Embedding Infrastructure IP for SOC Yield Improvement

Yervant Zorian
Virage Logic
46501 Landing Parkway
Fremont, CA 94538
+1-510-360-8035
zorian@viragelogic.com

ABSTRACT
In addition to the functional IP cores, today's SOC necessitates embedding a special family of IP blocks, called Infrastructure IP blocks. These are meant to ensure the manufacturability of the SOC and to achieve adequate levels of yield and reliability. The Infrastructure IP leverages the manufacturing knowledge and feeds back the information into the design phase. This paper analyzes the key trends and challenges resulting in manufacturing susceptibility and field reliability that necessitate the use of such Infrastructure IP. It also describes several examples of such embedded IPs for detection, analysis and correction.

Categories and Subject Descriptors
B.8.1 [**Performance and Reliability**]: Reliability, testing and fault tolerance.

General Terms: Measurement, Performance, Design, Reliability, and Verification.

Keywords: Semiconductor IP, Embedded Test & Repair, Yield Optimization, Test Resource Partitioning.

1. INTRODUCTION
Every new semiconductor technology node provides further miniaturization and higher performance, thus increasing the functions that electronic products could offer. Although adding such new functions do benefit the end-user, but they also necessitate finer and denser semiconductor fabrication processes, which make chips more susceptible to defects. Today's very deep-submicron semiconductor technologies of 130 nanometers and below are reaching defect susceptibility levels that result in lowering the manufacturing yield and reliability, and hence lengthening the production ramp-up period, and therefore the time to volume (TTV). The very deep submicron impact on yield, reliability and TTV is very critical for the semiconductor industry. It puts the conventional IC realization flow at an impasse.

In fact, every single phase in the IC realization flow Figure (3) impacts yield and reliability. This includes the design phase, prototyping or production ramp up, volume fabrication, test, assembly, packaging, and even the post-production life cycle of the chip. In order to optimize yield and reach acceptable TTV levels, the semiconductor industry needs to adopt advanced yield optimization solutions. These solutions need to be implemented at different phases of the chip realization flow.

The conventional semiconductor manufacturing infrastructure, i.e. the external equipment and processes, alone are insufficient to handle such advanced yield optimization solutions; supplemental on-chip infrastructure is needed. To optimize yield and reliability, the industry has recently introduced a range of embedded intellectual-property (IP) blocks, called infrastructure IP [6]. These are meant for inclusion into IC design and utilized during the different phases of product realization.

Semiconductor IP is well known for the last decade. Most of the known IP blocks, though, are functional ones, such as embedded processor, embedded memory, embedded analog, or embedded FPGA cores. Whereas, Infrastructure IP (I-IP) is not functional, i.e. does not contribute to the normal functionality of a given IC. Rather, I-IP is embedded in an IC solely to ensure its manufacturability and lifetime reliability [6]. This role is similar to the infrastructure elements of a building, such as wiring networks or plumbing, which are independent from the actual function of the building.

This paper introduces the basic types of infrastructure IP and presents their effectiveness in improving yield and reliability. The paper is organized in the following way. Section 2 discusses the trends and challenges of today's very deep submicron technologies concentrating on the ones that have major impact on yield and reliability. Section 3 introduces the architecture of I-IP. Sections 4-8 demonstrate examples of I-IP solutions used at different phases of the IC realization and use flow. These solutions are based on on-chip resources for embedded diagnosis, timing measurement, debugging, test, repair, and fault tolerance.

2. TRENDS AND CHALLENGES
Every phase in the product realization flow affects yield. The following subsections describe the key trends and challenges of today's very deep submicron technologies that have considerable impact on yield.

2.1 Yield Learning and Time to Volume
The semiconductor fabrication process is constantly evolving to implement the advances necessary to realize the improvements in fabrication technology. This introduces new materials and techniques into the fabrication process, which lead to a new

generation of yield limiting faults [1]. Furthermore, the increase in design complexity and the shrinking geometries in very deep submicron technologies made devices more susceptible to systematic and random defects. Typically, yields are low for the initial lots of an IC design. This is called the ramp up period, as shown in Figure (1). During which the yield problems are learned (identified and diagnosed). As a result, the yield slowly ramps to a mature level. Then traditionally the volume production starts.

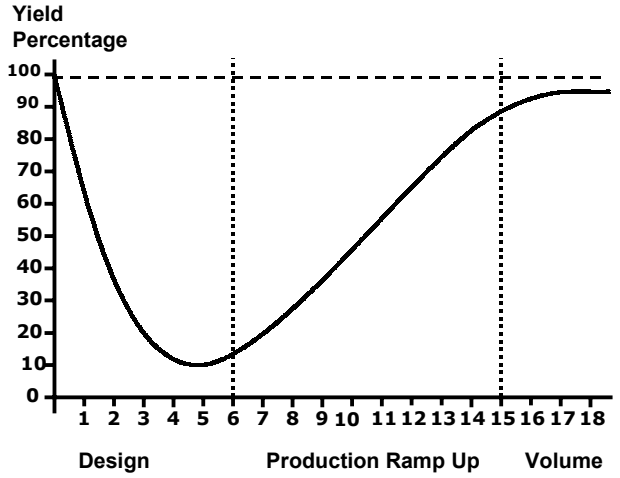

Figure (1) Yield Learning Curve (Source: HPL)

However due to the increase in time-to-market pressures, foundries are often forced to start volume fabrication on a given semiconductor technology before reaching the traditional defect densities, and hence the yield maturity levels necessary prior to volume production. Hence, improving yields as quickly as possible is an important component in lowering costs and improving profitability. The yield learning curve can be considerably improved, if the yield optimization can start at the design stage. This can happen by applying knowledge from the fabrication process into design. To optimize the design and obtain better yield, foundries should diagnose yield problems during the early development process. Collecting the manufacturing data, like defect distribution data, is done using special infrastructure IP blocks called embedded process monitoring IP, which will be discussed in Section 4.

Another trend is the supply chain disaggregation in the semiconductor industry due to the fabless model. The resulting supply chain includes fabless houses, functional-IP suppliers, pure-play foundries, and contract manufacturers for test, assembly and packaging services. In this disaggregated model, functional-IP providers must supply SoC designers with IP blocks optimized for functionally as well as manufacturability and yield. As a result, the advanced IP suppliers have begun assuming yield optimization responsibilities for their corresponding functional IP blocks, and thus embedding the necessary I-IP to ensure its manufacturability.

2.2 Memory Dominance

There is a clear trend to integrate large quantities of memory on a chip. Memories are designed with aggressive design rules and tend to be more prone to manufacturing defects. The overall yield of an IC or System-on-Chip (SoC) design relies heavily on the memory yield. Even though intrinsic or native memory yield may be unsatisfactory, memory and therefore, overall die yield, can be improved. The lower curve on Figure (3) shows the memory yield as a function of aggregate memory bit count. Traditionally embedded memories were testable but not repairable because embedded memories did need redundant elements for repair. The yield challenge in today's memories is addressed by offering memories with redundancy, i.e. spare elements. However, having redundancy only does not resolve the problem. The know-how of how to detect the defects in a memory and how to allocate the redundant elements require manufacturing know-how in terms defect distributions. In order to achieve the higher curve in Figure (2), i.e. optimized yield solution, one needs to utilize an infrastructure IP to contain the know-how, without which the yield can only be improvement to a limited level, as in Figure (2). An infrastructure IP to achieve optimized yield is discussed in Section 5, and is known as embedded test and repair IP.

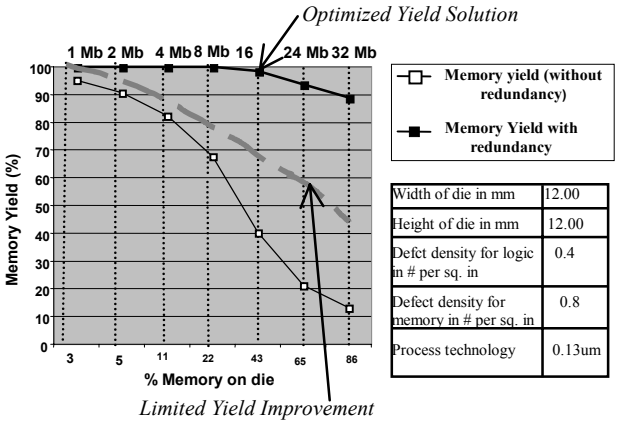

Figure (2) Memory Yield Optimization Curve (Source: Virage Logic Corp)

2.3 Diagnosis and Failure Analysis

The traditional failure analysis is comprised of fault localization, silicon deprocessing and physical characterization and inspection steps. The migration towards smaller geometries severely challenges this physical failure analysis process. Because of the increased sensitivity to failure, which requires finding smaller, more subtle defects; tighter pitches, which require greater spatial resolution; and increasing numbers of metal layers, which along with flip-chip packaging force the use of backside analysis [itrs]. These trends combined make the physical failure analysis process difficult to rely upon [6]. The key alternative is gathering failure data by using embedded diagnosis I-IP, such as signature analyzers, dedicated test vehicles or on-chip test processors, and then analyzing the obtained data by off-chip fault localization methodologies and tools. Examples of such embedded diagnosis I- IP are described in Section 6.

2.4 High Performance Circuits

Increasing performance in SOC designs require increased accuracy for proper resolution of timing signals. While semiconductor off-chips speeds have improved at 30% per year, tester accuracy has improved at rate of 12% per year. The tester timing errors are approaching the cycle time of the fastest device [5]. Yield losses due to tester inaccuracy are becoming a problem when using a

traditional functional test methodology. The yield loss in this case is due to additional guard-bending performed during the test stage. The alternate solution to address this challenge is to use Infrastructure IP for measuring and analyzing timing specifications. This alleviates the yield loss on high-speed devices due to tester timing accuracy. Section 7 describes an effective embedded timing IP.

2.5 Transient Errors

Drastic decreases in device dimensions and power supply have significantly reduced noise margins and challenged the reliability of very deep-submicron chips. Soft errors, timing faults, and crosstalk are major signal integrity problems [2]. Hence, both the logic block and the embedded-memories require self-correcting intelligence, such as an infrastructure IP for robustness. Section 8 introduces examples of embedded Fault Tolerance IP.

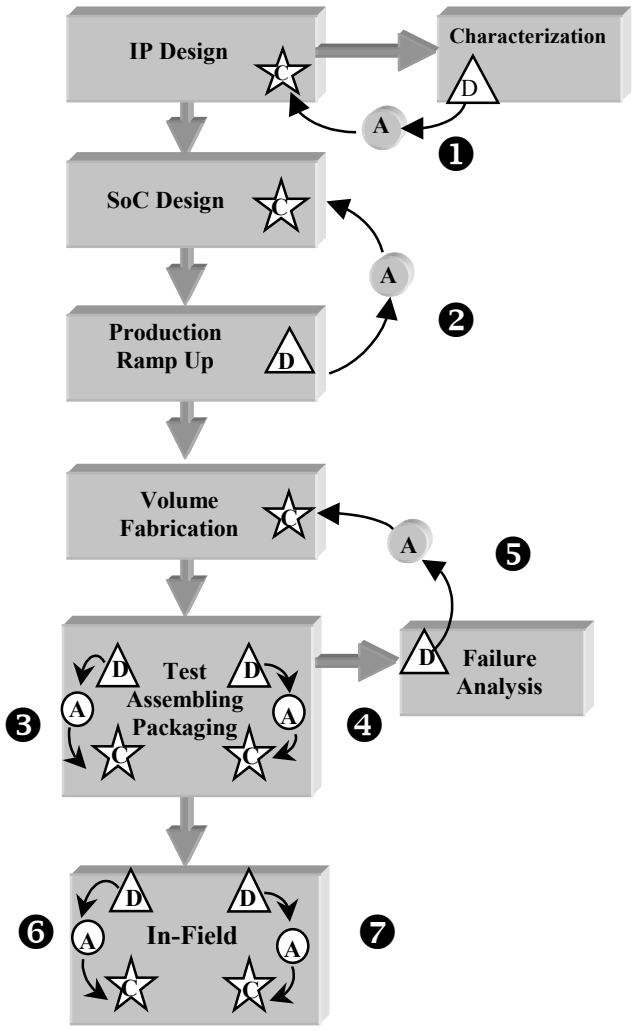

Figure (3) Yield Feedback Loops in IC Realization Flow

3. I- IP ARCHITECTURE

The trends and challenges described in Section 2 affect different phases of the IC realization flow. In order to improve yield and reliability a number of yield feedback loops need to be leveraged. Typically, a feedback loop, i.e. a yield enhancement solution, encompasses three consequent components: Detection, Analysis and Correction (DAC). Naturally, all three components are needed to achieve yield improvement. Due to the challenges of very deep submicron technology several of the feedback loop components (DACs) are embedded into the SOC design as Infrastructure IP. In fact, some of the feedback loops are completely embedded, such as loops 4, 5, 6, and 7; whereas others such as 1, 2, and 3 have only their detection monitors (D) only on-chip.

In the first wave, the infrastructure IP blocks were embedded mainly at the top level chip design and provided direct monitoring and control to the peripheries of different functional blocks. However, the trend has been towards integration of infrastructure IP with the functional IP. The following five sections will discuss examples of I-IP used in the yield feedback loops of Figure (3).

4. EMBEDDED PROCESS MONITOR IP

In loops 1 and 2, the D components are specialized I-IP solutions that monitor the process characteristics and collect device attributes. These process monitors are also called test vehicles or test chips. Such an I-IP can be used during the process development phase, or later during production, as shown in Figure (4). It may be incorporated in the IP design, as in loop 1 or in the SOC design, as in loop 2. It may be a full chip or an IP in a product chip [1]. The attributes collected by I-IP are fed to external analysis engines (A) and the results are utilized for modifications (C) to optimize yield. The A and C in loops 1 and 2 are performed not by an I-IP, but by external resources.

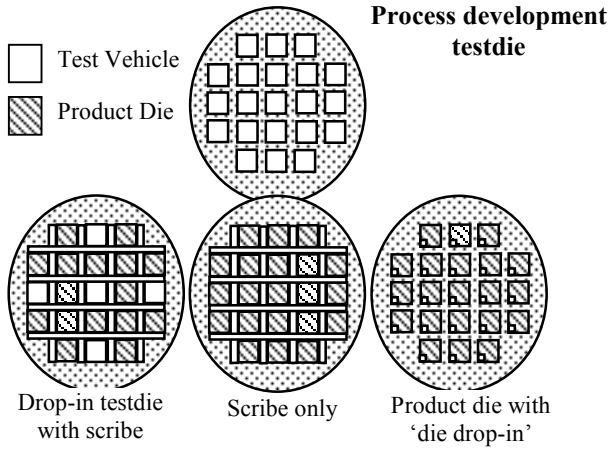

Figure (4) Embedded Process Monitor IP Application Modes (Source: HPL)

5. EMBEDDED TEST & REPAIR IP

The feedback loops 3 and 6 in Figure (3) correspond to this embedded test and repair IP. This I-IP is meant to address the embedded memory yield challenges, described in Section 2.2. The function of this I-IP is to perform embedded test, diagnosis,

redundancy allocation and repair signature generation. This I-IP is a comprehensive one, which includes the whole feedback loop, i,e. D-A-C. During the testing phase loop 3 is operational, which means that the embedded memories achieve optimized yields during manufacturing. Loop 6 becomes operational upon power up, i.e. for in-field test and repair. The type of I-IP requires integration the functional IP, i.e. the embedded memory. This composite IP, in addition to the embedded memories with redundancy, consists of a processor for embedded test and repair, intelligent wrappers associated with each memory, and a fuse box to permanently store the repair information on-chip.

6. EMBEDDED DIAGNOSIS IP

The feedback loop 5 of Figure (3) is meant to detect the root cause of failures using embedded diagnosis IP, and then to transfer the gathered data to external resources in order to perform analysis and correction steps to the fabrication process. Several examples of embedded diagnosis IPs can be mentioned here. In case of embedded memories, the dedicated processor in Section 5 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. Using this information, the fabrication process may be corrected.

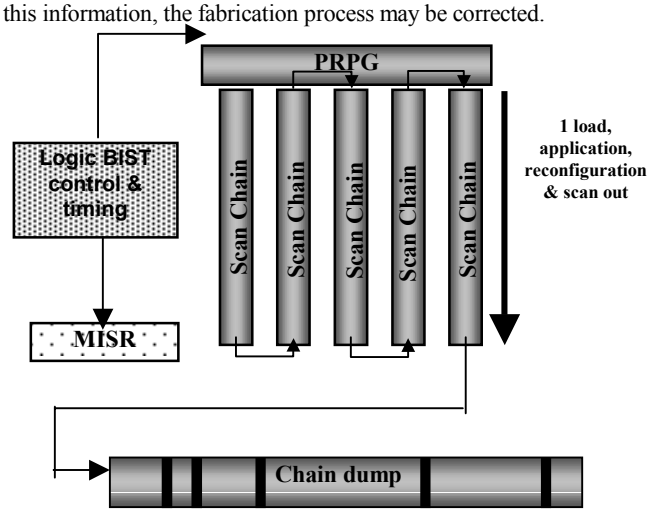

Figure (5) Diagnosis for Logic Blocks (Source: LogicVision)

In the case of random logic blocks, the embedded test and diagnosis IP is comprised of scan chains and test points incorporated into the random logic block itself, and an embedded signature analyzer on the peripheries of the random logic IP, as in Figure (5). This is another example demonstrating the integration of infrastructure IP with functional IP. The infrastructure IP operates with external software to allow interactive diagnostic modes. This allows flip-flop level diagnosis. A detailed description of this scheme can be found in [4].

The need to locate faults with further granularity is rising. Using dedicated test vehicles as, described in Section 4 and Figure (4), provides defect distribution data for each process layer and silicon structure [1]. Localization of performance fails is especially important. In addition, certain I-IP and external analysis tools do jointly locate defects to single transistor level. Some also handle realistic physical defects, including resistive bridges, resistive contacts/vias and opens.

This type of infrastructure IP may also use the IEEE proposed standard, P1500, which allows standard accessibility and isolation to individual functional IP blocks [3]. Designers implement this basic infrastructure IP around the peripheries of their individual functional IP blocks.

7. EMBEDDED TIMING IP

Because the timing specifications are often very stringent in today's SoCs, external instrumentation is not enough to ensure accurate measurement, as discussed in Section 2.4. An example of embedded timing IP is introduced in [5]. In this article, the I-IP achieves effective accuracy and is implemented via the feedback loop 4, as shown in Figure (3). This IP distributes multiple probes over different parts of an SOC to collect the necessary timing information. A central I-IP core controls the probes and transfers the information to a timing processor for analysis.

8. EMBEDDED FAULT TOLERANCE IP

The growing environmental susceptibility increases the reliability risks during the SOC life cycle, as described in Section 2.5. An Infrastructure IP is needed to operate in the field during the normal mode operation of the SOC. With this type of I-IP, the transient errors, including the soft errors, are detected, analyzed and corrected on-line, as shown in the feedback loop 7 of Figure (3). One well-known solution to perform embedded fault tolerance at the block level is described in [2]. The I-IP in this solution is fully integrated with the functional IP or SOC design. This allows for timing and area optimization and provides protection throughout the life cycle.

9. SUMMARY

This paper discusses the very deep submicron trends and challenges affecting manufacturing yield and reliability. It covers a wide range of infrastructure IP solutions to address these challenges — some used in the design phase, some in manufacturing; and others, in the field.

10. REFERENCES

[1] Bordelon, J. et al. "A Strategy for Mixed Signal Yield Improvement", IEEE Design & Test of Computers, Vol 19, No 3, May-June, 2002.

[2] Dupont, E., "Embedded Robustness IP", IEEE Design & Test, Vol 19, No 3, May-June, 2002.

[3] IEEE P1500 Web Site. http://grouper.ieee.org/groups/1500/.

[4] Pateras, S., "IP for Embedded Diagnosis", IEEE Design & Test of Computers, Vol 19, No 3, May-June, 2002.

[5] Tabatabaei, S., Ivanov, A., "Embedd Timing Analysis", IEEE Design & Test, Vol 19, No 3, May-June, 2002.

[6] Zorian, Y., "What is Infrastructure IP?", IEEE Design & Test of Computers, Vol 19, No 3, May-June, 2002.