526e-4637-8ba8-9b746bd49618>

[104] TSMC 3DFabric™ alliance. Web Link. [Online] Available: https://www.tsmc.com/english/dedicatedFoundry/oip/3dfabric_alliance