Embedded memory test and repair optimizes SoC yields

Sandeep Kaushik and Yervant Zorian, Synopsys - July 17, 2012

As process technologies continue to shrink and memory size and design complexity grow, it has become increasingly difficult to achieve high manufacturing yield. Embedded memories are the most dense components within a system-on-chip (SoC), accounting for more than 50 percent of the chip area. Implemented using aggressive design rules, embedded memories tend to be more prone to manufacturing defects and field reliability problems than any other core on the chip. Therefore, the overall yield of an SoC depends heavily on the memory yield, and securing high memory yield is critical to achieving lower silicon cost.

New systematic defects are often manifested as yield-limiting faults resulting from known factors such as reduced feature sizes. Additionally, the use of high-density integration and packaging technologies such as 3D-IC, which require complex manufacturing processes and have associated physical access limitations, further compounds the problem. These new and emerging challenges make it critical that embedded memory test and repair solutions keep up with technology advances in order to consistently provide superior test quality and yield optimization. This article will describe embedded memory test solutions, including fault detection in very deep submicron technologies, repair at the manufacturing level, and diagnosis for process improvement and field repair capabilities, that address today’s design yield and reliability needs.

Keeping up complexity
Today’s demanding applications require SoCs that are bigger and faster that are more area, timing, and power sensitive than ever before, resulting in a shift from the logic-dominant chips of the past to memory-dominant ones. Figure 1 shows embedded memory projections from Semico Research Corporation. In 2008, embedded memories accounted for more than half of the die area in a typical SoC. It’s predicted that the amount of space they occupy on the die will continue to increase, reaching up to 70 percent by 2017.
Applications that require lots of memory are served by designs that embed large numbers of memory bits per chip, creating more powerful SoCs, but this has the associated problems of increased die size and poor yield. As design applications require more memory, it is essential to implement a comprehensive embedded memory test, repair, and diagnostic solution to help achieve high yield.

In addition to needing the capability to deal with increasing numbers of on-chip memory instances, the embedded memory test solution also must be able to handle aggressive design specifications, such as hierarchies, size, performance, area, and power consumption. As embedded test solutions often have a negative impact on performance, it is important that high-performance cores be built with carefully planned memory BIST MUX logic and an integrated memory test bus to minimize these effects. The bus consists of pipelining, latency, and setup for the memory signals of all the memory instances in the core and eliminates the need for a traditional BIST wrapper for each memory.

An embedded memory test solution will interact with this test bus and add the required logic to use the test bus interface for embedded memory test. Additionally, an effective embedded memory test solution will understand the functional pipelining with predetermined controllability and observability logic on the timing-sensitive datapaths, and it will have the flexibility to add the required number of pipeline stages on the memory BIST paths to meet the performance targets of the design. To further address design complexity, a test and repair solution that is integrated with embedded memories, with most of the timing-critical BIST wrapper logic hardened in the memories, makes it possible to achieve faster design closure while meeting the performance, power and area characteristics of the design.

An embedded memory test solution with a comprehensive set of test algorithms that are optimized to provide out-of-box fault coverage for embedded memories in each advanced node helps improve yield ramp-up time while minimizing the required test time. The embedded test solutions developed for 90-nm technology nodes will not deliver the same level of test quality for 28-nm technology nodes, because memory defects and failure mechanisms change as process technologies shrink. The following three-step flow can be used to test algorithms covering a wide range of faults associated with advanced technology nodes:

1. Memory layout to electrical circuit extraction using memory scrambling information
2. Electrical circuit to fault modeling extraction using SPICE simulations. A comprehensive set of faults can be injected on electrical circuits including the memory array, address decoder, sense amplifier, and write driver to validate the coverage of test algorithms.

3. Fault modeling to test algorithm extraction using test algorithm generator tools that can generate minimal March test algorithms for detection of a given set of faults.

**Manufacturing repair**

With the overall yield of an SoC being largely dependent on memory yield, it is important to implement techniques to improve it. Although the yield of native memory may be inadequate, embedded memory yield can be improved through the use of redundancy or spare elements.

In Figure 2, the purple lines represent memory yield as a function of the aggregate memory bit-count in an SoC. In this example, the yield for 24 Mbit of embedded memory is close to 20 percent for new processes, represented by the longer purple line, assuming chip dimensions of 12cm X 12cm with a memory defect density of 0.8 and a logic defect density of 0.4. Through redundancy, the yield can be improved, but determining the type and quantity of redundant elements needed for a given memory requires both memory design knowledge and failure history information for the process node under consideration. Yet, simply providing the right redundant elements is not sufficient. Both the ability to detect and locate the defects in the memory and an understanding of how to allocate the redundant elements require manufacturing knowledge of defect distributions.

In order to achieve the optimized yield solution represented by the orange line in Figure 2, test and repair algorithms that contain these capabilities must be utilized.

![Figure 2. Using redundant elements to improve yield of embedded memory](image)

Conventional memory test algorithms detect memory failures to determine whether or not a chip is defective. For repairable memories, however, fault detection is not enough. Repairable memories need fault localization to determine which cells must be replaced. The greater the fault localization coverage, the higher the repair efficiency and therefore the obtained yield. It is very helpful to localize the exact coordinates of failed bits and associated fault classification to understand the root cause of failures. **Title-1**

**Soft error correction**

The continued scaling of complementary metal-oxide semiconductor device technologies has lead to ongoing device shrinkage and a decrease in the operating voltage of the device transistors (Vdd).
Scaling has meant denser circuitry overall, thinner silicon (e.g., silicon-on-insulator) in logic applications and less charge on capacitors for volatile memory. Even low-energy alpha particles can flip a memory bit or alter timing in a logic circuit of these smaller, lower voltage chips, making them more susceptible to soft errors. In many cases, these soft errors are self-inflicted because alpha particles are commonly generated in materials adjacent to the chip, solders and in the packaging.

With growing circuit complexity and the converging demands for high reliability and safety, transient errors are no longer restricted to space applications, but are impacting today’s consumer electronic applications, such as biomedical, automotive and networking applications. With this greater susceptibility to soft errors and an increased focus on field reliability, an embedded memory test solution that provides error correcting code (ECC) to detect multi-bit errors and correct transient errors enables designers to meet the target failure-in-time (FIT) rate for their applications.

**Test Considerations for 3D-ICs**

In the quest to integrate more functionality into increasingly smaller form factors with higher performance, lower power, and reduced cost, semiconductors that have traditionally used two-dimensional planes are now taking advantage of the third, vertical dimension. This latest evolution is referred to as a three-dimensional stacked IC, or 3D-IC, and consists of a single package containing a vertical stack of naked dies that are interconnected by means of through-silicon vias (TSVs), as shown in Figure 3.

TSV-based 3D-IC and silicon interposer based 2.5D-IC technologies, shown in Figure 3 and Figure 4, allow the semiconductor industry to continue its pursuit of more functionality, bandwidth, and performance at smaller sizes, power dissipation, and cost; even in an era in which conventional feature-size scaling becomes increasingly difficult and expensive. As with all ICs, 3D-ICs and 2.5D-ICs need to be tested for manufacturing defects.
In general, the test content of 3D-ICs is similar to that of conventional 2D ICs, because the same defects can occur during manufacturing, resulting in the same faults, fault models, and test patterns. However, the newer 3D-IC technology does have the potential for intra-die defects and TSV interconnect faults that require new tests. Intra-die defects may result from the 3D-IC processing step of wafer thinning. While testing for manufacturing defects in TSV-based 3D-ICs is in the early
stages, preliminary results indicate degradation of some I-V characteristics, shifts in device performance, and limited yield losses.

The TSV-based interconnect is a new structure present in 3D-ICs and may be prone to new types of defects. It is therefore necessary to identify these defects and determine how they behave as faults as well as how to model and test for them. Defects might occur in the TSV fabrication or in the TSV bonding between layers of the 3D stack.

Although the actual defect mechanisms are different, the resulting faults for TSV-based interconnects are similar to those associated with wiring interconnects, including opens, shorts, and delay faults. This makes it possible to leverage existing test algorithms to detect TSV interconnect faults through a set of digital test patterns. These test algorithms require full controllability at all interconnect inputs and full observability at all interconnect outputs.

Just as system architects can redesign and re-optimize their system architecture based on the 3D-IC technology, design-for-test (DFT) architects can redesign and re-optimize their DFT architecture, especially as it relates to test resource partitioning, or determining which DFT resource to put into which die. This can be demonstrated with an example of a 3D-IC product consisting of a memory die stacked on top of a logic die. The memory die provider may only make standalone memories, which often do not come with BIST. However the memory I/Os can be accessed by the test equipment and test costs are typically reduced by multi-site testing, or the testing of many memory chips in parallel.

With a 3D-IC chip, die-level testing is conducted as it would be for a standalone memory. On the other hand, the stack test for the same memory die is conducted in a manner similar to testing for an embedded memory, for which BIST is the DFT methodology of choice. To support the use of BIST in these memories, one option is to have the memory die come ‘3D-prepared’ with an on-chip BIST engine, which offers the benefit of ensuring that no proprietary memory test content will have to be released. However, the memory BIST, which is a logic element, might be difficult to implement in the technology of the memory die.

An alternative solution to eliminate this issue is for the memory supplier to provide a description of a memory BIST engine that is implemented in the bottom logic die. In this case, the memory BIST operation is controlled from within the logic die with the stimuli and responses flowing into and out of the memory die through TSV-based interconnects. Any embedded memory test solution used with 3D-IC chips requires complete test, repair and diagnostics support for various external memory types, such as SRAM, DRAM, DDR2, DDR3, LPDDR2, and external memory interfaces such as DDR PHY and Wide I/O memory that might be implemented on the chip, in order to test the external memories and TSV interconnects.

**Process improvement**

Many embedded memory test algorithms are targeted at catching random defects. However, as a result of shrinking geometries and the introduction of new material in the fabrication process, systematic defects now have a significant impact on process yield, necessitating the implementation of other methods to address manufacturing defects in embedded memories.

Adequate diagnosis and failure analysis is necessary to discover the root causes of the yield-limiting factors and accordingly update the manufacturing process. To expedite diagnostics, the embedded memory test solution needs to rapidly, cost-effectively, and accurately identify, analyze, isolate, and classify memory faults as designs are readied for transition from first silicon to volume manufacturing. A solution with automatic test vector generation, fault analysis, and root-cause failure guidance based on silicon test results would enable test and product engineers to rapidly
analyze failures manifested in embedded memories and inspect the physical location and class of each fault to determine the root cause.

**Improving time-to-volume**

SoC complexity is increasing as are time-to-market pressures, forcing SoC providers to find ways to get to volume production faster. Time-to-volume (TTV) consists of SoC design time and production ramp-up time. Reducing SoC design time is not a recent discussion. Reusing pre-designed cores and ensuring ease of integration is a viable way to address the growing SoC design gap and it is common for SoC designers to obtain embedded memories from IP providers.

Traditionally, yield optimization is done during the ramp-up period following the design stage. However, because of tight time-to-market windows, the ramp-up period of an SoC may start before the traditional defect densities and corresponding yield maturity levels are achieved, which results in a longer ramp-up time. Yet it is possible to reduce the ramp-up period, and therefore TTV, if the yield optimization effort starts before the ramp-up period. For embedded memories this can be done if a memory IP provider conducts yield optimization during the IP design and characterization phases before SoC ramp up by:

- Fabricating memory IP test chips, characterizing them and applying knowledge from the fabrication process to improve the yield of the memory IP block, which will result in silicon-proven memory IP before the SoC production ramp up starts.
- Designing all necessary memory repair functions into the memory IP and the SoC in advance, which augments the volume of fault-free SoCs and simplifies the ramp-up effort.
- Designing all necessary diagnosis and failure analysis functions into the memory IP, which will be the basis of the process improvement that will be performed during the ramp-up period.

Today’s embedded memories require solutions capable of addressing yield and reliability needs such as fault detection, repair at the manufacturing level, diagnosis for process improvement and field repair capabilities, all with minimal impact on the functional design. In addition, they must be able to address the test, repair and diagnostics requirements of emerging technologies such as 3D-ICs, which bring with them new types of defects. At the same time, these solutions need to minimize the manufacturing cost and reduce TTV. For example, Synopsys’ DesignWare® STAR® Memory System aims to be a comprehensive solution for embedded memory test, repair and diagnostics to optimize SoC yield, while staying within cost and schedule parameters.

**References**


**About the Authors**

Sandeep Kaushik is the Sr. Product Marketing Manager for the embedded memory test and repair products at Synopsys. Sandeep brings over 12 years of experience in the field of Design-for-Test. He holds a Bachelor’s Degree in Electrical Engineering from the Indian Institute of Technology in Delhi,
India and a Master's Degree in Electrical Engineering from Stanford University. He is currently working towards his Master's Degree in Business Administration from the Haas School of Business, University of California, Berkeley.

Dr. Yervant Zorian is the Chief Architect at Synopsys for embedded test & repair products. Formerly, he was Distinguished Member of Technical Staff AT&T Bell Laboratories, Vice President and Chief Scientist of Virage Logic and Chief Technologist at LogicVision. He received an MS degree in Computer Engineering from University of Southern California, a PhD in Electrical Engineering from McGill University, and an MBA from Wharton School of Business, University of Pennsylvania. He is currently the President of IEEE Test Technology Technical Council (TTTC), the Vice General Chair of Design Automation Conference (DAC), the Editor-in-Chief Emeritus of Design & Test of Computers, the founder & chair of IEEE 1500 Standardization Working Group, and an Adjunct Professor at University of British Columbia.