

# Test Strategies for 3D Die-Stacked Integrated Circuits

Dean L. Lewis  
dean@gatech.edu

Hsien-Hsin S. Lee  
leehs@gatech.edu

School of Electrical and Computer Engineering  
Georgia Institute of Technology  
Atlanta, GA 30332  
<http://arch.ece.gatech.edu>

## ABSTRACT

*3D integration technology is a radical new chip assembly technology that promises greater numbers of devices on chip, increased performance, and reduced power consumption. However, in order for this technology to be economically viable, we must be able to test each die before it is bonded to the rest of the die to form a stacked chip. Pre-bond test presents interesting new challenges because chip functionality and connectivity is partitioned across different pieces of silicon. We will explore these challenges and present test strategies to address them. Our solutions are simple extensions of current scan-based test technology, enabling simple integration of 3D into current test systems. Our results show that full pre-bond test can be achieved at equal or lower cost than testing an equivalent planar design.*

## 1. INTRODUCTION

3D die stacking is a promising new technology that enables the tight integration of multiple silicon die in a vertical stack [3][6]. In its simplest form, multiple planar designs are stacked vertically, significantly increasing on-chip device count. In addition, these planar layers do not need to be homogeneous as they could be designed separately using different design flows. As such, 3D stacking makes the integration of heterogeneous devices onto a single package possible. At the opposite end of the spectrum, individual circuits may be split across multiple die in the stack. Such designs promise simultaneous reductions in delay, power, and area. Unfortunately, 3D die stacking also presents tough new challenges to industry including power delivery, routing, 3D aware design flows, and thermal dissipation, along with many others. One chief challenge is testability, specifically testing each individual die pre-bond (i.e. before the die are bonded together to form a complete chip stack). Before bonding, design functionality is partitioned across multiple die. Depending on the partitioning styles, only partial circuits may exist on any single layer. A simple solution is “bond and pray,” where the stack is fully assembled and then tested as a complete design. This is not a practical solution because the manufacturing yield will fall off exponentially as the number of stacked die increases. This in turn imposes an economic limit on the number of die that can be stacked.

In this paper, we will explore the range of challenges presented by pre-bond test and discuss several techniques that can enable this crucial technology. We will consider the different granularities of 3D partitioning as well as the special needs of the so-called “hardcore” hardware. We will also

present case studies representative of these varying challenges and demonstrate the application of such test strategies.

## 2. TECHNOLOGY PARTITIONING

We begin at the granularity of technology partitioning. At this granularity, die manufactured in different technologies, for example high-speed CMOS and high-density DRAM, are bonded together in a 3D stack. Each individual die is effectively a planar design where TSVs replace what would normally be off-chip connections. Of course, TSVs are orders of magnitude smaller than board or package wiring, so 3D significantly reduces latency and increases bandwidth between the components.

The test challenges here are minimal. Each design is completely functional, and each die can be individually tested just as they would be in standard system with off-chip DRAM. Traditional boundary scan can be employed to test the quality of the actual TSV bonds. One small concern: once the die are bonded into a stack, all but the topmost die are physically inaccessible to probing. Thus, the test hardware must be designed in such a way as to allow logical access to these lower layers. Fortunately, current test hardware designs are hierarchical in nature [2], so 3D test can be seen as simply another level in the hierarchy.

## 3. ARCHITECTURAL PARTITIONING

The next smaller granularity is the block level. At this granularity, the die are manufactured in the same technology, but different functional units are partitioned across the stack. For example, we might stack the arithmetic units on top of the register file to reduce the length of the operand and result buses. Such a partitioning scheme increases the test difficulty; a planar test strategy that calls for one functional unit to supply test data to a neighboring functional unit is no longer a viable strategy if these units exist on different die. Once again, extensions to current scan-based test techniques can be applied to achieve the required test coverage. In this case, it comes down to establishing controllability of input signals from TSVs and observability of output signals to TSVs. Such functionality can be provided with well-studied test techniques like PRPG, MISR, and BILBO hardware [8].

### 3.1 Architecture Test Strategy

To enable test for this granularity, we adopt the *scan island* test architecture as implemented in the Alpha 21364

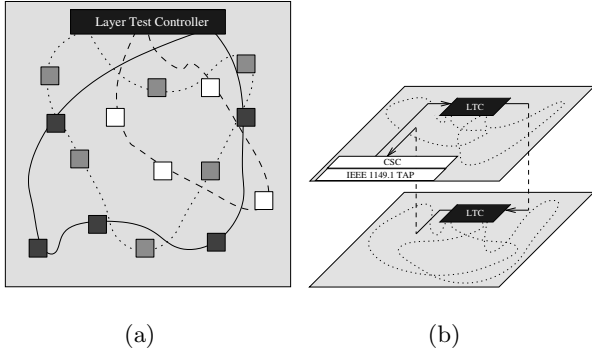


Figure 1: A generic 3D implementation of the scan island architecture. (a) shows individual scan registers connected in series to form scan chains and these scan chains connected to the LTC. (b) shows the LTCs connected in series to the CSC for post-bond test.

processor [2]. In this design, the processor is subdivided into logically-independent islands with the use of scan chains. The scan chains all connect to an *island scan port* (ISP). The ISPs are connected in series to the *central scan controller* (CSC), which then interfaces to the ATE for in-house test or the IEEE 1149.1 TAP for board-level test.

Applying the concept of scan islands to 3D, we find that each individual silicon layer is already a perfectly isolated island pre-bond. Thus we adopt the scan island test architecture for 3D test (Figure 1). ISPs are replaced with *layer test controllers* (LTC). Pre-bond each LTC interfaces directly with the ATE to run the layer test. Post-bond the LTCs are connected serially to the CSC, just as in the 21364. This hierarchical approach gives us full test access at each stage of manufacturing.

Of course, there’s no reason for each layer to be limited to a single island. As appropriate for the test requirements of a specific design, a layer may be subdivided into multiple scan islands. Each ISP would then connect in series to the LTC.

With this general test architecture in place, we must consider the individual cross-layer signals. Just as the Alpha test team inserted scan registers to isolate islands, we insert them to establish controllability and observability of TSV signals. In the worst case, two registers are required per TSV, one on the sending layer to observe the test output and one on the receiving layer to provide the test input. Figure 2 shows an example of this design applied to a high-speed pipelined adder. Note the insertion of pass FETs to disable the test signals post-bond. One distinct advantage of 3D is that it is possible to avoid adding any extraneous gates to the operational path. The cost of this design is that the test registers become useless post-bond. Traditional mux insertion is another option, of course. If we choose this route, the pair of control and observation latches on each TSV can be used to enable boundary-scan-like test of the TSVs themselves post-bond. This may be desirable as it enables very quick verification of the TSV connections.

These varying test strategies highlight the flexibility of our test architecture. It is up to each 3D test team to decide which option is best for a particular product.

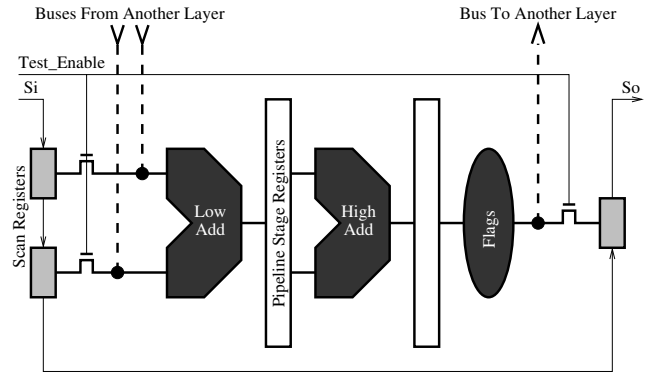


Figure 2: Shown is a three-stage pipelined adder which first adds the low-order bits, then adds the high-order bits, and finally computes the associated flags. Attached are injection and observation scan-flops which are integrated into one of the layer’s scan chains. Thick lines indicate multi-bit structures (e.g. thick lines represent buses and thick nFETs represent one nFET per bit in the associated buses).

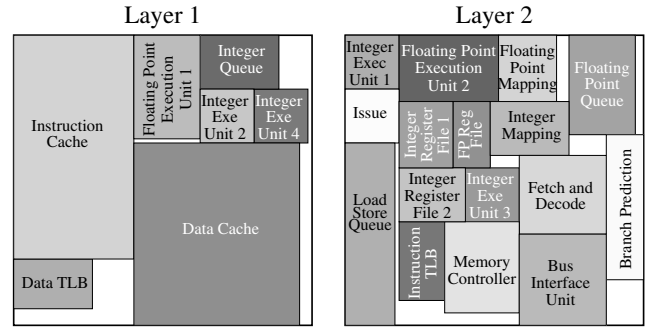


Figure 3: A floorplan for a two-layer die stack split by architectural block. The gray areas between and around blocks represents whitespace within the floorplan.

### 3.2 Architecture Experiments

To evaluate our test architecture, we took the widely-studied Alpha 21264 as a case study. To evaluate the overhead of each scan cell, we produced a layout for one in  $0.25\mu\text{m}$  TSMC technology, which closely matched the technology used in Alpha’s 21264A product. The design is a pair of 8T latches, consuming a total silicon area of  $75.8\mu\text{m}^2$ . To determine the total number of scan cells required, we employed a published 3D floorplanner[15] (Figure 3). From this floorplan, we counted the number of cross-layer signals and calculated the total overhead by assuming a worst-case two scan cells per TSV—this is a worst-case scenario because in a real 3D design many functional latches could also serve in the pre-bond test capacity. In total, 4794 cells were required, for a total silicon overhead of 0.165%, a negligible cost. Detailed results of this case study can be found in [10].

### 4. CIRCUIT PARTITIONING

The finest partitioning granularity is the circuit level. At this level we see a variety of partitioning schemes from simple sub-block partitions to very ambitious transistor-level partitions, where even individual circuits are split across

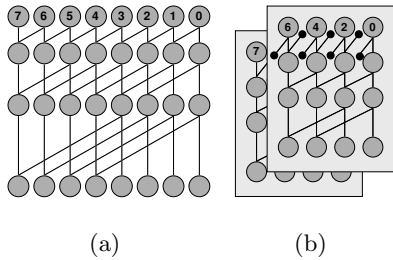


Figure 4: An eight-bit Kogge-Stone adder. (a) shows the planar implementation with its massive wiring area. (b) shows the 3D design with the significant wiring reduction; the black dots represent TSVs.

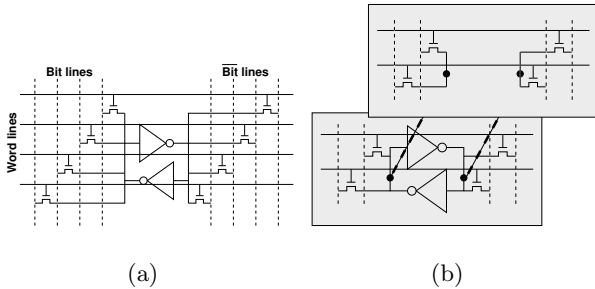


Figure 5: A four-port SRAM cell. This cell is laid out in an array to form a four-port register file. (a) shows the planar implementation with its massive wiring area. (b) shows the equivalent 3D design. Note that the lengths of the bitlines, wordlines, and internal nets have all been significantly reduced.

several layers. Here pre-bond test becomes trickiest, if not completely impossible. Unlike technology- and block-level partitioning, the sheer number of TSV connections can overwhelm the latch-insertion techniques in the previous section; the number of latches required is simply too great. Fortunately, we see an interesting trend in testing these circuit-partitioned designs. In general, test cost increases super-linearly with circuit complexity. As a result, the cost of separately testing each sub-block or sub-circuit individually plus the cost of testing the TSV connections post bond is usually similar to, and can be significantly less than, the cost of testing the equivalent planar design.

## 4.1 Circuit Design

For circuit partitioning, we examine a bit-sliced Kogge-Stone adder and a port-split register file [14]. In the bit-sliced adder design, the odd bits are computed on one layer and the even bits in the other (Figure 4). Data is shared between the layers only in the first stage of the computation when the neighboring bits are summed together. Such a design is fairly described as sub-block partitioning, a coarse-grained example of circuit partitioning.

On the opposite end of the circuit-partitioning spectrum is the port-split register file (Figure 5). This 3D design targets the many-ported register files required for today’s very wide out-of-order microprocessors, some of which require as many as twenty access ports. Port-splitting effectively targets the quadratic growth in SRAM cell size, significantly

reducing the size compared to a planar design. A smaller cell size means shorter word- and bitlines, which means smaller drivers, so port-splitting is a win in every part of the register file design. Port-splitting is a wonderful showcase of the power of 3D design.

## 4.2 Circuit Test

With two very different 3D designs, we require two very different test strategies. Testing the Kogge-Stone adder is straight-forward. With only two TSVs per adder bit, we can effectively apply scan test to this partitioning. Conveniently, the two signals that cross the layer boundary are already observable on the source layer. Thus, unlike above, we only need one scan register per signal to act as a test control on the receiving layer, reducing the scan test overhead by half.

Testing the port-split register file, on the other hand, requires a completely new approach. The bottom layer, which contains the actual storage cells, can be tested with any standard RAM test, e.g. walking ones. The other layers, obviously, cannot. To test these layers, we propose *transmit test*. The idea is simple; we place a test vector on one write port, pass it through the pass FETs, and read it from the read ports, all in a single test cycle. This fully exercises the address decoders, write drivers, pass FETs, and sense amplifiers, enabling full coverage of stuck-at faults. In order to run this test, at least one write port and one read port is required on each layer, which becomes a DFT requirement for the design team to meet. Fortunately, the number of read ports is usually well-balanced with the number of write ports (no worse than two-to-one), so meeting this constraint is not overly difficult. Of course, one port could always be made read/write just for the sake of test if required.

Running a transmit test is easy enough. We must be able to source test addresses to the decoders and test vectors to the write ports. We must also be able to read test vectors from the read ports. All of these requirements are already met by the test hardware that is in place for standard RAM test. The only change would be in the control logic, which would have to alter the timing of read and write enable signals to allow for test vector transmission.

## 4.3 Circuit Experiments

For our experiments, we produced two-layer implementations of the adder and register file. To evaluate our test strategies, we used two different tool flows. For delay and energy measurements, we used the 3DMagic [4] design tool, which can produce 3D VLSI layouts. These layouts were then extracted to HSPICE for simulation.

To evaluate fault coverage, we used the FlexTest tool from Mentor Graphics [5]. This tool takes VHDL models as input and both produces a set of test vectors and calculates the fault coverage. To determine the total cost of 3D test, we sum the cost of testing the bottom layer pre-bond, the top layer pre-bond, and the TSVs post-bond.

We will briefly summarize the results here; a full analysis is available in [11]. Confirming the prior work, both 3D designs out-performed their planar counterparts in area, delay, and energy simultaneously. Most interesting, however, is that both 3D designs were cheaper to test as well. The planar Kogge-Stone adder required 313 test patterns, while the 3D equivalent required slightly less, 301 patterns—146 for the top layer, 145 for the bottom, and ten for the TSVs.

The results are even better for the register file. Using Suk

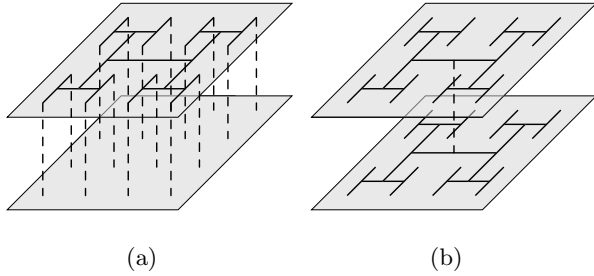


Figure 6: Design of a 3D clock tree. (a) shows a 3D clock tree optimized for performance; the wirelength and thus power consumption is minimized. (b) shows a 3D clock tree naively optimized for pre-bond test. Each layer has a complete tree, but post-bond a lot of unnecessary wire is driven, wasting significant power.

and Reddy’s Test B[1], the planar register file requires 8192 accesses. Our 3D register file, on the other hand, requires only 4864 test accesses—4096 to apply Test B to the bottom layer, 256 to execute a transmit test on the top layer, and 512 accesses to test the TSVs—a 40% reduction in test cost. This results showcases how 3D can not only significantly improve circuit performance but in some cases significantly reduce test cost as well.

Of course, a transmit test will not work if it takes too long or consumes too much power. And there is reason to worry because the transmit test requires a very long bit line to be charged through two small pass FETs in series. To test the practicality of transmit test, we simulated the test environment in HSPICE as well. Transmission takes approximately  $1.5ns$ , compared with  $1.4ns$  for a planar read and  $1.0ns$  for a 3D read. The energy results are similar;  $0.16pJ$  for transmission versus  $0.15pJ$  for a planar read and  $0.13pJ$  for a 3D read. These results demonstrate the feasibility of transmission test in a real test environment.

## 5. IMPLICATION TO CLOCK AND POWER DISTRIBUTION

Functional logic is not the only part of the design affected by 3D integration. We must also consider the chip’s hardware: its power and clock networks. Thankfully, to minimize IR drop, modern processors employ dense grids of power and ground wiring. Such grid designs ensure that these critical nets are fully connected within each and every die in the stack. Unfortunately, the clock distribution network is not so simple.

Previous design work has shown that the optimal 3D clock distribution is one in which the clock is distributed on a single layer and provided to other layers through TSVs as the leaves of the clock tree[13] (Figure 6(a)). Such a design minimizes the wirelength of the clock tree, thus also minimizing power consumption. However, it also creates thousands of disconnected clock domains on each layer (excepting the distribution layer) pre-bond. Without a functional clock, all of the test methods described above are completely useless. The simplest solution is to distribute the clock through a fully connected tree on each layer; these layers can then be fully connected with a single central TSV (Figure 6(b)). Unfortunately, this design is very wasteful due to the redundant

nature of the trees.

In a pre-bond testable 3D design, we have to have an operational clock pre-bond, but we don’t want to pay the massive power cost of the redundant trees. We propose two potential solutions, which are currently works-in-progress. The first and simpler design starts with an optimized test-unaware 3D clock tree as described above. Then, to enable pre-bond test, each layer—excepting the layer that is already fully connected—is augmented with a pre-bond test tree that connects all the leaf nodes together. This tree includes a gating signal; pre-bond the tree is on to enable test, post-bond it is off so power is not wasted switching the redundant distribution wiring.

The second design is more complicated but also more useful. Basically, in this design, even the distribution portion of the original 3D tree can be gated. The effect of this slight alternation is that the distribution wiring on any layer can be used to clock *all* of the leaf nodes throughout the stack. Thus, if one distribution network fails, e.g. due to electromigration, another layer can be switched on to allow the stack to continue operating. Thus we use the redundancy in the previous design to increase product lifetime. The downside is the complexity of the design. It is of the utmost importance to minimize skew in the clock network. In the first design, this is fairly simple because there is only one post-bond clock to design. In the second design, there are  $n$  trees to design, much more difficult. Additionally, each distribution net will intersect the leaves of the clock tree at slightly different points (plus or minus a couple TSV delays), exacerbating the skew problem. Whether the benefits of the more complicated option justify the costs or if redundant low-skew trees are even possible remains to be seen. However, either design promises a combination of pre-bond testability and optimal post-bond operation.

## 6. RELATED WORK

3D test is still in the very early stages. Below are contributions others have made to this emerging field. Mak identified a list of challenges facing 3D test going forward[12]. Wu *et al.* studied scan chain ordering in a 3D stack to minimize wirelength[16]; this work does not consider pre-bond test. Jiang *et al.* studied the total test time of 3D systems, factoring in pre-bond test[7]. More recently, Lee and Chakrabarty identified the research challenges still yet to be addressed in 3D-ICs[9].

## 7. SUMMARY

In this paper, we have identified a variety of challenges facing 3D test engineers and presented a number of potential solutions. Taken together, these test strategies provide a framework for enabling pre-bond test in 3D integrated designs. More importantly, these test techniques can be easily integrated into existing test plans and executed on existing test systems. This greatly reduces the barrier to 3D adoption in industry. With these test solutions, we can finally realize the amazing potential 3D integration promises the semiconductor industry.

## 8. ACKNOWLEDGMENT

This work was supported in part by C2S2, a center of the Focus Center Research Program, a Semiconductor Research Corporation Program and an NSF grant CCF-0811738.

## 9. REFERENCES

- [1] M. S. Abadir and H. K. Reghbati. Functional testing of semiconductor random access memories. *Computing Surveys*, 15(3), September 1983.
- [2] D. K. Bhavsar and R. 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.
- [3] 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.
- [4] S. Das, A. Chandrakasan, and R. Reif. Design Tools for 3-D Integrated Circuits. In *Asia South Pacific Design Automation Conference (ASP-DAC)*, pages 53–56, 2003.
- [5] M. Graphics. <http://www.mentor.com>, 2007.
- [6] 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.
- [7] L. Jiang, L. Huang, and Q. Xu. Test Architecture Design and Optimization for Three-Dimensional SoCs. In *Proceedings of the Design Automation and Test in Europe*, 2009.
- [8] B. Konemann, J. Mucha, and G. Zwiehoff. Built-in logic block observation techniques. In *Test Conference*, 1997.
- [9] H.-H. S. Lee and K. Chakrabarty. Test Challenges for 3D Integrated Circuits. *To appear in IEEE Design and Test of Computers, Special Issue on 3D IC Design and Test*, Sep/Oct 2009.
- [10] D. L. Lewis and H.-H. S. Lee. A Scan-Island Based Design Enabling Pre-bond Testability in Die-Stacked Microprocessors. In *IEEE International Test Conference (ITC)*, October 2007.
- [11] D. L. Lewis and H.-H. S. Lee. Testing Circuit-Partitioned 3D IC Designs. In *IEEE Computer Society Annual Symposium on VLSI (ISVLSI)*, May 2009.
- [12] T. M. Mak. Test challenges for 3D circuits. In *Proceedings of the IEEE On-Line Testing Symposium*, 2006.
- [13] J. Minz, X. Zhao, and S. K. Lim. Buffered clock tree synthesis for 3d ics under thermal variations. In *IEEE/ACM Asia South Pacific Design Automation Conference*, 2008. To appear.
- [14] K. Puttaswamy and G. 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.
- [15] E. Wong and S.-K. Lim. 3d floorplanning with thermal vias. In *Design, Automation, and Test in Europe Proceedings*, pages 878–883, 2006.
- [16] X. Wu, P. Falkenstern, and Y. Xie. Scan Chain Design for Three-dimensional Integrated Circuits (3D ICs). In *Proceedings of the International Conference on Computer Design*, 2007.