A Scan-Island Based Design Enabling Pre-bond Testability in Die-Stacked Microprocessors

Dean L. Lewis   Hsien-Hsin S. Lee

School of Electrical and Computer Engineering
Georgia Institute of Technology
Atlanta, GA 30332
TEL: (404) 894-9483
{dean, leehs}@ece.gatech.edu

ABSTRACT

Die stacking is a promising new technology that enables integration of devices in the third dimension. Recent research thrusts in 3D-integrated microprocessor design have demonstrated significant improvements in both power consumption and performance. However, this technology is currently being held back due to the lack of test technology. Because processor functionality is partitioned across different silicon die layers, only partial circuitry exists on each layer pre-bond. In current 3D manufacturing, layers in the die stack are simply bonded together to form the complete processor; no testing is performed at the pre-bond stage. Such a strategy leads to an exponential decay in the yield of the final product and places an economic limit on the number of die that can be stacked.

To overcome this limit, pre-bond test is a necessity. In this paper, we present a technique to enable pre-bond test in each layer. Further, we address several issues with integrating this new test hardware into the final design. Finally, we use a sample 3D floorplan based on the Alpha 21264 to show that our technique can be implemented at a minimal cost (0.2% area overhead). Our design for pre-bond testability enables the structural test necessary to continue 3D integration for microprocessors beyond a few layers.

Keywords

Design-For-Testability, Die Stacking, 3D integration

1. INTRODUCTION

In a continuing effort to keep up with the relentless march of Moore's law, processor designers keep pushing the limits of technology further and further. Unfortunately, each push inevitably costs more than the last — more money and more time. To make matters worse, the returns from each push are steadily diminishing. Each new technology generation consumes more power, becomes less reliable, and fails to achieve an ideal performance improvement. Consequently, researchers continue to seek out innovative new technologies orthogonal to technology shrinks. 3D integration — also known as *die stacking* — is a very promising technology that enables IC design (as complex as a microprocessor design) in the third dimension, continuing the scaling trajectory predicted by Moore's Law for a few more generations.

For 3D die-stacked microprocessors, prior research thrusts proposed and studied several methods for partitioning their functions [1, 2, 3, 4, 5, 6, 7, 8]. These partitioning schemes range from simple die-stacking of memory chips on top of a processor to partitioning a microarchitectural block or even a single circuit (such as an adder) across different die layers. All these techniques aim to reap the benefits made available by the shortened wire lengths of 3D integration. While they appear to be feasible as demonstrated, one major challenge has yet to be addressed, i.e., how do we test these individual die separately prior to bonding the layers together? Note that, without a viable pre-bond testing strategy, manufacturing yield will decay exponentially with the number of layers integrated, washing out all of benefits of 3D integration. To address this concern, a systematic and generic design-for-testability (DFT) method needs to be realized at the early architecting stage, which is the primary goal of this work. To the best of our knowledge, this paper is the first one that proposes and evaluates an applicable methodology to enable pre-bond testability for 3D die-stacked microprocessors.

The rest of this paper is organized as follows. Section 2 gives an overview of 3D integration and its application to microprocessors; Section 3 explores the motivation for enabling pre-bond test and the associated challenges; Section 4 will lay out the general architecture of our test strategy and address some important related issues; Section 5 presents our experimental setup and results; Section 6 concludes the paper with a summary and discussion of results.

2. OVERVIEW OF 3D-IC TECHNOLOGY

3D integration is an emerging technology that allows semiconductor die to be bound together to form a tightly integrated stack. Opening design to the third dimension provides several advantages. First, it enables the integration of heterogeneous components such as logic and DRAM memory [2] or analog and digital circuits [9]. Secondly, it increases routability [10]. Last but not least, it can substantially reduce wire length, which contributes to long communication latency and high power consumption. Recent work in this field has already demonstrated significant improvements in both performance and power consumption [11] and lead to other interesting applications, such as online profiling [12] and network-in-memory [4]. Even greater returns are expected as researchers further explore the opportunities afforded.

2.1 3D Die Bonding

Figure 1 shows simple two-layer die stacks. The two die

1-4244-1128-9/07/$25.00 ©2007 IEEE

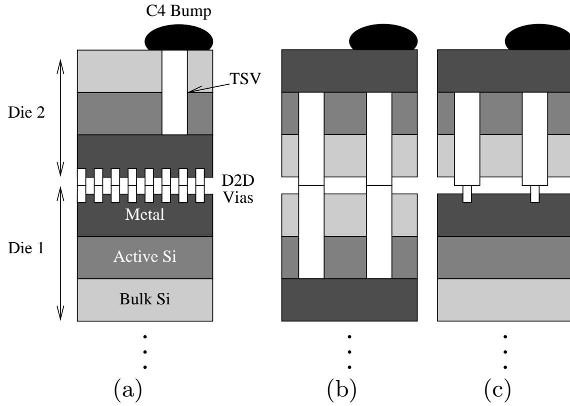

Figure 1: Three die stacks, each comprised of two layers using three possible bond styles: (a) face-to-face, (b) back-to-back, and (c) face-to-back

communicate through an array of die-to-die (D2D) vias, which come in two flavors: faceside and backside. Faceside vias are manufactured on top of the metal layers with size and pitch on the order of a few microns [13]. Backside vias — also called *Through Silicon Vias (TSVs)* — are etched through the active and bulk silicon with size and pitch on the order of tens of microns. Exposing backside vias requires that the die be thinned from several hundred microns to only a few tens of microns thick. With these vias exposed, the die can be bound to the other die in the stack [1]. There are three possible bonds: face-to-face (Figure 1(a)), back-to-back (Figure 1(b)), and face-to-back (Figure 1(c)). Face-to-face is the superior interface because it enables a significantly higher via density. However, back-to-back and face-to-back interfaces are required to stack beyond two layers.

Utilizing these different interface options, designers continue to push further into the third dimension. Some embedded applications already utilize a die stack with eight layers [14]. Given the disparity between faceside and backside via densities, face-to-face bonds are more appropriate for small granularity partitions while back to back bonds are better suited to coarse-grained partitions. When the stack is complete, the requisite C4 solder bumps can be placed on the TSVs of a backside layer (as shown in Figure 1) or on the top metal layer of a faceside layer (just as in planar designs).

2.2 3D Partitioning Granularity

Die stack technology may be used to partition a design at three general levels of granularity. The coarsest level is the technology level. Disparate technologies like high-speed CMOS and high-density DRAM both have their own dedicated and highly-optimized manufacturing processes. Many problems arise when attempting to integrate such technologies onto a single die, requiring sophisticated manufacturing tricks to achieve economically viable integration quality [15]. Die stacking allows each technology to be manufactured on its own layer in its own process. After each layer is manufactured, a separate integration process bonds these layers together. The result is the best of both worlds: each layer is manufactured at the highest possible quality level and, simultaneously, the two technologies are tightly integrated. This improves both the performance of the system and the form factor.

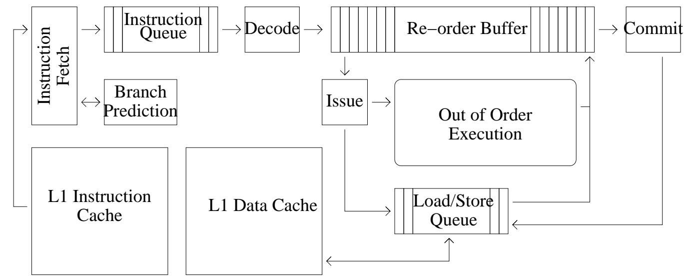

Figure 2: A generic out-of-order processor architecture.

The next finer level of partitioning is the architectural level. Unlike technology partitioning, both layers are manufactured using the same process. The goal of architectural partitioning is to spread the functional blocks of a design across the available layers in such a way as to minimize the length of the interconnect buses. By reducing bus length, the resistance and capacitance seen on these buses is reduced, consequently reducing power consumption and improving performance. Architectural partitioning makes much better use of the large number of D2D vias available than technology partitioning.

The finest partitioning granularity is the circuit level. Here, the transistors that make up a functional block may exist on different layers. Circuit partitioning has its own levels of granularity. At one extreme, blocks are simply split along logical boundaries into sub-blocks (e.g. a design could place half the banks of a cache on one layer and the other half on a different layer — so called bank-stacking [4, 5]). At the other extreme, individual circuits are split across the layers (e.g. in a register file, read and write ports may be spread across different layers, connected to the actual memory inverter pair through D2D vias; this is known as port-splitting [11]). This granularity best utilizes the available D2D vias and thus shows the best power and performance improvements.

3. MOTIVATION

From a quality perspective, 3D designs face the same problem plaguing IC boards and multi-chip modules (MCM): exponentially decreasing yield resulting from the integration of many distinct components. However, unlike these technologies, the current test strategy employed in industry is bond-and-pray; that is, no testing is performed on the layers before they are bonded together. This works well enough at low layer counts. But to achieve very high integration in the third dimension — tens or hundreds of layers — bond-and-pray will not suffice. Worse yet, the fine-grained partitioning that achieves the highest performance and lowest power and area consumption described in Section 2.2 only makes testing more difficult, if not impossible. Work in 3D-IC continues in the area of design complexity, both in the area of fundamental design (determining which designs best exploit the potential of a die stack) and in the area of CAD tools to assist designers in 3D floorplanning, routing, etc. The loss of testability, however, has gone thus far unaddressed. It is possible for 3D designers to rely on current planar techniques for test once the individual die layers are bonded together. Unfortunately, this imposes a practical limit on the number of die that can be stacked due to the exponential decay in chip yield [16]. To overcome this limit, pre-bond testability becomes a neces-

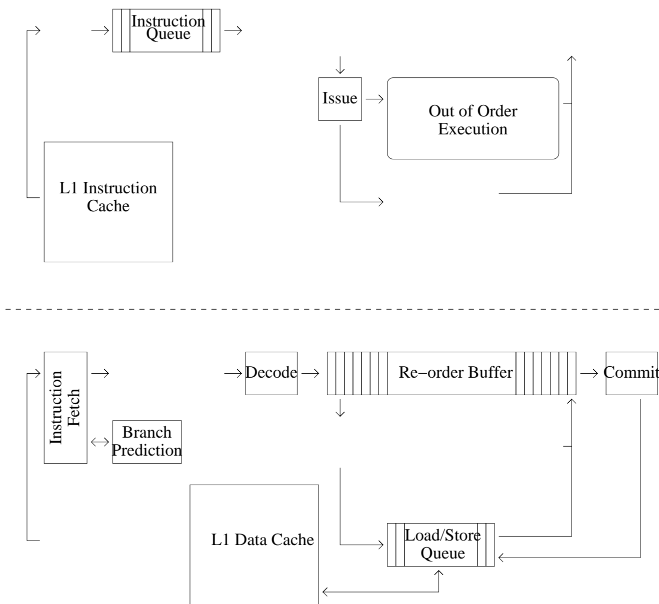

Figure 3: Example partition a generic out-of-order processor across two layers.

sity for achieving high enough yield to make such levels of 3D integration commercially viable.

The primary testability challenge posed by 3D integration is that each layer, before bonding has occurred, exists in an incomplete state. The severity of this incompleteness depends on the partitioning granularity. At the technology level, there is no problem, as each layer can be independently tested using the test methods developed for board and MCM integration.

Starting with the architectural level, however, trouble arises. Figure 2 shows a block-level model of a generic out-of-order processor. In a traditional planar design, all these microarchitectural blocks are placed and manufactured on a single die, so the traditional test methodologies [17, 18, 19, 20, 21] were developed for this case. In a 3D die stacked processor design, these blocks can be manufactured separately on different die layers. Before these layers are bonded together, each layer contains only half of the functional blocks in the case of a two-layer stacking. Thus, as seen in Figure 3, the complete microprocessor we had before is now incomplete[1] and thus cannot be tested by traditional methods.

At the circuit level, testing becomes even more challenging. Now even the functional blocks are incomplete, and, worse, the circuits themselves may be incomplete and functionally broken. This leads to a paradox of sorts in that we want to test broken circuits to see if they function correctly. Enabling test in such circumstances may be as simple as duplicating missing hardware, or it may require completely new DFT hardware. In-depth exploration of these challenges and their solutions is left to future work of this finest grained partitioning. In this paper, we focus on the architectural level partitioning.

The simple brute-force solution would be to probe each D2D via individually, providing or observing test values as necessary. Unfortunately, this is not a viable solution. The number of vias on a given layer will vary from hundreds for technology partitioning to hundreds of thousands for circuit partitioning. Such a massive number of test connections is well beyond the capabilities of modern testers [22]. Additionally, the process of actually making a connection with a test probe is a very stressful and damaging procedure [22]; the structure of the D2D vias [1] would be damaged to the point where the two die could not be successfully bonded post-layer-test. In order to test die layers pre-bond, a more practical solution is required.

Beyond this challenge of design incompleteness, several secondary concerns have been identified. First is the question of how pre-bond test fits into the larger testability picture. The hardware required for pre-bond test could potentially just be left sitting there after bonding has occurred, but this would be very wasteful. The pre-bond test hardware should be integrated into the post-bond test strategy.

The second challenge is the state of the fundamental support nets. Such nets include power, ground, and clocks. It could be the case that these nets are complete within each layer. It could also be that they exist in an island pattern; each net could be locally completely connected, but the wires connecting these local nets may exist on a separate layer. If these nets are non-functional, test of the logic they support is impossible.

The final challenge is that of test pads. In a traditional planar design, bond pads serve double-duty as test probe touchdown points and wire bond contacts for interfacing the chip to the outside world. In a 3D design, only bond pads on the top layer can play both roles. Any pads placed on lower layers only provide a test interface and are left hanging afterwards. Thus, test pads must be used very judiciously to control the area cost.

4. HARDWARE

There are several key hardware components necessary to realize structural test of individual die layers pre-bond. These include a general test architecture, specialized scan registers, the necessary support nets (power, ground, and clocks), and an interface to the outside world. In the following sections, we will discuss each component and the associated design tradeoffs, respectively.

It is important to note that a die will only be tested *once* before being bonded. Thus, any additional pre-bond test hardware must focus on post-bond reusability. Hardware that is not reused will be completely wasted post-bond and thus must be avoided as much as possible. Thus, the proposed solutions will emphasize reusability heavily to mitigate the cost of enabling pre-bond test.

4.1 Layers as Scan Islands

The Alpha 21364 utilized a test strategy of design segmentation [17]. Each segment was coined a *test island*, and they were isolated from their neighbors with specially designed border registers. During normal operation, these registers allowed data to flow freely between islands. In test mode, these registers closed the borders of the islands, replacing incoming values with test values from the scan chain. By segmenting the design into these islands, the complexity of the design was greatly reduced, making testing a faster and easier task.

Comparing this approach to 3D designs, it is clear that each layer, before bonding, exists as a perfectly isolated test island — a condition the Alpha designers were not able to achieve [17]. Thus we adopt this general test strategy for developing our pre-bond test methodology. The scan chains on each layer

[1]For example, instructions decoded in the bottom layer are dispatched from the instruction queue located in the top layer.

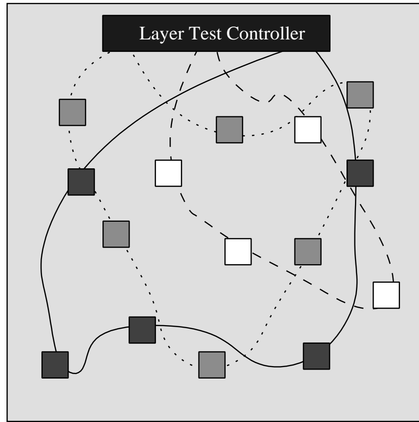

Figure 4: Implementation of scan chains on a single layer. Shown are generic scan registers, three chains connecting these registers, and one LTC controlling the chains.

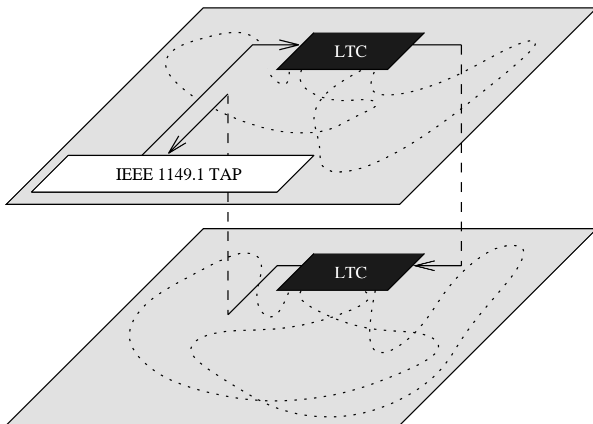

Figure 5: Integration of layer-level scan into chip-level scan with IEEE 1149.1 TAP. Shown are two layers (each with their associated chains), the IEEE 1149.1 TAP, and the routing that forms a serial test loop.

are managed by test logic termed the *Layer Test Controller (LTC)*. Figure 4 shows a generic chip layer with scannable registers hooked up into three scan chains controlled by an LTC. The choice of which registers are scannable and how these registers are wired into chains is design-dependent, involving a trade-off between functionality (speed, power, and area), test cost (time and power), and test coverage [23, 24, 25, 26].

The LTC provides scan chain access to the next step up in the test hierarchy. This is one of two different test mechanisms, depending on whether the layers have been bonded together. For pre-bond die, the LTC interfaces directly to the external Automatic Test Equipment (ATE) via probing. After bond, the LTC on each layer becomes part of a chip-level test chain, connecting all the test structures on the chip to a standard IEEE 1149.1 test controller as shown in Figure 5. This test controller is then accessed as normal via probing or dedicated package pins.

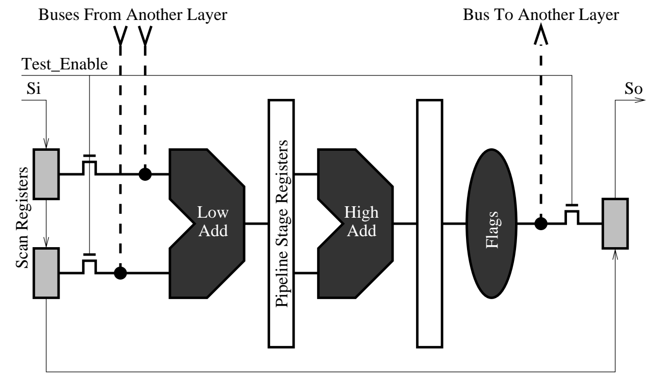

Figure 6: Shown is a three-stage pipelined adder which first adds the low-order bits, then adds the high-order bits, and finally computes the associated flags. Attached are injection and observation scan-flops which are integrated into one of the layer's scan chains. Thick lines indicate multi-bit structures (e.g. thick lines represent buses and thick nFETs represent one nFET per bit in the associated buses).

The actual design of the LTC is dependent on the application and the goal of the designers. At one extreme, the LTC could be as simple as a few wires that connects the various chains into a single, very long chain and provides one scan-in connection and one scan-out connection. At the other extreme, the LTC could be a full IEEE 1149.1 TAP. The most likely design would be somewhere in between including multiplexors, demultiplexors, and a bypass register. See Section 5 for the particular LTC design we used in our experiments.

4.2 Layer Border Scan Flops

To complete the scan island architecture, test values must be provided on the layer inputs, and test values must be observed on the layer outputs. There are two cases that must be considered. In the first case, a layer input directly drives a register or a register directly drives a layer output. To provide test value injection and observation, these registers need only be made part of a scan chain.

In the second case, D2D vias connect directly to logic, either as source or sink. Additional scan registers are required to inject and observe values on these lines. Figure 6 shows a three-stage staggered adder similar to the one implemented in the Pentium 4 processor [27]. In this adder, we assume the block providing the input and the block processing the output are placed on different layers. In order to inject and observe test values pre-bond, scan registers have been added to provide this functionality. The injection and observation scan registers shown are intentionally designed to be very light-weight in terms of area. The cost of this design choice is functionality; in particular, the value injection registers function properly only before bonding has occurred. Post-bond, turning on the pass FETs with the Test_Enable signal would cause contention between the injection registers and whatever entity is sourcing values on the neighboring layer. This limitation could be overcome by simply converting the pass FETs into 2-to-1 multiplexors that select between the scan values and normal operation values. Unfortunately, this functionality would come at an area and performance cost. Alternatively, these injec-

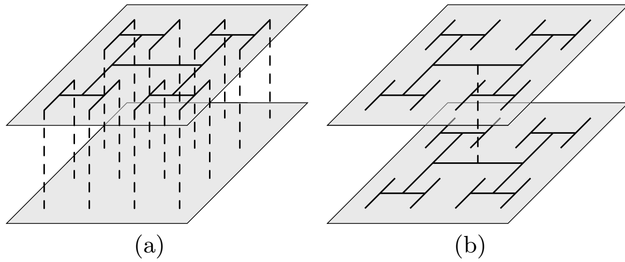

Figure 7: Two optimized clock trees. (a) is designed for shortest wire length and thus least power consumption. (b) is designed for maximum pre-bond testability.

tion registers could be reused, potentially as PRPGs, MISRs, BILBO registers, etc [28][29]. Such a reuse would be application specific and is not considered in this work.

4.3 Supporting Nets

Before the layers can be tested, they must be activated. This requires complete, functioning support nets: power, ground, and clock. This is given in planar designs, but this is not necessarily the case in 3D designs. There is the potential for these nets to be designed in such a way that they are connected locally but not globally; they could exist as local domains. As an example, Figure 7(a) shows a simple H-tree design for clock distribution in an ideal 3D die stacked processor. The clock tree exists almost entirely in the upper layer while vias provide local clock connectivity for the bottom layer. This is very efficient in terms of wire length and power — if these layers were placed side-by-side on a planar die, the clock tree would be twice as large and consume much more power. Allowed unlimited vias for clock routing, this is precisely the style of tree produced by an automated, balanced-skew 3D clock router [30]. Unfortunately, this design leaves the bottom layer completely untestable pre-bond.

To enable simple pre-bond test, the clock tree shown in Figure 7(b) would be the best. The H-trees on each layer are connected by a single via, requiring only a single ATE probe to provide the clock during pre-bond test. Overall, however, this design is just as poor as the power-optimal design. The bottom tree replicates the top tree, wasting significant amounts of power and routing area while providing no benefit post-bonding. A middle ground must be found between these two approaches.

One potential solution is a redundant design. In such a design, a fully-optimized 3D clock tree is designed that results in a tree similar to Figure 7(a). Next, a redundant clock tree is designed only for the second layer that will distribute the pre-bond test clock to that layer. Finally, an enable line is added to every buffer in the redundant tree. Asserting this signal will allow the ATE to activate this tree for testing the second layer pre-bond. Once the two layers have been bonded together, this line will be deasserted, disabling the redundant tree. From post-bond on, the optimized top-layer tree will clock the chip.

Such a design enables pre-bond testability in all layers while maintaining the power reduction benefits of a 3D tree. Routing area on the bottom layer is sacrificed, but this area would have been consumed by clock routing in the equivalent planar design anyways, so this design is no worse off than its planar equivalent in this regard. Thus, this 3D clock tree design is sufficient for the scope of this paper. Full exploration of clock tree design possibilities is left to future work.

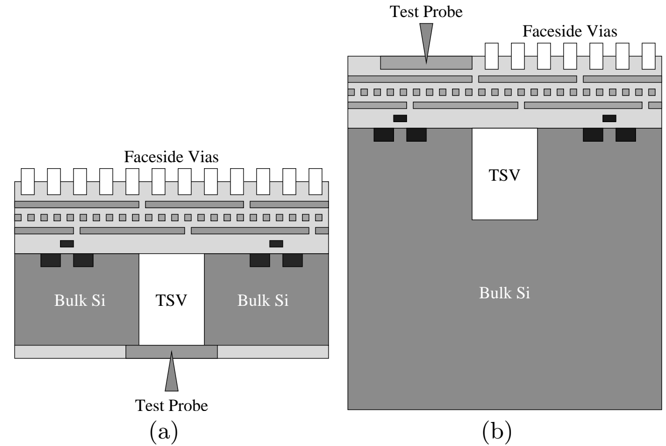

Figure 8: Two options for pre-bond test interfacing. (a) shows a thinned wafer in which backside vias are used to access the layer. (b) shows a pre-thinning wafer in which faceside test pads are used to access the layer.

Concerning power and ground nets, these will not suffer from the same problems. As a general design rule, power and ground are routed on the lowest metal layers (Metal 1 and 2 typically); this is due to their high utilization. Because of this, it is quite likely that the power and ground nets will be complete across the entirety of each and every layer. So unlike the clock network, the power and ground nets require no special treatment to enable layer test.

4.4 Test Pads

So now the hardware is in place to structurally test each layer prior to bonding, but a physical interface must be provided to access this hardware. There are two available options: backside probing and faceside probing, shown in Figure 8(a) and Figure 8(b), respectively. The backside probing method appears to be more attractive because backside vias are much less dense than faceside vias, so placing test pads on the backside of a layer creates less disruption of inter-die communication. There are two disadvantages though. First, the bottom die in a stack cannot be probed this way because it is never thinned. Thus a backside test interface cannot be standardized across the design. Second, the actual probing of a chip is a very stressful process involving a very large contact force between the probes and the silicon to ensure good electrical conduction. Since backside probing requires a *very* thin wafer, there is a strong possibility the die will not survive contact with a normal probe card. A new probe card could be developed with significantly tighter error margins, but the cost of such development would be prohibitive.

A more viable alternative is to use faceside probing. Faceside test can be performed without thinning the die, preventing destruction of the die during test. Also, faceside test can be applied universally to all die in a stack, so no die need receive special treatment. Unfortunately, a single test pad does consume area equivalent to that of a few hundred faceside vias. But with room for tens of millions of vias on the face of

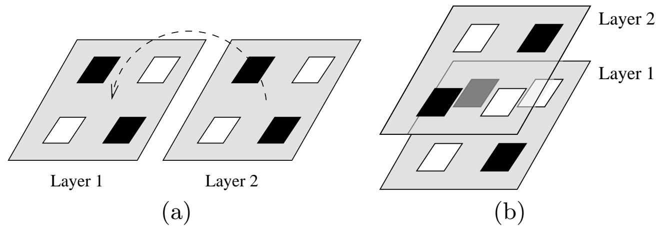

Figure 9: By arranging test pad connections to power (white) and ground (black) as shown in (a), flipping layer 2 onto layer 1 and bonding them in a face-to-face connection results in the formation of decoupling capacitors as shown in (b).

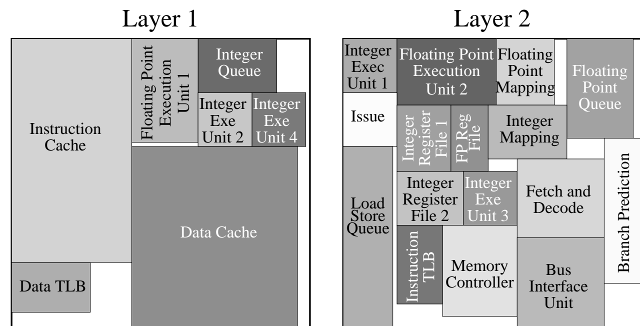

Figure 10: A floorplan for a two-layer die stack split by architectural block. The gray areas between and around blocks represents whitespace within the floorplan.

a single die, it should be possible to place these test pads in areas of low inter-die communication density, minimizing the impact.

Now something needs to be done with these test pads after bonding. One naive solution is to simply leave them hanging, but there are more elegant solutions. In the case of test pads that connect to the LTCs, these pads can be disconnected from the LTC via a transmission gate or a fuse. Transmission gates can also disconnect the clock trees from their associated test pads. In this case, disconnection is critical because leaving the test pads hanging introduces extra capacitance on the trees that (a) must be accounted for in the design of the tree (in order to produce a balanced, skew-free tree) and (b) wastes extra power, a problem aggravated by the fact that the clock is always switching. However, disconnecting the test pads is not always the best that can be done.

For the case of the power and ground nets, these will actually benefit from connected test pads. The capacitance introduced by the pads will naturally act as decoupling capacitance (decap), smoothing out current fluctuations (known as the $\frac{di}{dt}$ problem) in power supply network that represent a major concern in modern processor design [31]. But more can be done. Figure 9 shows two die with power and ground test pads. By arranging the test pads as shown in (a), subsequent bonding will form standard parallel-plate capacitors between the test pads, as shown in (b). Thus the test pads can provide even more decap, which helps justify the cost of including them for test.

5. EXPERIMENTS

Our experiments are based on the architecture and technology of the Alpha 21264. In order to evaluate the cost of implementing our pre-bond test strategy, we need to know the area consumed by a scan cell and the number of scan cells required in a 3D-integrated design.

To determine a realistic size for the scan cell, the scan cell was laid out using 0.25 micron TSMC design rules. This technology generation was selected to match, as closely as possible, that used to manufacture the 21264A. The actual design of the scan cells is based on the 8T latch. Each cell requires $75.8\mu m^2$ of silicon.

To determine the number of cells required by our technique, a sample 3D floorplan (Figure 10) for a 21264 was designed by a published 3D floorplanner [32]. From this floorplan we extracted the number of signals crossing between the die. Table 1 lists all of the inter-die buses, the number of signals comprising that bus, and the cost of adding the necessary scan cells. Note that each signal requires two scan cells: one on the source side to observe the test output and another on the sink side to provide a test input.

The bottom row in Table 1 gives the final area cost of injecting and observing test values on D2D via signals. This cost is 0.165% of the area of the sample floorplan in Figure 10. However, the floorplan contains 8.56% whitespace, so the scan flops do not require an expansion of the chip footprint. Additionally, the area consumed by the scan flops is only 0.173% of the die size of the original Alpha 21264A, which results in a negligible expansion of the die footprint.

Our experiments assume a simple LTC design. The LTC provides parallel access to sixteen scan chains per layer. Additionally, the LTC contains sixteen one-bit bypass registers. Finally, sixteen multiplexors and demultiplexors are included to allow selection between the scan chains and the bypass registers. Together, this allows for sixteen scan chains per layer — thirty two chains in the chip — which is comparable to modern designs [18]. This design requires thirty three test pads per layer: $S_i[15, 0]$, $S_o[15, 0]$, and a select signal. The area cost of such an LTC is insignificant compared to the cost of the injection and observation scan cells.

This area cost represents the worst-case cost we should expect for implementing this test technique for two reasons. First, academic layouts produced under publicly available DRC rules are much larger than functionally-equivalent industrial designs produced under highly-optimized and proprietary DRC rules [33]. Second, we assume a worst-case scan cell scenario in which *every* D2D via requires the addition of two scan cells that serve no purpose beyond pre-bond test value injection or observation. In a real design, many of these cells could be unnecessary — if the D2D via directly sources and/or sinks a scannable flip-flop — or could be reused as part of the post-bond test strategy, as discussed in Section 4.2. For these reasons, we expect an actual application of our technique in an industrial design to cost even less area than the results reported here.

6. CONCLUSION

In this paper, we presented a general DFT technique for enabling pre-bond testability for 3D die stacked microprocessors. 3D processor designers place functional blocks across

SOURCE	Die Layer	SINK	Die Layer	SIGNAL COUNT	AREA (μm^2)
Instruction Cache	1	Instruction TLB	2	40	6065
Instruction TLB	2	Instruction Cache	1	174	26384
Instruction Cache	1	Fetch and Decode	2	128	19409
Fetch and Decode	2	Instruction Cache	1	42	6369
Integer Mapping	2	Integer Queue	1	200	30326
Integer Queue	1	Issue	2	196	29720
Integer Register File 1	2	Integer Execution Unit 2	1	150	22745
Integer Execution Unit 2	1	Integer Register File 1	2	71	10766
Integer Execution Unit 2	1	Integer Mapping	2	14	2123
Integer Execution Unit 2	1	Branch Predictor	2	93	14102
Integer Register File 2	2	Integer Execution Unit 4	1	150	22745
Integer Execution Unit 4	1	Integer Register File 2	2	71	10766
Integer Execution Unit 4	1	Integer Mapping	2	14	2123
Integer Execution Unit 4	1	Branch Predictor	2	93	14102
Floating Point Register File	2	Floating Point Execution Unit 1	1	154	23351
Floating Point Execution Unit 1	1	Floating Point Register File	2	71	10766
Floating Point Execution Unit 1	1	Floating Point Mapping	2	14	2123
Load/Store Queue	2	Data TLB	1	66	10008
Load/Store Queue	2	Data Cache	1	180	27294
Data Cache	1	Load/Store Queue	2	144	21835
Data Cache	1	Memory Controller	2	166	25171
Memory Controller	2	Data Cache	1	166	25171
TOTAL				2397	363,461

Table 1: This list consists of the buses that cross from one layer to another. Listed are the source block and layer, the sink block and layer, the number of signals, and the area penalty paid to include scan flops as in Figure 6.

different die layers in order to improve processor performance and power consumption. Unfortunately, the testing of such an incomplete processor is a major challenge. In this work, we leverage a test architecture similar to that previously employed in 2D planar designs [17]. However, this is, to the best of our knowledge, the first time such a DFT technique has been applied to and analyzed step-by-step for 3D-IC architectures. Our design focused on simple, straightforward solutions to the problems 3D test poses, and our experimental results show that basic, structural pre-bond test is not only possible but practical at a negligible cost. This is a testability technique that can be integrated into comprehensive test strategy with minimal extra effort.

7. ACKNOWLEDGMENT

This research is supported by the C2S2 and GSRC centers of the Focus Center Research Program. The authors also thank Michael Healy, Dae Hyun Kim, Sung Kyu Lim, Gabriel Loh, and Kiran Puttaswamy for their inspirational discussion.

8. REFERENCES

[1] S. Gupta, M. Hilbert, S. Hong, and R. Patti. Techniques for Producing 3D ICs with High-Density Interconnect. In *VMIC '04: Proceedings of the 21st International VLSI Multilevel Interconnection Conference*, Waikoloa Beach, HI, USA, 2004.

[2] B. Black, M. Annavaram, N. Brekelbaum, J. DeVale, L. Jiang, G. H. Loh, D. McCauley, P. Morrow, D. W. Nelson, D. Pantuso, P. Reed, J. Rupley, S. Shankar, J. Shen, and C. Webb. Die Stacking (3D) Microarchitecture. In *Proceedings of the 39th International Symposium on Microarchitecture*, 2006.

[3] Bryan Black, Donald Nelson, Clair Webb, and Nick Samra. 3D Processing Technology and Its Impact on iA32 Microprocessors. In *Proceedings of the 22nd International Conference on Computer Design*, pages 316–318, 2003.

[4] F. Li, C. Nicopoulos, T. Richardson, Y. Xie, N. Vijaykrishnan, and M. Kandemir. Design and Management of 3D Chip Multiprocessors using Network-in-Memory. In *Proceedings of the International Symposium on Computer Architecture*, 2006.

[5] Kiran Puttaswamy and Gabriel H. Loh. Implementing Caches in a 3D Technology for High Performance Processors. In *Proceedings of the International Conference on Computer Design*, 2005.

[6] Kiran Puttaswamy and Gabriel H. Loh. Dynamic Instruction Schedulers in a 3-Dimensional Integration Technology. In *Proceedings of the ACM/IEEE Great Lakes Symposium on VLSI*, 2006.

[7] Kiran Puttaswamy and Gabriel H. Loh. The Impact of 3-Dimensional Integration on the Design of Arithmetic Units. In *Proceedings of the International Symposium on Circuits and Systems*, 2006.

[8] Kiran Puttaswamy and Gabriel H. Loh. Thermal Herding: Microarchitecture Techniques for Controlling Hotspots in High-Performance 3D-Integrated Processors. In *Proceedings of the 13th International Symposium on High-Performance Computer Architecture*, pages 193–204, 2007.

[9] S. Bhansali, G. Chapmann, E. Friedman, Y. Ismail, P. Mukund, D. Tebbe, and V. Jain. 3-D Heterogeneous Sensor System on a Chip for Defense and Security Applications. In *Proceedings of SPIE Volume 5417*, pages 413–424, 2004.

[10] V. Pavlidis and E. Friedman. 3-D Topologies for Networks-on-Chip. In *International SOC Conference*, pages 285–288, 2006.

[11] Kiran Puttaswamy and Gabriel H. Loh. Implementing

Register Files for High-Performance Microprocessors in a Die-Stacked (3D) Technology. In *ISVLSI '06: Proceedings of the IEEE Computer Society Annual Symposium on Emerging VLSI Technologies and Architectures*, pages 384–389, 2006.

[12] Shashidhar Mysore, Banit Agrawal, Navin Srivastava, Sheng-Chih Lin, Kaustav Banerjee, and Timothy Sherwood. Introspective 3D Chips. In *Proceedings of the International Conference on Architectural Support for Programming Languages and Operating Systems*, 2006.

[13] Tezzaron. http://www.tezzaron.com/technology/fastack.htm. 2006.

[14] Samsung. http://www.samsung.com/presscenter/pressrelease/pressrelease.asp?seq=20060413_0000246668. 2006.

[15] Y. Nunomura and N. Manijikian. M32R/D-Integrating DRAM and Microprocessor. *IEEE MICRO*, 17(6):40–48, 1997.

[16] Yangdong Deng and W. P. Maly. 2.5-Dimensional VLSI System Integration. *IEEE Trans. VLSI Syst.*, 13(6):668–677, 2005.

[17] Dilip K. Bhavsar and Richard A. Davies. Scan islands - a scan partitioning architecture and its implementation on the alpha 21364 processor. In *VTS '02: Proceedings of the 20th IEEE VLSI Test Symposium*, page 16, Washington, DC, USA, 2002. IEEE Computer Society.

[18] M. Riley, L. Bushard, N. Chelstrom, N. Kiryu, and S. Ferguson. Testability Features of the First-Generation Cell Processor. In *Proceedings of the International Test Conference*, 2005.

[19] P. Tan, T. Le, K.-H. Ng, P. Mantri, and J. Westfall. Testing of UltraSPARC T1 Microprocessor and Its Challenges. In *Proceedings of the International Test Conference*, 2006.

[20] T. McLaurin. The Challenge of Testing the ARM Cortex-A8^{TM} Microprocessor Core. In *Proceedings of the International Test Conference*, 2006.

[21] P. Parvathala, K. Maneparambil, and W. Lindsay. FRITS - A Microprocessor Functional BIST Method. In *Proceedings of the International Test Conference*, 2002.

[22] O. Weeden. Probe Card Tutorial, http://www.keithley.com/data?asset=13263, 2003.

[23] S. Makar. A Layout-Based Approach for Ordering Scan Chain Flip-Flops. In *Proceedings of the International Test Conference*, 1998.

[24] D. Berthelot, S. Chaudhuri, and H. Savoj. An efficient linear time algorithm for scan chain optimization and repartitioning. In *Proceedings of the International Test Conference*, 2002.

[25] M. Hirech, J. Beausang, and X. Gu. A New Approach to Scan Chain Reordering Using Physical Design Information. In *Proceedings of the International Test Conference*, 1998.

[26] Y. Bonhomme, P. Girard, L. Guiller, C. Landrault, and S. Pravossoudovitch. Efficient Scan Chain Design for Power Minimization During Scan Testing Under Routing Constraint. In *Proceedings of the International Test Conference*, 2003.

[27] Glenn Hinton, Dave Sager, Michael Upton, Darrell Boggs, Doug Carmean, Alan Kyker, and Patrice Roussel. The Microarchitecture of the Pentium 4 Processor. *Intel Technology Journal*, 5(1), 2001.

[28] Miron Abramovici, Melvin A. Breuer, and Arthur D. Friedman. *Digital Systems Testing and Testable Design*. IEEE Press, New York, 1990.

[29] B. Konemann, J. Mucha, and G. Zwiehoff. Built-In Logic Block Observation Techniques. In *Proceedings of the International Test Conference*, 1979.

[30] Mohit Pathak and Sung-Kyu Lim. Thermal-aware Steiner Routing for 3D Stacked ICs. In *To appear in the IEEE International Conference on Computer-Aided Design*, 2007.

[31] Y.-S. Chang, S. Gupta, and M. Breuer. Analysis of Ground Bounce in Deep Sub-Micron Circuits. In *VLSI Test Symposium*, pages 110–116, 1997.

[32] E. Wong and S.-K. Lim. 3D Floorplanning with Thermal Vias. In *Design, Automation, and Test in Europe Proceedings*, pages 878–883, 2006.

[33] Personal Communication with Josh Fryman. Intel corporation, 2006.