Innovative Test Techniques for Advanced Technology Nodes

Darayus Adil Patel\textsuperscript{1,2}, Sylvie Naudet\textsuperscript{1}, Arnaud Virazel\textsuperscript{2}, Alberto Bosio\textsuperscript{2}, Patrick Girard\textsuperscript{2}

\textsuperscript{1}STMicroelectronics
850, rue Jean Monnet, Crolles, FRANCE
\textsuperscript{2}LIRMM
University of Montpellier / CNRS, Montpellier, FRANCE

E-mail: darayus-adil.patel@st.com

Résumé

The introduction of nanometer technologies, has allowed the semiconductor industry to keep pace with increased performance-capacity demands from consumers. This has brightened the prospects for future industry growth; however, new technologies come with new challenges. Semiconductor test costs have been growing steadily. Testing techniques for integrated circuits are today facing many exciting and complex challenges. In the era of large systems embedded in a single system-on-chip (SOC) and fabricated in continuously shrinking technologies, it is important to ensure correct behavior of the whole system. This work addresses the calls for the design and implementation of novel test strategies which provide high quality test. This paper within the context of semiconductor test, briefly describes the significant contributions made in the diverse areas of Reliability, Self-Heating, Ultra-Low Voltage Processors, Adaptive Voltage Scaling, Analysis of Timing Margins & IP Qualification.

1. Introduction

The ongoing evolution in microelectronics allows the semiconductor industry to create nanoscale devices in combination with gigascale complexity. Test is becoming a dominant factor in overall manufacturing cost. Furthermore, the semiconductor industry is extremely competitive and is asking for the best quality and reliability levels at the lowest cost. Although important advances have been made in the last period, existing test solutions are still unable to exhaustively cover all types of defects in new technologies (CMOS 28/14nm). Consequently, new technologies require innovative solutions to cope with new failure mechanisms under the constraints of higher density and complexity, cost and time to market pressure, product quality level and usage of low cost test equipment [1].

At STMicroelectronics, the development of new technology platforms is done through the validation of various test chips. From a test perspective, the main objective while developing test chips is therefore to anticipate the capture of all defects that might affect a prototype including a high density of gates, with all constraints related to test chips designed for the development of technology platforms. This work addresses the great need & scope to improve & optimize existing test structures & methodologies with regards to test time, silicon area occupied by test structures and the ability to analyze the impact of deep submicron effects. Additionally advances and innovations introduced in myriad domains of electronic design, reliability management, manufacturing process improvements etc. that call for the development of advanced, modular and agile test methodologies have been effectively covered within the scope of this work.

The rest of this paper is organized as follows: Section 2 discusses the Silicon Test, Verification & Reliability analysis of innovative test structures. Section 3 provides an overview of the Analysis & Mitigation of Self-Heating in FDSOI. Section 4 presents the Design and Performance Parameters of an Ultra-Low Voltage, Single Supply 32bit Processor implemented in 28nm FDSOI Technology. Section 5 discusses the Simulation & Silicon Verification of Adaptive Voltage Scaling for Real Applications. Section 6 talks about At-Speed test for IP Qualification. In Section 7 the Analysis of Timing Margins for Accurate Sign-Off is discussed. Finally Section 8 concludes this paper.
2. Silicon Test, Verification & Reliability Analysis of Innovative Test Structures

Considering the need for growing margins in nominal operating conditions to accommodate process variations in recent technologies, it is desirable to cut unused voltage margins while operating away from worst-case situations. To exploit unused margins, but avoid aggressive voltage scaling, the supply voltage has to be adjusted in a closed-loop control. To reach the real-time feedback of the chip’s operating condition, various classes of monitors [2, 3, 4] have been proposed in order to measure the circuit path delay. However the current State of the Art solutions are limited by the inherent complexity for the necessary recovery mechanisms, lack of real-time capability and the minimum delay constraint, which is similar to the hold-time constraint in latch-based designs. An innovative in-situ pre-error monitor namely the Canary Flag has been developed within ST [5] which allows achieving high monitoring performance with minimum overhead, timing impact and failure rate with respect to previous solutions [6, 7, 8]. Canary flags are critical path timing monitors and can be used to provide a pre-error warning mechanism before the timing margins on critical paths are violated leading to system failure. Fig. 1 shows a detailed schematic of the Canary Flag monitor. The main idea is here to probe the incoming Data right after the Master latch (and not prior as all previous approaches) in order to get rid of the additional load related to the monitor on the already critical path. Through adequate sizing of the delay element, it is possible to minimize the impact on the flip-flop setup time. The delay element is composed of minimum size inverter and a multiplexer so to allow a Test Mode (TM). The delayed signal is then passed through an additional latch, equivalent to the Slave latch of the reference flip-flop. In the implementation shown, a second latch has been introduced to act as an additional delay element, thus enabling two different monitor detection windows at the price to larger monitor area overhead. The monitor corresponding timing diagram is shown in Fig. 2. From left to right, the Data (D) is more and more delayed with respect to the clock (CLK) though the flip-flop output Q remains correct. In the first configuration, no warning flags are emitted since the remaining slack is large enough. For slightly reduced slack (intermediate situation), only the Flag2 is raised (corresponding to the large timing window) while Flag1 remains at zero. Finally, when the slack is further reduced, both Flag2 and Flag1 are raised. A comprehensive test solution enabling the silicon characterization of Canary Flags has been developed and realized. Additionally an adaptive voltage & wearout management strategy using the Canary flags has been implemented. Using this technique, overall up to 40% static and 25% dynamic power reduction has been demonstrated over a large range of DVFS operating conditions while operating in a failure-free mode. The developed test methodology used in conjunction with Canary flags offers new perspectives towards product hardening with respect to an adaptive approach to real user-based workloads.

3. Analysis & Mitigation of Self-Heating in FDSOI

Elevated chip temperatures are true limiters to the scalability of computing systems. Excessive runtime thermal variations compromise the performance and reliability of integrated circuits. To address these thermal issues, state-of-the-art chips have integrated thermal sensors that monitor temperatures at a few selected die locations. These temperature measurements are then used by thermal management techniques to appropriately manage chip performance [9, 10]. A novel test strategy that enables the measurement of dynamic self-heating properties of FDSOI technology at a system level has been put in place. The developed test method utilizes an on-chip thermal sensor coupled with a representative high speed digital block capable of operating in the Ghz range using the frequency generated via an on-chip PLL. Results obtained at wafer level from the implemented test methods help to perform a comparative analysis of the self-heating properties of FDSOI with respect to FinFets. The developed methodology puts in place the basic building blocks for characterizing heat variations in complex Multi-Processor System on Chips (MPSoCs) using multiple on-chip thermal sensors. The ongoing work aims to put in place a comprehensive reliability and aging management strategy utilizing the Thermal sensor data in consort with other parameters like critical path degradation, input voltage and frequency.


Growing complexity and performance requirements of modern multimedia devices is pushing chipmakers to achieve better logic performance especially at low voltages. Herein the FDSOI technology has conclusively
Figure 3: Average microprocessor minimum voltage extracted at 1MHz

demonstrated its ability to achieve high speed at low operating voltages [11]. In the race for higher energy efficiency driven by the IoT (Internet Of Things) and battery-powered systems, lowering the device supplies to Ultra Low Voltage (ULV) has been intensively explored. This is the strongest lever to achieve low energy operation at the expense of low frequency and increased variability [12, 13]. A 32b SPARC microprocessor aiming low operating energy and stand-by power designed with specific standard cells, a 10 Transistors memory cell, and adapted place-and-route tooling margins has been realized [14]. A comparative analysis with similar implementations has been done highlighting the performance gain and power savings that are achieved by our design methodology and implementation technology. Wafer-level tests showed that our ULV adapted microprocessor has an operating range that is functional down to 0.33V and that the ULV able cache can save from 30% to 62% energy. The minimum voltage population extracted on 193 dies at 1MHz is displayed in Fig. 3.

5. Simulation & Silicon Verification of Adaptive Voltage Scaling for Real Applications

Modern CMOS processes vary in performance from die to die and wafer to wafer. Process variation causes measurable and predictable variance in the output performance of all devices but becomes particularly important at smaller process nodes. The extent of the variation is not always known since designers are provided a worse case set of parameters to use for timing closure simulation with the operating voltage which takes into account the slowest silicon. With the emphasis on lowering power consumption a concern for system designers, Adaptive Voltage Scaling (AVS) is now a key power management feature that enables power reduction through voltage reduction. In the most basic form, Dynamic Power consumption for digital circuits can be considered as $P = CV^2F$ where $C$ is the switching capacitance, $V$ is the supply voltage, and $F$ is the clock frequency. The leakage power is also voltage dependent. Therefore significant power savings can be made through AVS which delivers accurate voltage to each individual device across temperature, process corner, and frequency variations. AVS is applied by reducing the voltage of fast silicon, to a point where speed target (Fmax) is met, thereby reducing power consumption as shown in Fig4.

A novel time conserving simulation flow for digital circuits has been developed. The said simulation flow enables a 15X gain in CPU time for Adaptive Voltage Scaling simulation for digital circuits [15]. The major gain is due to use of CustomSim-VCS capabilities, which allows optimizing testbench behaviour, thanks to communication between bench and fast spice environments. This has made possible the detailed performance assessments that have previous been considered as prohibitive and opens the door to investigate scaling of this approach to complete industrial
IP or System on Chip designs. A summary of the simulation flow enabled using Synopsys tools and the resulting runtime improvements achieved are shown in Fig. 5.

Secondly, a comprehensive test characterization methodology has been implemented for a representative high speed digital IP designed to address adaptive voltage scaling in 28nm bulk technology [16]. Performance characterization of this block enables the understanding and provides a deeper insight into cutoff frequencies and power supply saving dependencies versus process variations. The characterization methodology utilizes a low area ring oscillator based frequency generator, thereby laying the basis for the development of characterization solutions for future multi tile SOC designs which may employ individual per tile clock source that are synchronized via Network-on-Chip architectures. The silicon characterization has also led to a more profound appreciation of the accuracy and capabilities of a novel simulation methodology used to evaluate the extent of AVS in complex SOC designs. The observed results on silicon for the metrics of performance & power display a close correlation with the ones obtained via our simulation flow, thus providing designers with deeper insights into the gains with AVS from within the simulation environment. Fig 6 shows the normalized values of silicon results at 30°C for dynamic power saving through AVS for a chip fabricated using the Fast process corner.

6. At-Speed Testing for IP Qualification

Advanced nanometer technologies have led to a drastic increase in operational frequencies resulting in the performance of circuits becoming increasingly vulnerable to delay variations. At-speed testing is an effective approach to screen timing defects and improve test quality [17]. However, the increasing process spread in advanced nanometer nodes poses considerable challenges to predicting hardware performance from timing models. Thus there is a great need to qualify basic building structures on silicon in terms of critical parameters before they could be integrated within a complex SOC. Further, power consumption during test is significantly higher than during normal functional mode [18]. The excessive power consumption can result in parts being damaged during test, or in preventing good parts from passing test, leading to yield loss. This necessitates the need to develop power-aware designs & test methodologies. An innovative test-chip design test chip design and test methodology that enables the characterization of basic standard cells or complex IP structures and helps in detecting Small Delay Defects (SDD) at an early phase of technology development is presented. The focus is on detection and elimination of timing related defects at the standard cell level. The power-aware modular design structure & test methodology allow for the application of at-speed Functional & at-speed ATPG patterns. The proposed framework and methodology simplify the procedure for fault diagnosis and outlier localization using Functional at-speed patterns. The associated results & analysis help in further understanding the effects of PV across multiple corners on timing related defects & design margins.

This basic building structure referred to as a TILE and is shown in Figure 7. A tile consists of N different standard cells or complex IP structures between launch and capture flops. The capture flop is referred to as the Scan Flop in the figure as it forms part of the scan chain. The Circuit Under Test (CUT) refers to a single standard cell / IP as shown in Figure 7. The cuts embedded within a tile for the purpose of qualification differ in terms of geometry & characteristics such as layout, orientation etc. A cut is activated and tested by either a rising or falling transition. An NXI multiplexer is used to select & observe the output of any cut within the tile which is then captured in the scan flop. The scan flops of multiple tiles are stitched together to facilitate the shifting out of the cut’s capture response. The excitation stimuli may be either generated on-chip or loaded externally depending on the mode in which we are operating our structure. For retaining observability in case of any faults within the scan chain, redundancy has been added using a Bypass Mux structure. The launch & capture flops as well as the multiplexers employed within the tile have higher reliability as compared to the cuts i.e. they have been qualified earlier or if the tile is used for the qualification of an entirely new technology node, these cells employ a relatively relaxed geometrical layout. The tile can be clocked dynamically by either using the slow tester clock or fast clock generated using an on-chip PLL. Each tile is replicated several times within an instance and is then interfaced with an internal on-chip data generator and an On-Chip Clocking (OCC) mechanism.
An instance is further replicated inside a Block with the goal of increasing the cell population and the number of instances is configurable. For tackling peak power issues during capture, the at-speed pulses for each instance are mutually skewed such that at any given point in time, the capture pulses for just a single instance induce activity. This is achieved by setting the required length of the Clock Control Shift Register of the OCC during the design phase and then programming it accordingly to provide the required skew amongst instances. This approach induces a maximum switching of X across all instances at any point of time during capture. In order to guarantee this clocking scheme on silicon, clock skew of zero must be ensured at the input of each instance.

Figure 9 shows a representative graph of the First Fail Frequency derived via Fmax characterization across a population of dies embedded within the test chip at a single PVT point. Analysis of the data allows us to quickly understand the global speed limitation across all instances within our block. The quantum of within die variation and within wafer variation is easily discernable from the results. Within die or within wafer variation refers to the difference between the maximum & minimum values of the first failing frequency of instances within a die or across a wafer. Characterization for our block was performed at various PVT points to understand the impact of process variations across different process corners. The derived data is also used to compare the silicon results with the sign-off estimates and this information is vital for calibrating CAD models to accurately reflect the performance on silicon. The failure datalogs of the instances having the lowest first fail frequency are considered for deeper analysis to locate the cells with the lowest speed tolerance.

7. Analysis of Timing Margins for Accurate Sign-Off

A novel yet simple test circuit for accurate characterization of Setup and Hold margins while taking into account effects of PV is presented. The circuit consists of a data path between launch and capture flops whose skew can be accurately controlled to cause hold / setup violations. This design has been used for the development of the 28nm FDSOI node and the associated relevant results and analysis provide deeper insights into the effects of PV and the constituent components of the timing margins observed on silicon. Understanding the sources of extra margin aids in precisely establishing de-rate added during sign-off such that fail-free operation can be ensured while deriving maximum performance from a given circuit. The proposed design uses only standard cells, which allows for easy implementation. The presented test structure can be used as sensor inside the chip along with SOC designs to improve the frequency of a design in post-silicon tuning after reading results from the sensor.

A test circuit namely the Clock Data De-rate (CDD) block has been designed as a hard macro to characterize and analyze setup and hold margins. This is done by measuring the variation of delay in the Clock and Data paths to validate the de-rate added during sign-off. The basic diagram for which is shown in Fig. 10. The circuit consists of a data path commencing and terminating in Launch and Capture Flops. In order to derive the data path for our design, an analysis was done on real world SoC’s and a representative set of cells was selected from the ones that were most commonly employed and the ones that were susceptible to variability. To prevent the possibility of metastability at the Capture Flop impacting test results in a negative manner, a sensor has been used to flag the occurrence of metastability. Two variable delay generators employing ring oscillators have been used to add different delays in paths of launch and capture clocks. The amount of delay on each individual path can be controlled by setting the configuration of the Control Bus via top-level pins of the test-chip. By inducing different skew conditions, Setup / Hold failures can be created. Each path within the block is designed as a Ring Oscillator whose frequency is observable on the output ports after passing through an internal divider to account for the IO Pad and Tester limitations. As just a single transition is propagated along the clock and data paths, the flip-flop ring oscillator concept [19] has been used to find the
path delay along the clock and data paths. By considering the Setup / Hold margin at which design has been closed in conjunction with the data path delay and the skew at which failure occurs, we can accurately determine the extra margin along with its constituent components. Figure 11 shows the Within-Die variation of normalized extra margin along with its constituent components for the Hold Rise violations at the PVT point FF_1.3V_−40°C. Extra margin in Hold & Setup on silicon means the timing margin available in silicon over and above the sign-off passing value for setup & hold. From the graphical spread approximately 10% within-die variations in timing margins are observed. Such variations are expected in advanced technology nodes and require a deeper investigation & understanding of the constituent components of timing margins in order to enable the application of accurate de-rating factors. The graph of Figure 11 depicts the sources of extra margin available on silicon before Hold Rise violation occurs. Herein, the extra margin inside silicon for a single die is presented for analysis and a similar pattern is observed for different dies within the wafer. An analysis of test results reveals that Hold which is passing with 0ps of positive slack in sign-off is showing margins in Spice and Silicon. The reason for this extra margin is pessimism in clock and data de-rates.

8. Conclusion

This work presents an overview of the design and implementation of novel test strategies targeting advanced technology nodes for the purpose of:

- Verification of innovative test structures
- Analysis & mitigation of Self-heating
- Characterization & Performance evaluation of high-speed digital IPs
- At-speed IP Qualification
- Analysis of timing margins in silicon
- Reliability & process improvements

Références