Test Challenges for 3D Integrated Circuits

Hsien-Hsin S. Lee
Georgia Institute of Technology

Krishnendu Chakrabarty
Duke University

Editor's note:
One of the challenges for 3D technology adoption is the insufficient understanding of 3D testing issues and the lack of DFT solutions. This article describes testing challenges for 3D ICs, including problems that are unique to 3D integration, and summarizes early research results in this area.
—Yuan Xie, Pennsylvania State University

3D ICs promise to overcome barriers in interconnect scaling by leveraging fast, dense interdie vias, thereby providing an opportunity for continued higher performance using CMOS.[1] In addition, 3D ICs also enable the integration of heterogeneous fabrication processes on the same chip to make the form factor more compact, leading to a true SoC. However, 3D technology adoption is hampered by an insufficient understanding of 3D testing issues and by the lack of DFT techniques. Test techniques and DFT solutions for 3D ICs have remained largely unexplored in the research community, even though experts in industry have identified several test challenges related to the lack of probe access for wafers, test access to modules in stacked wafers and dies, thermal concerns, design testability, test economics, and new defects that arise from unique processing steps such as wafer thinning, alignment, and bonding. Despite recent advances in architectures,[2] design automation tools,[3] and yield enhancement techniques,[4] today's 3D chip designs do not consider test cost and the implications of design decisions on testability, thereby leading to a gap between anticipated (perceptual) benefits and practical value. Therefore, it is not surprising that, among all EDA challenges for 3D IC design, tools and methodologies for 3D IC testing are regarded as the number-one challenge.

Our goal in this article is to identify emerging test challenges for 3D ICs and to provide an overview of early and ongoing work on test strategies for 3D chips. Figure 1, for example, highlights the state of the art in 3D integration. It positions testing in the context of ongoing research activities. The bottom of the triangle in Figure 1a, which shows the manufacturing aspects specific to 3D ICs performed mostly by industry, serves as the foundation for research in design automation and architectures that explore 3D ICs. As Figure 1b shows, although 3D IC manufacturing technology is almost ready, certain design methodology challenges and most of the test issues remain largely unexplored.

3D ICs: Manufacturing

Researchers are investigating various 3D IC manufacturing processes that are particularly relevant to testing and DFT. In terms of the process and the level of assembly that 3D ICs require, we can broadly classify the techniques as monolithic or as die stacking.

Monolithic 3D ICs

For the monolithic manufacturing process using epitaxy, multiple device layers are grown on the same wafer in a serial manner. Once a layer of devices and their associated interconnect are completed, an isolation interlevel dielectric layer (for example, SiO_2) can be deposited and polished to allow another layer of devices and interconnect to continue to grow vertically. To electrically connect devices across separate processed layers, 3D vias are etched through the isolation layer, and metal fillings are deposited. The same process is repeated to fabricate a 3D IC consisting of multiple layers of devices.

Figure 2a shows the cross-section of a monolithic 3D IC with three device layers. Jung et al. have demonstrated such a technique by growing the pull-down NMOS transistors, the load PMOS transistors, and the pass NMOS transistors separately in a series of

0740-7475/09/$26.00 © 2009 IEEE Copublished by the IEEE CS and the IEEE CASS

Authorized licensed use limited to: Georgia Institute of Technology. Downloaded on November 21, 2009 at 00:51 from IEEE Xplore. Restrictions apply.

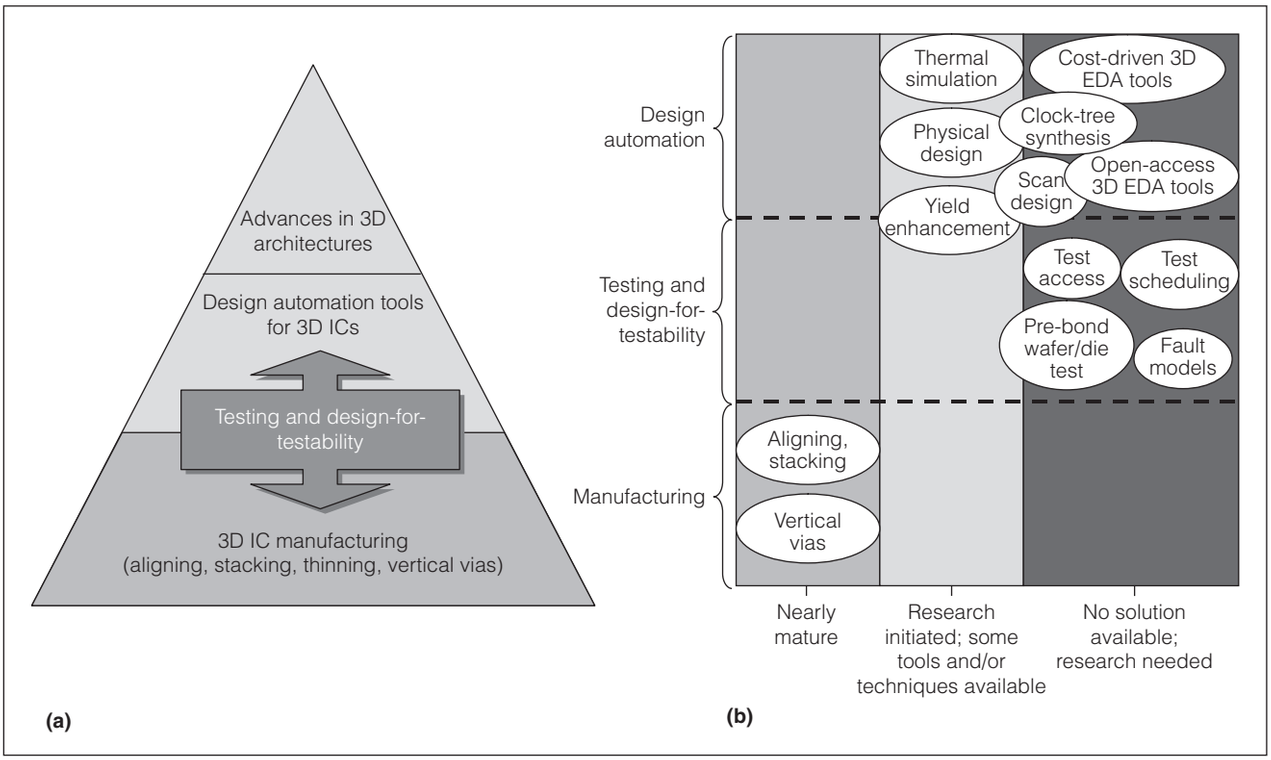

Figure 1. 3D IC testing and related challenges in the context of 3D integration: Role of 3D IC test (a); status of 3D IC R&D (b).

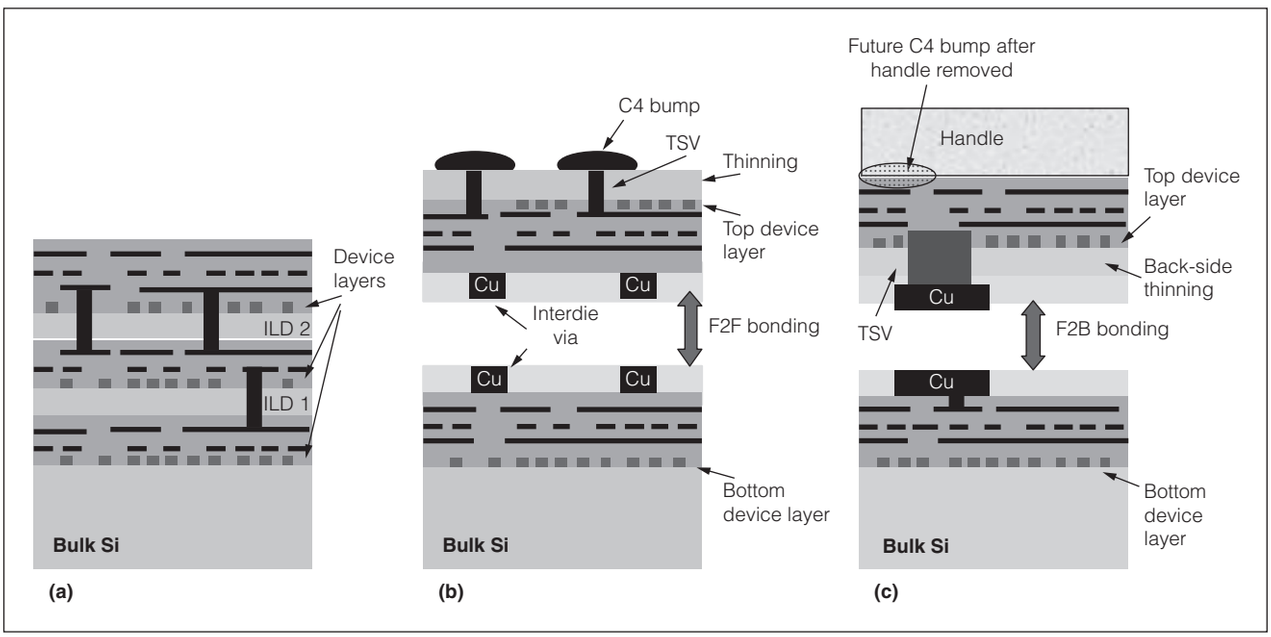

Figure 2. 3D IC fabrication methods: monolithic fabrication (a), face-to-face bonding (b), and face-to-back bonding (c). (F2B: face-to-back; F2F: face-to-face; ILD: interlayer dielectrics; TSV: through-silicon via.)

Authorized licensed use limited to: Georgia Institute of Technology. Downloaded on November 21, 2009 at 00:51 from IEEE Xplore. Restrictions apply.

processes.[5] Designed with a monolithic manufacturing process, the cell areas consumed by n-wells were completely eliminated, whereas the cross-coupling interconnect wires were substituted with 3D vias—the designers having thereby achieved an SRAM implementation of high density.

Die stacking

Another 3D integration technique is to stack individual 2D die layers vertically. In contrast to monolithic 3D manufacturing, which may require many changes in current process facilities, fabricating 3D ICs using die stacking technology can minimize the impact of altering existing manufacturing technology and equipment. With 3D die stacking, the candidate dies to be integrated onto the same package can be designed and manufactured separately, just as they are with a regular, existing 2D planar process—with additional manufacturing processes of substrate thinning and through-silicon via (TSV) filling, if needed. Then they are bonded together by precise alignment of interdie vias and the application of thermocompression. In general, die stacking presents three integration alternatives: wafer to wafer, die on wafer, and die on die, each with their respective pros and cons from a cost or yield perspective.

Depending on the stacking orientation—whether face-to-face or face-to-back bonding—the interdie vias that connect different die layers will have different fabrication procedures and requirements. With face-to-face bonding, the via stubs from each metal layer of two candidate wafers are bonded directly. The only TSVs required to go through the silicon bulk are those for I/O connections, as Figure 2b shows. Most of the fabrication steps will remain unchanged from the conventional 2D ones except for the final phase that requires the application of chemical-mechanical polishing (CMP) to thin the back side of the top die for exposing buried TSVs to connect to the I/O pads. Although this technique incurs minimal changes in the fabrication process, it is not scalable, and so limits the integration to a two-die-layer stack.

An alternate, yet complementary, stacking approach is to bond die layers face-to-back, wherein the backside silicon bulk of the top wafer is first thinned by CMP down to 10 to 100 microns, depending on the process of technology vendors, before applying thermocompression bonding with the front side of the bottom wafer. The 3D TSVs are etched and filled with tungsten plugs on the thinned back side, either using a via-first or a via-last process sequence. Figure 2c depicts this bonding technique with a via-last process applied after the back-end-of-line (BEOL) process. Because this fabrication technique involves stacking dies in a repeatable, uniform manner (face-to-back), it can be applied for stacking an arbitrary number of die layers, as long as other considerations such as thermal dissipation, power supply, and cost can be kept within bounds.

Compared to die stacking, the main advantages of a monolithic 3D process are the high via density, a potentially lower mask count, and reduced area. Unlike the die-stacking method, the via density of a monolithic process can scale at the same rate as the feature size. Reductions in both mask count and area lower nonrecurring engineering costs, although the manufacturing technology itself might need a more elaborate process and possibly new materials. In contrast, die stacking can leverage the conventional 2D process to a large extent and enable the integration of heterogeneous devices, realizing a true SoC, which is infeasible using epitaxy technology. Moreover, it is difficult to achieve good crystallized layers of silicon with any known deposition techniques when fabricating monolithic 3D structures, which results in poor electrical characteristics of transistors on monolithic 3D circuits and prevents them from achieving high-performance operation. Because the current trend in industry is in favor of die stacking, our subsequent discussions will focus on this area.

Test challenges for 3D integration

In traditional IC manufacturing, wafers are probed and individual dies tested (a process referred to as wafer sort) before they are packaged. In 3D integration, we are confronted with new challenges before bonding wafers. The yield of 3D ICs can be increased if we can bond pretested dies, or if we can sort the wafers first and stack matched dies (based on the speed or power consumption level) on top of each other. However, prebond testing of wafers in 3D integration faces several serious obstacles, as we explain here.[6]

Wafer probing

The connection from the tester to the wafer in wafer sort is made through a hardware interface called a probe card. The probing needles in probe cards are major performance limiters and cost

Authorized licensed use limited to: Georgia Institute of Technology. Downloaded on November 21, 2009 at 00:51 from IEEE Xplore. Restrictions apply.

contributors due to impedance discontinuity for high-frequency signals; therefore, tests must be applied at lower frequency using DFT techniques such as structural test and scan design. DFT techniques also lead to higher structural fault coverage. However, wafer sort does not prevent bad dies from being stacked on top of good dies in wafer stacking. Moreover, the wafer probe itself is a problem of face-to-face bonding. The bottom die has up to hundreds of thousands of copper pads, but their small size and large number make probing of signals difficult. The top wafer would be hard to be probed from the copper side—the TSVs are buried and C4 bump pads are not fabricated prior to bonding.

In face-to-back stacking, the top die is more testable than the bottom because the C4 bump pads can be fabricated on the top layer. However, the top wafer must be thinned, which introduces the problems of ultrathin wafer processing and limits the ease with which the wafer can be probed. Typically, the probe card applies weight in the range from 3 to 10 g per probe. Therefore, the probe weight per wafer can be as high as 60 to 120 kg, which is a serious issue for thinned wafers. Although it is unlikely that comprehensive wafer-probing solutions for wafer stacking will emerge in the near term,[6] efforts are under way in industry to address these challenges through techniques such as contactless testing and proximity I/O based on near-field wireless communication, inductive coupling, and capacitive coupling.

Known-good dies

The known-good-die (KGD) problem has been addressed in the past for package-level integration—for example, with multichip modules (MCMs) and systems in packages (SiPs). As in MCMs, a major concern in 3D integration is the quality of the incoming bare (unpackaged) dies and wafers prior to stacking. Testing is also needed to ensure the structural integrity and performance of the stacked modules. However, MCM testing and 3D IC testing differ considerably in KGD problems. First, wafer probing is a major limitation for thinned wafers and prebond wafers in 3D integration, so testing must be accomplished with very few probe contacts and few wafer touchdowns. Second, because of design partitioning in 3D ICs (both at the architectural and circuit levels), any given layer in a stacked 3D IC does not always include complete functionality. Consequently, the tests that can be applied at this stage are limited. Moreover, 3D ICs offer smaller form factors and higher performance than MCMs, so test solutions for 3D ICs must target more speed- and layout-related defects that arise at nanoscale dimensions. Test methods for 3D ICs must also handle more structural constraints. For example, limits on the number of TSVs and the availability of I/Os only on the layer with the C4 pads give rise to important test access constraints. Finally, the processing steps for 3D ICs gives rise to new failure mechanisms.

New defect types

Because the processing steps for 3D ICs are significantly different than for 2D ICs, new defect mechanisms (unique to 3D integration) must be addressed as part of a test strategy. The thinning, alignment, and stacking of the wafers add extra steps to the manufacturing process. During bonding, any foreign particle caught between the wafers can lead to peeling, as well as delamination, which dramatically reduces bonding quality and yield. In addition, to maintain good conductivity and minimize resistance, the interconnect TSVs and micropads between wafers must be precisely aligned. Edge effects must also be considered in 3D IC testing because the edges between bonded wafers are more susceptible to chipping, peeling, and delamination, a problem unique to 3D IC manufacturing. Cracking can occur during the stacking process, especially because of loading forces, back-side grinding, and die thinning. Finally, random open defects can result from dislocations, oxygen trapped on the surface, voids formation, and mechanical failures.

Thermal and power-delivery considerations

Thermal problems are greatly exacerbated in 3D ICs. It is a major challenge to maintain low junction temperatures, both during functional mode and in test mode. The oxides encapsulating each layer make upper tiers increasingly vulnerable to thermal effects. Thermal effects are therefore more serious problems in 3D ICs than in today's planar ICs. Thermal fatigue is also a potential problem for the C4 bumps. Hence, test patterns must be generated to target hot regions on different layers as well as the hot TSVs. Thermal problems are accompanied by power delivery concerns in 3D ICs (which can be viewed as two sides of the same coin). Test patterns are needed that can accurately evaluate the *IR* drop and *di*/*dt* effects. Although similar testing problems

Authorized licensed use limited to: Georgia Institute of Technology. Downloaded on November 21, 2009 at 00:51 from IEEE Xplore. Restrictions apply.

are being addressed for today's ICs, new methods for high-density 3D ICs require better scalability.

SoC test access and test scheduling

Modular testing, which is based on test access mechanisms (TAMs) and IEEE Std 1500 core test wrappers, provides a low-cost solution to the test access problem for a SoC; many I/O and scan terminals for the embedded cores can be accessed from a few chip pins. For today's 2D ICs, several optimization techniques have been reported in the literature for test infrastructure design to minimize test time. Similar techniques are needed for 3D ICs, but we are now confronted with an even more difficult test access problem: the embedded cores in a 3D IC might be on different layers, and even the same embedded core could have blocks placed on different layers. Only a limited number of TSVs can be reserved for use by the TAM. Although many TSVs can be integrated in a 3D IC, most are required for power, clock, and signal lines, and the need for a "keep-out" area requires optimization techniques that make judicious use of TSVs for test access. Wrapper design and optimization must also go beyond IEEE 1500 and consider how a core on multiple layers can be wrapped under TSV constraints. Finally, thermal constraints must be considered in TAM optimization and test scheduling. Prior work on thermal-constrained SoC test scheduling is not appropriate in this context, because heat spreading in the vertical dimension is limited, thermal TSVs allow more efficient heat distribution, and different thermal estimation techniques for 3D ICs must be used.

Economics of test

When deciding to adopt this emerging technology as a mainstream design approach, chip vendors must consider the cost of 3D integration. All the advantages of 3D ICs ultimately must be translated into cost-effectiveness when a design strategy must be determined. Testing is a key factor that affects the final IC product cost, and it could be a major portion of the total IC cost. In 3D IC design, various testing strategies and different integration methods could affect the testing cost dramatically, and the interaction with other cost factors could result in different trade-offs. For example, wafer-to-wafer stacking could save the KGD testing cost and improve the throughput, compared to die-to-wafer stacking; however, the overall yield might be far lower, so the total IC cost could be higher. Therefore, it is important to estimate test economics with a test cost model. It is crucial to develop a test cost model for 3D ICs, analyze the trade-offs associated with test strategies for complex 3D IC chips, and integrate them into a cost-driven 3D IC design flow to strike a balance between cost and other benefits (such as performance, power, and area).

KGD wafer-level test and burn-in

To make 3D ICs commercially viable, there is a pressing need for prebond test techniques to let bare dies be tested to isolate the KGDs prior to stacking. Accelerated test techniques shorten the time to failure for defective parts without altering the device failure characteristics. Burn-in is one such technique that is widely used in the semiconductor industry. Wafer-level test during burn-in (WLTBI) has recently emerged as an enabling technology to lower the cost of burn-in. It lets us obtain KGDs at low cost for die stacking, so it's especially relevant for 3D integration. In this approach, devices are subjected to burn-in and electrical testing while in the bare wafer form. Since the early 2000s, WLTBI has been adopted at several semiconductor companies—for example, Motorola (now Freescale Semiconductor), Samsung, and Panasonic. Companies such as Aehr Test Systems are marketing WLTBI testers to semiconductor companies. These full-wafer parallel-test systems can test thousands of chips simultaneously while providing traditional burn-in. Because burn-in takes several hours, this otherwise idle time can be used to apply functional and structural tests. Test equipments from Aehr Test Systems allow such tests under the higher than normal voltage and temperature environments during the burn-in period.

In the built-in test, burn-in method typically used, chip designers employ on-chip DFT infrastructure to achieve WLTBI. This technique allows all the dies on the wafer to be simultaneously contacted, with few probe contacts per die. The presence of sophisticated built-in DFT features on modern ICs makes *monitored burn-in* possible. Monitored burn-in is a process in which a device under test (DUT) is provided with input test patterns; the output responses of the DUT are monitored online, thereby leading to the identification of failing devices. Therefore, WLTBI has significant potential to lower test cost by removing the barrier between burn-in and test processes. With advances in probe technology for wafer processing in 3D ICs, it is also conceivable

Authorized licensed use limited to: Georgia Institute of Technology. Downloaded on November 21, 2009 at 00:51 from IEEE Xplore. Restrictions apply.

that WLTBI will alleviate some of the prebond test problems for wafer stacking.

Design for 3D IC testability

A key incentive for vendors of complex circuits such as microprocessors to adopt 3D integration is to gain the benefits of short latency, low power consumption, and an immense amount of bandwidth delivered by TSVs. Researchers have explored different architecture-stacking styles to unleash their maximal potential for 3D ICs. Most styles focused on novel mechanisms for partitioning and folding microarchitectural modules, memory components, and caches at different granularities to take advantage of the high-density TSVs. Researchers have examined the impact of access latency by simply folding the bit lines or word lines of a cache; for example, Puttaswamy studied more aggressive partitioning techniques to mitigate the area impact by implementing a heavily ported register file using a 3D port-splitting technique, which, in essence, stacks up area-consuming wires across multiple layers.[2] To a finer granularity at the logic level, researchers from MIT Lincoln Laboratory partitioned the chained inverters composing a ring oscillator into two wafer layers.[7] More recently, Park et al. showed how a bit-sliced 3D partitioning scheme can improve both the thermal gradient and performance for the hotspots caused by highly switched microarchitectural modules and on-chip routers.[8]

Prebond testability

Although these architectural partitioning methods attempted to maximize the benefits of 3D IC designs, they simultaneously introduced new manufacturing challenges and implications. To make 3D IC designs commercially viable, the most critical step is in the final integration—which must ensure that only KGDs will be bonded and packaged. As the number of 3D layers increases, a random bonding strategy will never be economically practical, because it could decrease the overall yield exponentially. To effectively address this issue, designers must ensure that each individual die layer is designed to be testable before bonding takes place. The primary challenge posed by many of these architectural partitioning proposals is that each layer, prior to bonding, can contain only partial circuits of the entire system. For example, in a planar processor design, the instruction queue, the ALU, and the physical register file are logically connected in series on a 2D floorplan, whereas with 3D partitioning they can be placed in different layers to take advantage of the communication latency from interdie vias. In the 3D case, for instance, the instruction queue and physical register file could be placed on different dies without being connected before bonding. Worse yet, circuit-level partitioning techniques such as 3D port-splitting will lead to functionally broken circuits before the bonding step as some layers will consist only of stacked wires and port-access NMOS transistors without the actual SRAM memory cells.[2] Enabling prebond tests for such 3D IC designs will require completely reconsidering conventional DFT strategies.

For a 3D processor partitioned at the microarchitectural level, each module is still complete (for example, a 32-bit adder); it's not unthinkable to test each layer containing these modules in their integral form except that the conventional scan chain structure used in a planar implementation requires certain restructuring. In fact, similar test strategies based on design segmentation have been proposed and used in the Alpha 21364. The goal of the modules' segmentation is to reduce the complexity of testing the entire design in order to improve test efficiency and throughput. In this framework, each segment forms a test island, isolated from its neighboring modules with specially tailored border registers. These border registers control the data flow during test mode (which closes the borders) and normal operation mode (which opens the borders). Accordingly, incoming test vectors can be provided directly during test mode by closing the island borders.

Consider a 3D design partitioned at functional module granularity: each module or subset of modules on one die layer can be treated as an isolated test island. Prebond test can be carried out by testing each island separately, with deliberately designed scan chains managed by a local layer test controller.[9] An LTC is designed to be interfaced to the ATE via probes for prebond tests. After bonding layers together, the LTC on each layer will connect each individual test structure to one standard IEEE 1149.1 JTAG port, implemented at only one particular die layer, for performing postbond test of the complete bonded circuit.

Figure 3 illustrates a scan-island-based design of a two-die-layered 3D IC, in which each scan chain on each die layer can link disjoint (that is, functionally not connected) microarchitectural modules.

Authorized licensed use limited to: Georgia Institute of Technology. Downloaded on November 21, 2009 at 00:51 from IEEE Xplore. Restrictions apply.

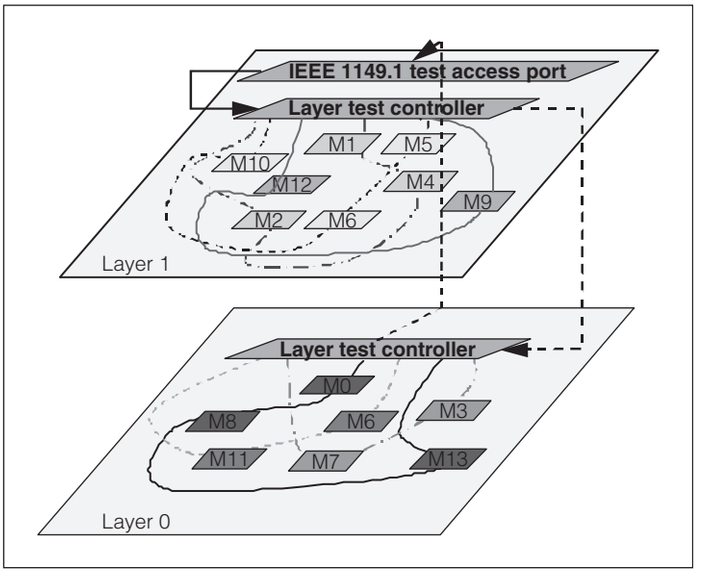

Figure 3. 3D scan-island design. (TAP: test access port.)

(In Figure 3, these are denoted as M1, M2, and so on.) Furthermore, a physical interface must be provided for probing during prebond test. Depending on a wafer's stacking orientation, the interface is implemented either through the interdie vias from the front side (top metal layer) for face-to-face bonding, or through the TSVs on the back side (thinned silicon bulk) for face-to-back bonding. Several challenges, however, complicate the design of test pads for supporting basic nets such as data signals, power, ground, and skew-free clocks. First, to ensure good quality of electrical conduction, a test probe's contact force might be too stressful for thinned wafers to survive. Second, the test pads can occupy a much larger area compared to the via's submicron diameter. Note that one single test pad can consume an area equivalent to hundreds of front-side vias, substantially diminishing the benefit of exploiting via density. Moreover, these prebond test pads will functionally become useless postbond, but could introduce unwanted extra capacitance for the clock tree. One way to take advantage of these test pads is to arrange them deliberately during the 3D floorplanning phase such that they could be reused as decoupling capacitance to mitigate the inductive noise.[9]

For circuit-level partitioning, the scan island-based approach is no longer applicable for prebond tests, because each layer essentially consists of incomplete circuits and cannot carry out the desired function. So, a drastically different DFT strategy is needed. By closely examining the previously proposed circuit-level partitioning schemes, we may find that the partitioning cut lines of a functional block do not follow a random pattern. They are either separated by grouping and slicing the bits of the same functional unit such as an adder to mitigate the thermal gradient,[8] or by moving an entire memory port to a different layer to reduce access latency and total chip area.[2]

For bit-sliced 3D partitioning, enabling prebond test involves no more than adding scan registers at the via side of each layer. After bonding, the path for supplying test data should be gated off. Even though this comes at the cost of extra die area, the total accumulative time for testing all bit-sliced circuit layers can be much faster than for its full-width counterpart because the number of bit patterns required to test each bit-sliced layer prebond is reduced exponentially.

A more challenging case is the partitioning style using port-splitting applied to memory-based structures, in which only one die layer has the access to the actual SRAM cells prior to bonding. Obviously, we can employ the existing memory test methods such as marching 0s and 1s to test the layer that contains the memory cells. For the other layers, however, new methods are required if we are to cope with their incomplete functionality. For the port-splitting design, in particular, prebond testability must be taken into account during the logic design phase. For instance, we can design at least one read port and one write port for each individual layer to enable prebond test. As Figure 4 shows, it's possible to place test data on the bit lines of the write port and then have the data immediately read back by the read port of the same layer as if the data were being supplied by a memory cell from a virtual device layer through interdie vias.

Reliability and testing of TSVs postbond

The TSVs are key components of a 3D IC and are used for providing power, as well as clock and functional signals. In addition, they also provide test access to logic blocks on different layers. Fabricated TSV dimensions have been reported in the range of 0.4 microns (small) and 200 microns (large), and process improvements can allow manufacturers to integrate more than several hundreds of thousands of TSVs in a single package.

One critical issue that requires taking DFT into account for 3D IC design is the assembly yield of the TSV postbond. Partly because of imperfect TSV

Authorized licensed use limited to: Georgia Institute of Technology. Downloaded on November 21, 2009 at 00:51 from IEEE Xplore. Restrictions apply.

etching (for example, underfill), ragged wafer surface, foreign particle contamination, and potential wafer misalignment, a certain number of TSVs in one wafer after thinning and polishing might not be completely exposed or aligned for making good contact with their counterparts on the other wafer. Nevertheless, TSVs make up a key test infrastructure. Any defective TSV will prevent test access of certain logic blocks. Under these circumstances, even a single TSV defect between any two layers can void the entire chip stack, reducing the overall yield.

To guarantee the full functionality of 3D TSV interconnects, designers must take the precaution of implementing redundancy and self-repairing mechanisms for each TSV—especially those for transmitting signals. TSVs for power delivery and clock distribution could be more defect tolerable as long as the power or clock distribution network remains a connected graph, and skew or current needed to drive loads are unaffected in the presence of defects. Self-repairing interconnects can be achieved by inserting redundant TSVs along with one or a cluster of the required TSVs and their selection logic.

During the test phase, the test structure (for instance, scan chains) will diagnose TSV defects, and the selection logic will isolate defects, reconfigure, and reroute the interconnects. Moreover, because TSVs take up chip area, especially because of the need for a keep-out area, the numbers and locations of redundant TSVs must be carefully evaluated and determined. Depending on the estimated TSV yield, several TSVs can share the same redundant TSV via a multiplexer or a crossbar on the assumption that not all of them would fail simultaneously to minimize the area overhead.

Testable clock tree design

To leverage the multifaceted benefits of latency, power, and high density from interdie vias, researchers have also proposed methods to maximize the use of these vias in place of long wires. In addition to reducing the wire length of global interconnects for generic functional data signals, these vias could be massively used to minimize the wire length and power for forming a zero-skew clock tree for a 3D IC as well. By doing so, the clock signals are directly relayed from other die layers by routing themselves through the interdie vias, reducing traveling distance by several orders of magnitude. Although this approach improves the efficiency of constructing the clock distribution network—that is, better routability—it inadvertently introduces another problem that makes prebond testability much more challenging. The results of such 3D routing algorithms often generate a clock distribution network, in which most of the die layers would contain a plethora of isolated clock islands that are not fully connected.

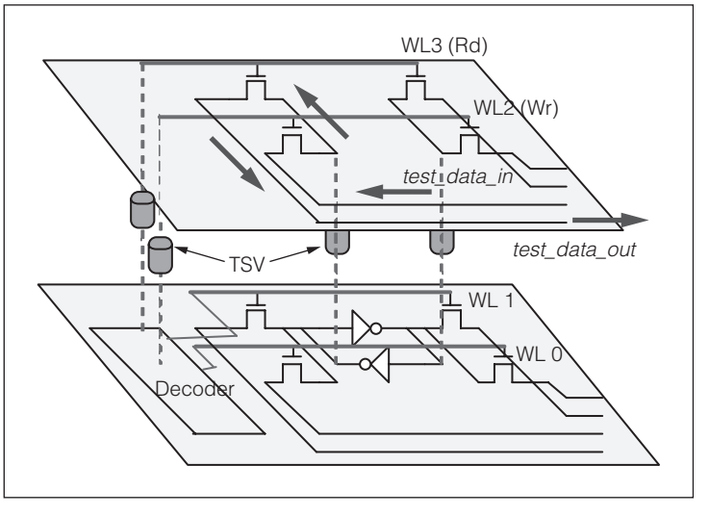

Figure 4. 3D memory port-splitting (four ports, two for each die layer). (WL: word line.)

Testing each prebond layer requires multiple synchronous clock input sources—that is, using multiple test probes to feed those disjoint clock islands during the prebond test phase. Nonetheless, applying multiple test probes could also undermine the requirement of maintaining a zero-skew clock tree. As discussed earlier, prebond test is indispensable for the commercial success of 3D ICs, due to the yield concern. As a result, with respect to 3D clock tree design, there are two goals to achieve. First, each prebond layer under test should contain a fully testable, zero-skew clock tree. Second, a zero-skew clock tree also must be guaranteed for the final integrated 3D chip after bonding.

One simple way to address this problem yet still take advantage of via routing is to employ a redundancy design by providing a configurable, complete 2D clock tree for each layer for prebond test purposes, while simultaneously maintaining a 3D clock routing tree for normal operations postbond. The 2D trees are merely used for prebond test and will be gated off afterward. One drawback of this approach is that the routability could be worsened rather than improved as a result of the redundant clock trees. However, this approach will not be

Authorized licensed use limited to: Georgia Institute of Technology. Downloaded on November 21, 2009 at 00:51 from IEEE Xplore. Restrictions apply.

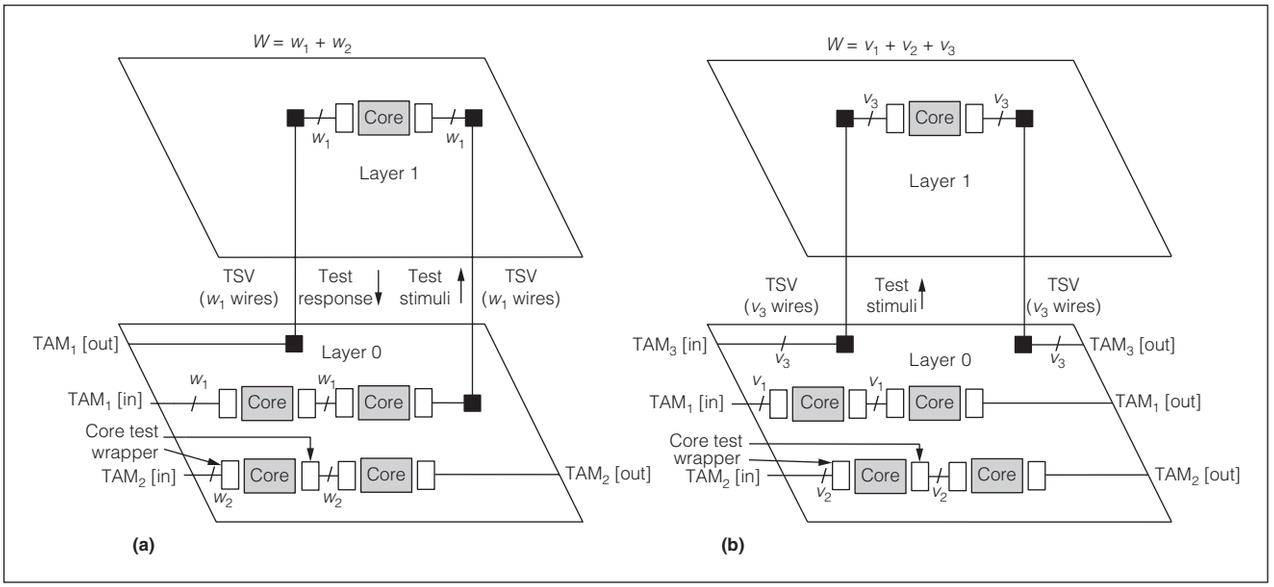

Figure 5. Two scenarios for test access mechanism (TAM) optimization for a 3D-core-based SoC: Approach 1 (a); Approach 2 (b).

worse than its 2D planar counterpart, which consumes these areas to plan its complete clock tree. Note that this method does not automatically solve the potential skew issues for both pre- and postbond trees. Therefore, new algorithms must still be developed to simultaneously guarantee the zero-skew clock networks for both scenarios.

Test access optimization

Test access is a major problem for stacked 3D ICs. Test access is provided to the scan and I/O terminals of logic cores, memory BIST controllers, TAP controllers, and other test structures. Recent research on this issue has addressed the dual problems of designing scan chains for cores placed on multiple layers[10] and the optimization of the TAM for a core-based 3D IC.[11]

Consider the example shown in Figure 5, which shows five (wrapped) cores in a 3D IC. Four cores are placed in Layer 0, and the fifth core is placed on Layer 1. Suppose a total of $2W$ channels (that is, number of wires from the TAM) are available from the tester to access the cores, out of which W channels must be used to transport test stimuli, and the remaining W channels must be used to collect the test responses. We also have an upper limit on the number of TSVs that can be used for the test access infrastructure. The TAM design needs to be optimized to minimize the test application time.

Therefore, the objective here is to allocate wires (bit widths) to the different TAMs used to access the cores. In Figure 5a, we have two TAMs of widths w_1 and w_2 ($w_1 + w_2 = W$)—and the core on Layer 1 is accessed using the first TAM. A total of $2w_1$ TSVs are needed for this test infrastructure design. Figure 5b presents an alternative design in which three TAMs are used, and the core of Layer 1 is accessed using the third TAM ($v_1 + v_2 + v_3 = W$). Note that both $2w_1$ and $2v_3$ are no more than the number of TSVs dedicated for test access. Optimization methods and design tools are needed to determine a test access architecture that leads to the minimum test time under the constraints of wire length and the number of TSVs. Wu et al. presented techniques to address this problem.[11] An alternative technique does not impose any limits on the number of TSVs used for the TAM, but considers prebond test considerations and wire-length limits.[12] Continued advances are needed to incorporate thermal constraints that limit the amount of test parallelism, and the effective use of prebond test access hardware for testing after assembly.

3D ICs HAVE EMERGED as a key enabling technology to extend the scaling trajectory predicted by Moore's law. However, their success and commercial viability will depend to a significant extent on their production yields and the corresponding test solutions for

Authorized licensed use limited to: Georgia Institute of Technology. Downloaded on November 21, 2009 at 00:51 from IEEE Xplore. Restrictions apply.

guaranteeing them—which, at this point, are left largely unexplored in the research community. In this article, we have presented several challenges and ongoing research efforts regarding testing 3D ICs. From wafer-probing techniques, test access mechanisms, KGD screening, and prebond DFT techniques to understanding new types of defects caused by TSV filling, alignment, and bonding, a broad range of open issues have a pressing need for novel ideas. Research solutions are needed to holistically address these test challenges before 3D integration technology can be widely adopted and become a market success. ■

Acknowledgments

We thank Yuan Xie from Penn State University, and Sungkyu Lim and Gabriel Loh from the Georgia Institute of Technology. We also thank PhD students Xiaoxia Wu and Yibo Chen (Penn State), Dean Lewis and Xin Zhao (Georgia Tech), and Sudarshan Bahukudumbi (Duke), for their contributions to various aspects of this article. This work is supported in part by the C2S2 center of the Focus Center Research Program and National Science Foundation contract CCF-0811738.

■ References

1. K. Banerjee et al., "3-D ICs: A Novel Chip Design for Improving Deep-Submicrometer Interconnect Performance and Systems-On-Chip Integration," *Proc. IEEE,* vol. 89, no. 5, 2001, pp. 602-633.
2. K. Puttaswamy, "Designing High-Performance Microprocessors in 3-D Integration Technology," doctoral dissertation, School of Electrical and Computer Engineering, Georgia Inst. of Technology, 2007.
3. B. Goplen and S. Sapatnekar, "Placement of 3D ICs with Thermal and Interlayer via Considerations," *Proc. 44th Design Automation Conf.* (DAC 07), ACM Press, 2007, pp. 626-631.
4. C. Ferri, S. Reda, and I. Bahar, "Parametric Yield Management for 3D ICs: Models and Strategies for Improvement," *ACM J. Emerging Technologies in Computing Systems,* vol. 4, no. 4, 2008, article no. 19.
5. S.-M. Jung et al., "The Revolutionary and Truly 3-Dimensional 25F^2 SRAM Technology with the Smallest S^3 (Stacked Single-Crystal Si) Cell, 0.16um^2, and SSTFT (Stacked Single-Crystal Thin Film Transistor) for Ultra High Density SRAM," *Proc. Symp. VLSI Technology,* IEEE Press, 2004, pp. 228-229.
6. T.M. Mak, "Testing of 3D Circuits," *Handbook of 3D Integration: Technology and Applications Using 3D Integrated Circuits,* P. Garrou, C. Bower, and P. Ramm, eds., Wiley-CVH, 2008.
7. J.A. Burns et al., "A Wafer-Scale 3-D Circuit Integration Technology," *IEEE Trans. Electron Devices,* vol. 53, no. 10, 2006, pp. 2507-2516.
8. D. Park et al., "MIRA: A Multi-layered On-Chip Interconnect Router Architecture," *Proc. 35th Int'l Symp. Computer Architecture* (ISCA 08), IEEE CS Press, 2008, pp. 251-261.
9. D.L. Lewis and H.-H.S. Lee, "A Scan-Island Based Design Enabling Pre-bond Testability in Die-Stacked Microprocessors," *Proc. Int'l Test Conf.,* IEEE CS Press, 2007, pp. 1-8.
10. X. Wu et al., "Scan-Chain Design and Optimization for Three-Dimensional Integrated Circuits," *ACM J. Emerging Technologies in Computing Systems,* vol. 5, no. 2, 2008, article no. 9.
11. X. Wu et al., "Test-Access Mechanism Optimization for Core-Based Three-Dimensional SoCs," *Proc. IEEE Int'l Conf. Computer Design,* IEEE CS Press, 2008, pp. 212-218.
12. L. Jiang, L. Huang, and Q. Xu, "Test Architecture Design and Optimization for Three-Dimensional SoCs," *Proc. IEEE/ACM Design, Automation and Test in Europe Conf.* (DATE 09), IEEE CS Press, 2009, pp. 220-225.

Hsien-Hsin S. Lee is an associate professor in the School of Electrical and Computer Engineering at the Georgia Institute of Technology. His research interests include computer architecture, 3D integration, and cybersecurity. Lee has a PhD in computer science and engineering from the University of Michigan, Ann Arbor. He is a senior member of the IEEE and a member of the ACM.

Krishnendu Chakrabarty is a professor of electrical and computer engineering at Duke University. His research interests include VLSI testing, digital microfluidic biochips, circuits based on DNA self-assembly, and wireless sensor networks. Chakrabarty has a PhD in computer science and engineering from the University of Michigan, Ann Arbor. He is a Fellow of the IEEE and a Distinguished Engineer of the ACM.

■ Direct questions and comments about this article to Hsien-Hsin S. Lee, School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332; leehs@gatech.edu.

Authorized licensed use limited to: Georgia Institute of Technology. Downloaded on November 21, 2009 at 00:51 from IEEE Xplore. Restrictions apply.