

Guest Editors' Introduction: Design for Yield and Reliability

Yervant Zorian
Virage Logic

Dimitris Gizopoulos
University of Piraeus

■ **TODAY'S SEMICONDUCTOR** fabrication processes for nanometer technology allow the creation of very high-density and high-speed SoCs. Unfortunately, this results in defect susceptibility levels that reduce process yield and reliability. This lengthens the production ramp-up period and hence affects profitability. The impact of nanometer technology on yield and reliability creates a dilemma for users of the conventional chip realization flow. Each chip realization phase affects manufacturing yield and field reliability. To optimize yield and reach acceptable reliability levels, the industry uses advanced optimization solutions, designed in and leveraged at different phases of the chip realization flow. Recognizing the importance of this topic, *IEEE Design & Test* has dedicated this special issue to design for yield and reliability solutions.

The ability to achieve acceptable levels of yield and reliability is likely to worsen as SoCs move to more aggressive technologies. To help solve this challenge, it is important to first identify the potential sources of yield loss and understand the faults that result in reliability failures.

Yield-loss factors

We can categorize the factors causing yield loss into four classes. The first is *systematic* yield loss. This typically results from the fabrication process and can be associated with a set of chips on a wafer. Because of a systematic process variation, these chips become nonfunctional.

The second class is *parametric*. This is not defect-related; rather, it is the design's sensitivity to process, temperature, and supply voltage variations that can affect the circuit's performance. Because we typically characterize a design over a parametric window, the foundry process must remain within this characterized window. Examples of factors causing parametric yield loss include variations in channel length, width, doping, and gate oxide thickness.

The third class is *defect-induced*. Here, the yield loss results from susceptibility to shorts and opens caused by

particles, contamination, cracks, voids, scratches, and missing vias.

The last class is *design-induced*. This occurs mainly because of physical problems that are highly layout-dependent and can have optical, electrical, or chemical effects. An example of a design-induced yield loss is photolithography with subwavelength feature sizes, resulting in feature corruption. Another example is chemical mechanical polishing (CMP) in wafer fabrication for metal layer planarization, where overpolishing of wide metal wires causes surface dishing. With the continuous advancements in semiconductor fabrication technologies, the semiconductor industry will see increasing levels of process variations, noise, and defect densities. These will add even more risk to the four previously mentioned yield-limiting factors.

Reliability faults

Similar to yield-loss factors, there are different types of reliability faults that are manifested in the field during the life cycle of the semiconductor product. The first type is *permanent faults*. These faults reflect an irreversible physical change. Improved semiconductor design and manufacturing techniques have decreased the rate of occurrence for this fault type.

The second fault type is *intermittent faults*. These faults occur because of unstable or marginal hardware activated by environmental changes such as lower voltage or temperature. Intermittent faults often become permanent faults. Typically, identifying a fault as intermittent requires failure analysis. This includes verifying if the fault occurs repeatedly at the same location, tends to result in bursts of errors, or if replacing the circuit removes the fault. Process variation is the main root cause of intermittent faults. Here are a few examples of the impact of process variation:

- Variation in etching rate can cause residual-induced

failures, which create smaller vias with higher resistance. Over time, this can turn into a permanent open fault.

- Variation in etching rate can also result in residuals on interconnects, which can cause an intermittent contact. This situation might eventually turn into a permanent short.
- A similar variation in layer thickness can cause electromigration in metallic or dielectric layers. This results in a higher resistance that manifests as intermittent delays. Over time, the high-resistance in interconnects can become permanent opens.
- The variation in layer thickness can also result in adjacent or crossing conductor signals, which cause intermittent contacts. This can, over time, turn into permanent shorts.

The third fault type is *transient faults*, also known as soft errors. They typically occur because of temporary environmental conditions. Possible causes of transient faults are neutron and alpha particles; power supply and interconnect noise; electromagnetic interference; and electrostatic discharge.

Trends in nanometer technologies are having a very negative impact on reliability because of shrinking geometries, lower power voltages, and higher frequencies. This is increasing process variation and manufacturing residuals, and, as a result, increasing the likelihood of intermittent faults. The same smaller transistors and lower voltages result in higher sensitivity to alpha particles and neutrons, thus causing significantly higher rates of particle-induced transient faults. Smaller interconnect features, causing the Miller effect, and higher performances, causing the skin effect, result in a higher number of transient timing errors. Finally, the increased coupling capacitance between adjacent conductors causes higher crosstalk noise and results in crosstalk-induced transient delays.

Given these nanometer trends, conventional techniques for improving yield and screening for reliability failures face serious limitations. For instance, the effectiveness of using I_{DDQ} , burn-in, and voltage stress during manufacturing for reliability screening faces challenges in the continued device scaling with each process generation. Increases in a device's quiescent current in the off state are raising the level of background current to the milliamperes, and in some cases the amperes, range. These levels of background current increase the difficulty of identifying microampere- to milliamperes-level I_{DDQ} fault currents. At the same time, the effectiveness of voltage and temperature acceleration methodologies used by burn-in

and voltage stress is declining because of the reduced margin between operational and over-stress conditions. As indicated in the *International Technology Roadmap for Semiconductors*, the increasing cost and declining effectiveness of conventional techniques for latent defect acceleration combine to create one of the most critical challenges for future process generations.

Solutions to optimize yield and improve reliability necessitate on-chip resources, known as infrastructure (embedded) intellectual property (IP) for yield and reliability. A wide range of such solutions, which we summarize next, are used in nanometer SoCs. A subset of these solutions is the topic of this special issue. The process of incorporating such functions into the design is known as *design for yield* and *design for reliability*.

Phases of the SoC realization flow

Infrastructure IP functions are useful in different phases of the SoC realization flow. These phases include process characterization; IP design and qualification; SoC design; silicon debugging; manufacturing test and packaging; process improvement; and field repair. Each of these should include yield and reliability feedback loops. A *feedback loop* includes three functions, namely *detection*, which provides the ability to identify the specific yield or reliability problem; *analysis*, which allows yield prediction (often using yield modeling); and *correction*, which improves reliability or optimizes yield.

Such feedback loops are becoming increasingly vital as processes move to smaller and smaller process nodes. For increased effectiveness, some of these loops reside fully on-chip as infrastructure IP, whereas others reside partly in off-chip equipment and partly on-chip. As you will notice, some of the design for yield and design for reliability loops focus on fault avoidance and others on fault tolerance.

Process characterization

During this step, the foundry characterizes the process and issues electrical design rules, including design for manufacturability (DFM) rules. Some of these rules are required and others are recommended. Often, during the early stages of a process node, the rules change frequently to optimize yield and reliability. The IP or SoC designers determine compliance to required or recommended DFM rules. Besides setting the rules, a foundry influences yield and reliability by selecting process material. For instance, using copper instead of aluminum is advantageous for reliability because copper provides a higher electromigration threshold.

Similarly, using silicon-on-insulator technology lowers circuit sensitivity to particle-induced transients. The foundry can further protect silicon from transients or substrate noise by introducing deep n-well in its process as a design-for-reliability solution. Similarly, to reduce particle-induced transients and reduce the FIT (failure in time) rate, the foundry can adopt a memory bit cell specifically designed with enlarged vertical capacitance and create its corresponding process steps.

IP design and qualification

To ensure adequate yield and reliability levels, the IP provider must qualify his hard IP core for each new technology node. With today's aggressive technologies, this effort has become quite challenging and sometimes requires either reiterating the IP design for yield or modifying the process. Typically, an IP provider complies with foundry-specific DFM rules, and uses manufacturing-friendly layout practices to guarantee the quality of his IP core. This might require tradeoff decisions to optimize for area, performance, power, yield, or reliability. An IP provider must optimize signal integrity to avoid yield loss. For logic blocks, he might choose to use latches that tolerate transient pulses for higher reliability. For higher yield, he might choose to limit lower-level metal layers for transistors and keep the upper-level metal for wiring. Instead of traditional synthesis, he might decide to use a yield-driven synthesis tool on his logic block (as described in the article by Nardi and Sangiovanni-Vincentelli, pp. 192-199). As for embedded memory cores, an IP provider might have to provide high-reliability solutions. In this case, he has the option of designing a special memory bit cell with increased capacitance using regular process steps, adopting tiling modification (horizontal versus vertical) in the memory design, or even adding error-correcting code blocks into the memory design. If he is required to provide high-yield solution, then he must augment the memory IP with the specialized embedded test and repair resources (as described in Vardanian, Shoukourian, and Zorian, pp. 200-206). In addition to the design techniques mentioned here, most IP providers use silicon testchips for their IP blocks. These blocks tape out and are characterizable on different process nodes, across multiple foundries, and for early and mature processes.

SoC design

Recently developed techniques, resources, and tools focus on design for yield and reliability. Their aim is to shield SoC designers from the complexities of the silicon-based manufacturing process. One set of design-for-yield

solutions is based on layout modification by, for example, optical techniques, such as phase-shifting masks (PSMs). PSM uses optical interference to resolve features to half-wavelength sizes. The second optical technique is optical proximity correction (OPC). In this technique, the designer applies the minimum amount of OPC necessary to meet the lithography goals. Another technique that can improve yield is CMP. Here, to avoid the thickness variation problem, back-end designers insert dummy metal fill—consisting of tiles inserted into empty areas of a chip—to even out the interconnect pattern density. This insertion occurs as a post-processing step. The article by Carballo and Nassif, pp. 183-191, describes some of these techniques.

Another layout modification technique is critical-area analysis (CAA). This consists of extraction from layout and then analysis; it is based on efficient algorithms for extraction from layout design, based on yield-relevant attributes. Another layout modification technique is modifying interconnect spacing, that is, wire spreading. Inserting redundant contacts and vias for additional improvement in yield can augment this technique. Also, the SoC designer can choose to replace cells with yielding variants while preserving the footprints of the original cells.

In addition to these design techniques, SoC designers might need to improve the yield further by modifying the design based on an understanding of the process node's yield-limiting factors. They can accomplish this by using test vehicles placed on a testchip or directly on the SoC, and extracting information from the test vehicle realized in silicon. The purpose of such test vehicles is to obtain design validation and reliability characterization; they must especially reveal marginalities in process, voltage, and temperature.

As for reliability, the SoC designer might need to protect the IP blocks by shielding them using metal layers, such as the ground/power mesh for memories. They also might design additional blocks such as those for error correction that protect valuable data (in memories, data paths, buses, and so forth).

Silicon debugging and process improvement

Adequate diagnosis and failure analysis techniques are essential to discovering the root causes of yield-limiting factors and performing appropriate process improvement steps. The migration towards smaller geometries severely challenges the physical failure analysis process. The key alternative is to gather failure data by using embedded diagnosis infrastructure IP—such as signature analyzers, dedicated test vehicles, or on-chip test processors—and

Back to Basics: The IC Industry's Reintegration and Dynamics of a Newly Emergent Full-Chip, Design-for-Manufacturability Infrastructure

Cary Vandenberg, HPL Technologies

Advances in semiconductor technology (specifically, microlithography) and related economics developments strongly point to the rapid emergence and significant potential of the new silicon-infrastructure, or design-for-manufacturability (DFM), market. The DFM market extends from design implementation, mask synthesis, and advanced lithography techniques, to yield enhancement and fab management. In 2003, analysts projected the market to grow 64% to \$220 million, versus a stagnant overall electronics design market with an expected growth of 2%. By other, broader, definitions of DFM, estimates of the new market's size are as high as \$2.5 billion. In any case, the challenges are significant, and the opportunity is compelling.

Today, this DFM opportunity lies fragmented among the solutions developed by integrated device manufacturers (IDMs), EDA and equipment vendors, foundries, mask makers, yield management, and fabless companies. There are also many promising startups at various stages of business development in this new market segment. Right now, it's clear that DFM is very full-chip lithocentric—focusing on, for example, optical proximity correction (OPC), phase-shifting masks (PSMs), and immersion—but is rapidly expanding to be process-centric (focusing on, for example, chemical mechanical polishing and etching) in the near future.

Some of the ultimate goals in DFM include isolating designers from process details and making design foundry portable. In that regard, the industry sees a limited number of DFM platforms emerging. Currently, the strongest visible commercial platforms are software solutions from Synopsys (the most comprehensive EDA-lithography flow), Mentor Graphics (an expansion of the popular Calibre tool), HPL Technologies (the most comprehensive yield-management and design-relevant flow), and, partially, from Cadence (based on their popular Virtuoso platform). Other significant players are Magma Design Automation, PDF Solutions, and several of the DFM startups: All are actively positioning themselves as DFM platform providers. The ideal DFM platform would manage all data, including design, lithography/mask, and production manufacturing.

Looking further ahead, I consider the Think-Spice methodology (proposed by former Cadence executive Jim Hogan) as an attractive path, combined with intensive real-time yield and testing solutions. A proposed global solution to design intent preservation is to "think Spice again!"

and to produce a standard similar to the E10 equipment standards of 12 years ago. E10 allowed for a common equipment protocol for utilization, capacity, performance, and status, thus enabling more efficient and sophisticated material movement and process control. The equipment suppliers and yield management community would most probably supply a similar model-based expression of DFM. The manufacturing community, which would likely consider its products proprietary, would provide the actual content for such a design intent model. Much like Spice, a design intent model describes behavior in manufacturing—not actual measurements. The manufacturing companies would run sophisticated and continuous design intent process monitors.

For example: Technology development organizations are creating full-field testchips containing tens of thousands of discrete test elements, but evidence at 90 nm shows that these solutions are not sufficient. The problem has also grown severe in the manufacturing line, where only a few dozen tests will fit into a manufacturing scribbleline. The DFM information chain is broken.

New techniques placing intensive test circuits into scribe-lines, coupled with established high-level analysis solutions, are offering significant improvement in test coverage of known design-manufacturing yield loss mechanisms. Manufacturing lines are enabled to perform meaningful correlations of die yield loss signatures to a rich set of systematic yield loss tests, with additional improvements in levels of statistics and also in improved spatial resolution. Longer term, we see the concept enabling die self-repair.

When the model is available and populated, a design automation solution will be able to allow a designer to extract, analyze, simulate, and optimize the circuit prior to the handoff to manufacturing. This will provide some confidence that the design will yield.

Clearly, the next step (since full-chip OPC is already established, and full-chip Spice is receiving market recognition) is the emergence of a new, full-chip DFM infrastructure based on a back-to-basics design-manufacturing reintegration. Think-Spice DFM and extensive continuous wafer characterization, combined with proper yield management methodologies, would be a sensible way to get there.

Cary Vandenberg is president and CEO of HPL Technologies. Contact him at cvandenberg@hpl.com.

Revolutionizing Manufacturing Processes in Very Deep-Submicron Designs

Philippe Magarshack, STMicroelectronics

Until 130-nm technology arrived, predicting IC yield was fairly straightforward: If you could manufacture each structure, you could manufacture the entire chip. What you drew on the layout tool was what you got on the wafer. Designers had simple physical-design rules with simple yes/no, pass/fail criteria for the design rule checker (DRC).

However, some signs of trouble were already noticeable as early as in the 0.25- μm process node:

- The effects of chemical mechanical polishing (CMP), a technique required for planarization of metal layers, had already forced designers to cope with protection diode insertion to compensate for the so-called antenna effect, starting at 0.25 μm . But at the time, the situation was simple: The design rules required one diode per μm^2 of gate area.
- At 180 nm, because of the dishing effect of polishing wide, empty areas, designers had to introduce some dummy polygons in the layouts of the metal layers to keep an average layer density within pre-defined bounds.

However, in 130-nm design rules, manufacturers use some 193-nm lithography scanners and some 248-nm equipment. Therefore, the critical dimensions of the smallest geometries on-chip have decreased below the lithography's capabilities. To compensate for the induced optical inaccuracies, process developers have introduced so-called optical-proximity-correction (OPC) techniques to modify the mask data just before sending it to the mask shop.

These techniques can't correct all complex geometries: For instance, to match a particular transistor's gate width/length with that of a neighboring transistor, the layout review requires the addition of dummy poly lines beside both transistors to obtain the best matching possible. Designers can no longer ignore the new fab-related techniques such as OPC: To produce robust designs, they must have insight into the realities of lithography and the fabrication process steps.

In 130-nm designs, most layout design rules are still mandatory, but new, qualitative sets of rules are appearing. For instance: Whenever possible, you should insert double vias between metal layers. This is a rule that a junior designer will easily bypass, producing a design that

will, by construction, yield poorly on silicon. Therefore, there is currently a further evolution to mandatory, but statistical, design rules. For example: For the GDSII data to be accepted at the fab, you should double 80% of the vias on-chip.

As designs move to 90 nm and smaller, the need to manage yield issues and adopt design-for-manufacturability (DFM) methods will be even greater. Because designers must weigh improvement in manufacturability against the effects on the design's performance, size, and functionality, they will need to participate in analyzing the variables that contribute to these design specifications. This means that manufacturability must start at the chip design level. It is not enough however, as some have suggested, to merely include DFM specifications in automated place-and-route tools. This method does not address how various components will react to each other when integrated into a full chip. As in traditional DRC manufacturing methods, DFM methodology must incorporate a full-chip approach; that is, it must make data available in its full context. This means having access to yield-limiting issues in a cross-layer and cross-hierarchical sense. Being able to look across hierarchical boundaries to see how the data in one cell interacts with data outside the cell is essential. It might be possible to improve the manufacturability of one layer by manipulating another. Similarly, a cell with no known manufacturability issues might significantly affect the manufacturability of a full chip when it is placed into context.

This more comprehensive model requires a new infrastructure that supports a feedback loop between designer and manufacturer. The feedback loop should include a means of defining and relaying manufacturing constraints, verifying IC layouts, and discussing manufacturing-related issues to the designer. To make this feedback loop possible, a complete cultural revolution must take place whereby designers understand the constraints of the lithography experts, and conversely, the yield production engineer participates up front in the design and design-for-test reviews, inducing a design-for-yield mentality in the design teams.

Philippe Magarshack is vice president of the Central Research and Development Group at STMicroelectronics. Contact him at philippe.magarshack@st.com.

then analyzing the obtained data by off-chip fault localization methodologies and tools. The article by Appello et al. (pp. 208-215) describes an example of such a system for random logic blocks. In the case of embedded memories, the dedicated infrastructure IP, that is, the test-and-repair processor, can gather the failure data at every error occurrence and transfer it to external analysis software, which builds the failed-bit map of the memory and performs statistical and graphical analysis on it. This type of infrastructure IP can also use the proposed IEEE standard P1500, which specifies a standard for accessibility to and isolation of individual functional IP blocks.

Manufacturing test and packaging

An effective way to obtain yield improvement for memories is to use redundant or spare elements during manufacturing test. Historically, embedded memories have been self-testable but not repairable. Recently, embedded memories, like stand-alone memories, have been forced to use design-for-yield approaches such as redundancy.

Because the timing specifications are often very stringent in today's SoCs, external instrumentation is not enough to ensure accurate measurement. For example, embedded timing IP is used as a design-for-yield solution. This IP distributes multiple probes over different parts of a SoC to collect the necessary timing information. Accurate timing measurement reduces unnecessary guard-banding and hence increases the SoC yield.

To enhance the reliability, test engineers can take additional steps at this phase. For example, the use of burn-in test reduces intermittent faults by accelerating their occurrence and eliminating the corresponding chips. To lower the particle-induced transient faults, this phase can use traditional low-alpha-emission interconnect and packaging material.

Field repair

Nanometer technologies make devices more susceptible to post-manufacturing reliability failures. One way to address this problem is to use remaining redundant elements in memories to perform periodic field-level repair or power-up soft repair.

Several design-for-reliability solutions provide online repair capabilities, including error-correcting Hamming code for failures unaccounted for during process characterization; Berger code for concurrent error detection in RISC processors; space and time redundancy of hardware or software implementations; error detection and firmware error recovery, such as in Intel Itanium proces-

sors; redundant functional units, such as in IBM S/390, G5, and G6 processors; and a wide range of system-level fault-tolerance techniques. This special issue presents two novel techniques that belong to this class of solutions (the articles by Breuer, Gupta, and Mak, pp. 216-227; and by Mitra et al., pp. 228-240).

DESIGN FOR YIELD and reliability solutions are very critical in this nanometer era. The wide range of options summarized here and in the three side bars associated with this introduction allow designers and manufacturers to choose from different options. The industry needs tradeoff analysis and return-on-investment procedures to select the most adequate options. Some tradeoff factors to look at include area, performance, power, cost, yield, and FIT.

We thank all who have contributed to this special issue including the article and sidebar authors, the reviewers, and especially the *D&T* staff and the editor in chief. ■



Yervant Zorian is vice president and chief scientist of Virage Logic. He previously was the chief technology advisor of LogicVision and a Distinguished Member of Technical Staff at Bell Labs.

Zorian has an MSc in computer engineering from the University of Southern California, a PhD in electrical engineering from McGill University, and an executive MBA from Wharton School of Business, University of Pennsylvania. He is the IEEE Computer Society vice president for technical activities, and a Fellow of the IEEE.



Dimitris Gizopoulos is an assistant professor in the Department of Informatics at the University of Piraeus, Greece. His research interests include self-testing of embedded processors; SoC and online testing; and yield and reliability improvement.

Gizopoulos has a PhD in computer science from the University of Athens, Greece. He is the Tutorials and Education Group Chair of the Test Technology Technical Council, a Senior member of the IEEE, and a Golden Core member of the IEEE Computer Society.

■ Direct questions and comments about this article to Yervant Zorian, Virage Logic, 46501 Landing Parkway, Fremont, CA 94538; zorian@viragelogic.com.