Industrial Evaluation of Transition Fault Testing for Cost Effective Offline Adaptive Voltage Scaling

Mahroo Zandrahimi* Philippe Debaud† Armand Castillejo† Zaid Al-Ars*
*Delft University of Technology, The Netherlands
{m.zandrahimi, z.al-ars}@tudelft.nl
†STMicroelectronics, Grenoble, France
{philippe.debaud, armand.castillejo}@st.com

Abstract—Adaptive voltage scaling (AVS) has been used widely to compensate for process, voltage, and temperature variations as well as power optimization of integrated circuits. The current industrial state-of-the-art AVS approaches using Process Monitoring Boxes (PMBs) have shown several limitations such as huge characterization effort, which makes these approaches very expensive, and a low accuracy that results in extra margins, which consequently lead to yield loss and performance limitations. To overcome those limitations, in this paper we propose an alternative solution using transition fault test patterns, which is able to eliminate the need for PMBs, while improving the accuracy of voltage estimation. The paper shows, using simulation of ISCAS'99 benchmarks with 28nm FD-SOI library, that AVS using transition fault testing (TF-based AVS) results in an error as low as 5.33%. The paper also shows that the PMB approach can only account for 85% of the uncertainty in voltage measurements, which results in power waste, while the TF-based approach can account for 99% of that uncertainty.

I. Introduction

Power is one of the primary design constraints and performance limiters in the semiconductor industry. Reducing power consumption can extend battery life-time of portable systems, decrease cooling costs, as well as increase system reliability [1]. Various low power approaches have been implemented in the IC manufacturing industry, among which adaptive voltage scaling (AVS) has proven to be a highly effective method of achieving low power consumption, while meeting the performance requirements. Moreover, with the on going scaling of CMOS technologies, variations in process, supply voltage, and temperature (PVT) have become a serious concern in integrated circuit design. Due to die to die process variations, each chip has its own characteristics which leads to different speed and power consumption. The basic idea of AVS is to adapt the supply voltage of each manufactured chip to the optimal value based on the operation conditions of the system so that in addition to saving power; variations are compensated as well, while maintaining the desired performance.

A standard industrial approach for AVS is the use of on-chip PMBs to be able to estimate circuit performance during production. AVS approaches embed several PMBs in the chip architecture so that based on the frequency responses of these monitors during production, the chip performance is estimated and the optimal voltage is adapted exclusively to each operating point of each manufactured chip [2]–[7].

However, trying to predict performance of the many millions of paths in a given design based on information from a single unique path could be difficult and in many cases inaccurate. This results in high costs, extra margins, and consequently yield loss and performance limitations. This approach might work for very robust technologies and when only very few parameters influence performance, such as voltage, process corner, and temperature. However, in deep submicron technologies, as intra-die variation and interconnect capacitances are becoming predominant, it is more complex to estimate the performance of the whole design based on few PMBs. Hence, to improve the accuracy, we should use an alternative approach that increases the number of paths we take into account for performance estimation. Moreover, the more the characterization effort can be reduced, the more cost effective the AVS approach will be.

Previous work in this context, such as [9] and [10], propose techniques for generating optimal set of delay test patterns during the characterization process, which guarantees to invoke the worst-case delays of the circuit. These tests are applied on a small set of chips selected from a batch of first silicon to expose systematic timing errors that are likely to affect a large fraction of manufactured chips so it may be addressed via redesign before the design moves into high-volume manufacturing. However, they do not propose test generation for the purpose of application to AVS during manufacturing on every chip. Authors of [8] propose an efficient technique for post manufacturing test set generation by determining only 10% representative paths and estimating the delays of other paths by statistical delay prediction. This technique achieves 94% reduction in frequency stepping iterations during delay testing with a slight yield loss. However, the authors are only able to define static power specification for all manufactured chips, which is not able to address AVS utilization for each chip. [12] introduces a built-in delay testing scheme for online AVS during run time, which offers a good solution for mission critical applications. However, this requires significant software modifications, making it very expensive for non critical applications. [11] investigates the importance of delay testing using all voltage/frequency settings of chips equipped with AVS to guarantee fault-free operation. However, their approach does not enable setting optimal voltage and corresponding frequencies to enable AVS.

In this paper, we introduce a more accurate, cost effective approach for the estimation of AVS voltages during production (post manufacturing) using transition fault test patterns. We focus on test generation for application of AVS during the manufacturing process on every manufactured copy of the chip. Our work optimizes power based on the frequency specification defined at the design stage by setting optimal voltage for each chip to meet performance constraints. The contributions of this paper are the following:

- Proposing the new concept of using transition fault (TF) testing for AVS during production.
- A detailed investigation of the TF-based approach in terms of accuracy and effectiveness using 29 IS-

978-3-9819263-0-9/DATE18/©2018 EDAA

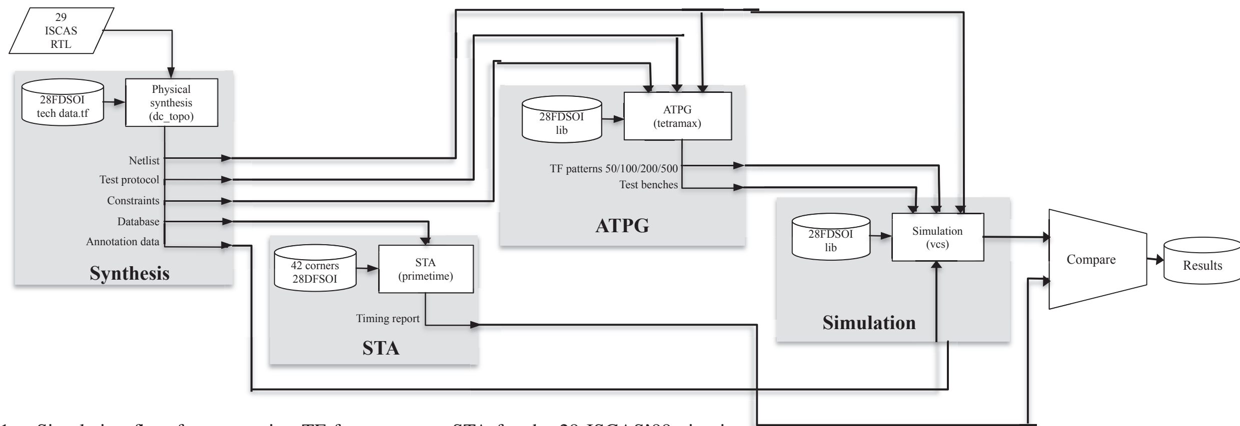

Fig. 1. Simulation flow for comparing TF frequency vs. STA for the 29 ISCAS'99 circuits

CAS'99 benchmarks with 28nm FD-SOI library for 42 different process corners.

This paper is organized as follows. Section II proposes the new approach of using transition fault test patterns for AVS. Evaluation of the proposed approach is presented in Section III using simulation results on ISCAS'99 benchmarks. Section IV presents the results of an industrial study performed on actual 28nm chips to validate the approach. Section V concludes the paper.

II. TF-BASED AVS

In this paper, we propose an innovative new approach for AVS using transition fault testing during production. Since transition fault testing covers many path-segments of the design, it can be a better performance representative than a PMB. Such an approach has a number of unique advantages as compared to PMB-based approaches. First, this approach can be performed *at no extra cost*, since delay tests are routinely performed during production to test for chip functionality. In addition, since delay testing is performed to explicitly test for actual chip performance, the expensive phase of correlating PMS responses to chip performance is not needed anymore, which reduces the length of the characterization stageand subsequently dramatically reduces cost and time to market. Moreover, as functional patterns are not used anymore, the delay testing approach could be a solution for general logic, and not only for CPU and GPU components, and last but not least, this approach makes using PMBs redundant, which saves silicon area as well as PMB design time.

TF testing is one of the three different types of delay test patterns, which includes small delay defect tests and path delay tests, in addition to TF tests [14]. TF test patterns target all gates and indirectly cover all path-segments. Hence, it covers all different kinds of gates and interconnect structures. Since several faults can be tested in parallel, we can achieve a high coverage with few patterns. However, automatic test pattern generation (ATPG) algorithms are based on heuristics like SCOAP [15], which tend to minimize computational effort. Thus, when several solutions are available for path sensitization, ATPG will use the easiest, which means that the algorithm tends to target the shorter paths rather than the optimal critical paths of the design [16]. On the other hand, we can alternatively use small delay defect testing, which sensitizes paths with smallest slacks, as well as path delay testing, which sensitizes a selected path. Among these two delay testing methods, path delay seems more promising since it sensitizes functional, long paths, which is an advantage over TF testing. However, in path delay testing the objective is to obtain a transition along those critical paths which are on average longer and more complex than the paths targeted in transition fault, thus reducing parallel testing capability and thereby reduces the overall coverage achieved. Therefore, we target TF test patterns in this paper for AVS during production since these give the highest path coverage of the three delay test alternatives.

The proposed flow performs a binary search to identify the minimum voltage (Vmin), at which the chip can pass all TF test patterns. The following steps are performed for each operation point of the chip: 1. Apply chip setup at nominal values and initialize variables, 2. Set supply voltage to Vmax and wait for stabilization, 3. Apply transition fault at speed test, 4. If the chip fails the test, discard it, otherwise, 5. compute new values and do a binary search to find Vmin. Conversion from Vmin to Fmax might be required depending on either performance estimation is done for yield enhancement or power optimization. "e" is an arbitrary value which is up to the users to define the resolution they want.

The basic requirement of using TF-based AVS is that there should be a reasonable correlation between TF frequency the chip can attain while passing all TF test patterns and the actual frequency of the chip. In this case, TF frequency could be a representative of actual chip performance. To investigate if this correlation exists, we perform simulations on ISCAS'99 benchmarks which contain 29 designs with different characteristics.

III. EVALUATION RESULTS

A. Simulation setup

This subsection explains the flow we used to explore if TF tests correlate with the actual frequency of the circuits. We use 28nm FD-SOI libraries to compare the transition fault maximum frequency versus the critical paths of ISCAS'99 benchmarks [18] using SYNOPSYS tools. ISCAS'99 contains 29 designs from small circuits with 21 cells to more complicated designs with almost 44K cells. 42 different corners of 28nm FD-SOI library have been used with different characteristics in terms of voltage, body biasing, temperature, transistor speed and aging parameters. We used Design Compiler in topographical mode for physical synthesis, Primetime for static timing analysis (STA), Tetramax for automatic test pattern generation (ATPG), and Vcs for back annotated simulation. Since

TABLE I. ERROR OF TF VERSUS STA

Benchmark	50p	100p	200p	500p	Benchmark	50p	100p	200p	500p
$b01$	1.00%	1.00%	1.00%	1.00%	$b15$	2.80%	2.75%	2.46%	2.06%
$b02$	3.81%	3.81%	3.81%	3.81%	$b15_1$	7.44%	7.38%	3.21%	2.57%
$b03$	4.04%	4.04%	4.04%	4.04%	$b17$	4.24%	4.21%	3.71%	3.68%
$b04$	2.97%	2.57%	1.70%	1.70%	$b17_1$	8.29%	5.26%	4.91%	4.91%
$b05$	1.28%	1.28%	1.21%	1.21%	$b18$	15.64%	12.25%	10.54%	6.47%
$b06$	3.64%	3.64%	3.64%	3.64%	$b18_1$	14.53%	7.89%	7.57%	7.47%
$b07$	5.83%	2.20%	2.20%	2.20%	$b19$	17.80%	15.90%	15.98%	12.42%
$b08$	2.84%	2.00%	2.00%	2.00%	$b19_1$	8.83%	8.82%	8.82%	8.82%
$b09$	7.50%	7.50%	7.50%	7.50%	$b20$	13.23%	12.53%	12.29%	10.00%
$b10$	0.05%	0.05%	0.05%	0.05%	$b20_1$	15.99%	15.70%	11.48%	9.94%
$b11$	2.19%	0.46%	0.20%	0.20%	$b21$	12.62%	12.82%	7.62%	7.62%
$b12$	1.82%	1.82%	1.82%	1.67%	$b21_1$	4.96%	4.47%	4.45%	3.42%
$b13$	2.35%	2.35%	2.35%	2.35%	$b22$	11.22%	10.38%	10.27%	10.27%
$b14$	18.55%	18.52%	18.52%	11.29%	$b22_1$	12.05%	12.01%	11.94%	8.54%
$b14_1$	19.23%	14.01%	14.01%	13.66%	-	-	-	-	-

functional pattens are not available for ISCAS'99 benchmarks, we use STA instead as a reference for comparison versus TF frequencies. This choice can be justified by noting that any set of functional patterns cannot be complete, since it is very tricky to select an application which reflects the real system performance specially for complex systems. Also, we note that identifying the most critical part of the application is not possible in most cases.

Fig. 1 shows the simulation flow containing 4 steps as follows:

- **Synthesis**: physical synthesis on 29 ISCAS'99 circuits using 28 nm FDSOI physical library to extract the netlists, and other reports required as an input for STA, ATPG and back annotated simulation. (29 netlists and other reports)
- **STA**: timing analysis using 42 corners of 28nm FD-SOI library to extract the critical timing of benchmarks in each corner. (42 corners*29 netlists= 1218 critical timing reports)
- **ATPG**: transition fault test pattern generation to extract test patterns and test benches for each benchmark. We generated 4 pattern sets (targeting only register to register paths) for each benchmark including 50, 100, 200, and 500 patterns. (29 netlists * 4 pattern sets = 116 pattern sets and test benches).
- **Simulation**: applying transition fault test patterns on back annotated simulation of each benchmark, and searching for maximum frequency at which each device passes the test. Transition fault frequency search is done using binary search and STA results as a starting point since TF maximum frequency cannot exceed critical timing.

Finally, we compared STA results versus transition fault frequencies of 29 ISCAS'99 circuits in 42 corners. The results are presented in the next subsection.

B. Simulation results

To understand if TF testing is a reasonable performance indicator that can be used for AVS during production, we compared the maximum frequency at which TF can be performed for each benchmark versus STA results.We estimated the performance of each benchmark in each of 42 corners both using STA and TF. In order to present the results, we define a parameter named $error$ which is measured for each benchmark. The concept relates to how much margin should be taken into account due to inaccuracies as a result of performance estimation using TF. To be able to measure $error$ for each benchmark, first we measured performance error for each corner by:

$$error_{corner} = (P_{STA} - P_{TF})/P_{STA} \quad (1)$$

where P_{STA} is the performance estimation using STA, and P_{TF} is the performance estimation using TF for the corresponding corner. Once $error_{corner}$ is calculated for all process corners, $error$ can be obtained for each benchmark by:

$$error = \max_{all\ corners} [error_{corner}] \quad (2)$$

Table I presents the $error$ for all benchmarks. We generated the results for 4 pattern sets including 50, 100, 200, and 500 patterns. As it can be seen in this table, depending on the size of each benchmark, with increasing pattern count, the $error$ is reduced. Therefore, depending on the time invested on testing during production, the accuracy of performance estimation using TF can be improved. As mentioned earlier, for some small benchmarks such as b01 with only 30 cells, the error remains unchanged since there are no more patterns that can be used to increasing the coverage.

TABLE II. AVERAGE ERROR PERCENTAGE OF TF VERSUS STA FOR 50,100,200, AND 500 PATTERN SETS

Pattern count	Coverage	error
50p	43.21%	7.85%
100p	49.82%	6.81%
200p	55.01%	6.18%
500p	62.97%	5.33%

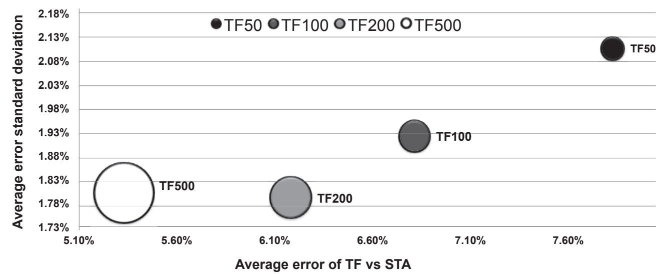

Fig. 2. Average error vs average standard deviation of error for 50, 100, 200, and 500 pattern sets; the size of the bubble represents the size of the pattern set used

As a conclusion, we presented the average error of all ISCAS'99 benchmarks for each pattern set in Table II. Increasing pattern count from 50 to 500 results in 19.76% increase in coverage, and thus 2.47% error improvement in average for ISCAS'99 benchmarks. According to these results, we can conclude that using transition fault testing for performance estimation achieves inaccuracy as low as 5.33%.

This measured $error$ means that in order to make sure the performance estimation using TF is accurate enough, a margin should be added on top of the estimated performance. If the inaccuracy of performance estimation using TF is predictable,

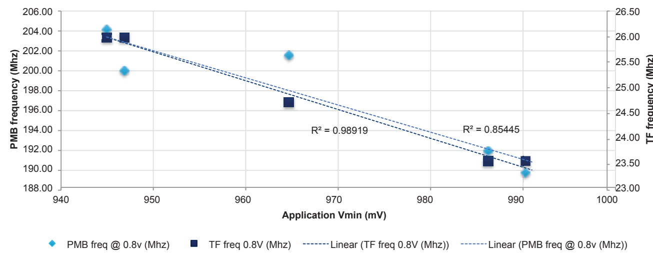

Fig. 3. Application Vmin versus TF and PMB results [19]

it is possible to come up with a safe margin. Figure 2 illustrates the average standard deviation of the estimation error plotted versus the average error measured using TF for all the circuits in the ISCAS'99 benchmark. The plotted measurements are clustered by the size of the test pattern set (reflected by the size of the circle in the plot). The figure shows that the larger the size of the used test pattern set, the more predictable the performance estimation will be. Therefore, depending on the time invested on testing during production, the accuracy of performance estimation using TF can be improved. However, moving from 200 to 500 patterns, the average standard deviation remained unchanged, which means that increasing pattern count up to a limit reduces the uncertainty. Further than that, the uncertainty remains unchanged even though the error is improved.

IV. Industrial case study

In this section, we compare PMB versus TF for AVS during production using measurements on real silicon. Our case study is a 28nm FD-SOI device on which a number of PMBs are distributed. These PMBs are ring oscillators designed based on the most frequently used cells extracted from the critical paths of various designs. During production, according to the frequency responses of PMBs, optimal voltage estimation is done for each chip. Alternatively, voltage estimation can be done using transition fault testing during production as well at no extra costs. Also, since transition fault testing represents a direct measurement of chip performance, the expensive correlation during the characterization phase is not needed anymore, which reduces time to market dramatically.

We have done silicon measurement on 5 chip samples. First, we have measured the real value of optimal voltage (Vmin) for each chip operating at its nominal frequency using functional patterns. Then, we set an arbitrary voltage for each chip (0.8v) and collected frequency responses from PMBs. We also measured TF maximum frequency for each chip using the same voltage settings (0.8v). To understand whether PMB or TF is more accurate for performance prediction, we have to identify which of them is more correlated with application Vmin. Therefore, we mapped both frequency response of PMB and the TF frequency to the Vmin of the chip in which that PMB is located. Then, we performed a linear least square regression analysis of the correlation between application Vmin and PMB frequency as well as the TF frequency, and measured the coefficient of determination (R^2) for both correlation functions. R^2 is a key output of regression analysis. It is interpreted as the proportion of the variance in the dependent variable (Vmin in this case) that is predictable from the independent variable (PMB frequency and TF frequency). Results are presented in Figure 3. R^2 for the correlation of application Vmin versus PMB is 0.85, while it has a value of 0.99 for the correlation versus TF, which means that Vmin estimation using the PMB approach can only account for 85% of the variability in the measurements, while Vmin estimation using the TF approach can account for 99% of that variability. These results confirm that we can achieve higher accuracy in Vmin estimation using TF [19].

V. Conclusions

This paper proposed a new approach that uses transition fault testing for AVS characterization during IC production, which serves as an alternative to the industry standard of using PMBs. This approach represents a powerful example of value-added testing, in which TF tests (already used during production) can replace a long and expensive process of PMB characterization, reducing cost and time to market dramatically. Moreover, since transition fault test patterns target all gates and indirectly cover all path-segments, it is a better performance representative than PMBs. As functional patterns are not used anymore, testing approach could be a solution for general logic, not only for CPU and GPU. According to simulation results of the 29 ISCAS'99 chips on 42 corners of a 28nm FD-SOI library, using transition fault testing for performance estimation ends up with an inaccuracy as low as 5.33% and a standard deviation of 1.8%. The paper also showed that the PMB approach can only account for 85% of the uncertainty in voltage measurements, which results in power waste, while the TF-based approach can account for 99% of that uncertainty.

References

[1] M. Zandrahimi and Z. Al-Ars, *A Survey on Low-power Techniques for Single and Multicore Systems*, in ICCASA, pp. 69-74, 2014.

[2] T. Chan and A.B. Kahng, *Tunable Sensors for Process-Aware Voltage Scaling*, in ICCAD, pp. 7-14, 2012.

[3] T. Chan, et al., *DDRO: A Novel Performance Monitoring Methodology Based on Design-Dependent Ring Oscillators*, in ISQED, pp. 633-640, 2012.

[4] A. Drake, et al., *A Distributed Critical-Path Timing Monitor for a 65nm High-Performance Microprocessor*, in ISSCC, pp. 398-399, 2007.

[5] TD. Burd, et al., *A dynamic voltage scaled microprocessor system*, in ISSCC, pp. 294-295, 2000.

[6] J. Kim and M.A. Horowitz, *An efficient digital sliding controller for adaptive power-supply regulation*, in IJSSC, vol. 37, no. 5, pp. 639-647, 2002.

[7] Q. Liu and S.S. Sapatnekar, *Capturing Post-Silicon Variations Using a Representative Critical Path*, in TCAD, vol. 29, no. 2, pp. 211-222, 2010.

[8] G. Li Zhang, et al., *EffiTest: Efficient Delay Test and Statistical Prediction for Configuring Post-silicon Tunable Buffers*, in DAC 2016.

[9] M. Sauer, et al., *On the Quality of Test Vectors for Post-Silicon Characterization*, in ETS 2012.

[10] P. Das, et al., *On Generating Vectors for Accurate Post-Silicon Delay Characterization*, in ATS 2011.

[11] N. B. Zain Ali, et al., *Dynamic Voltage Scaling Aware Delay Fault Testing*, in ETS 2006.

[12] K. N. Shim and J. Hu, *A low Overhead Built-In Delay Testing with Voltage and Frequency Adaptation for Variation Resilience*, in DFT 2012.

[13] M. Zandrahimi, Z. Al-Ars, P. Debaud, and A. Castillejo, *Challenges of Using On-Chip Performance Monitors for Process and Environmental Variation Compensation*, in DATE, 2016.

[14] M. Tehranipoor et al., Test and Diagnosis for Small-Delay Defects, in Springer Science+Business Media, LLC, 2011.

[15] L.H. Goldstein, E.L. Thigpen, SCOAP: Sandia Controllability/Observability Analysis Program, in DAC, 1980.

[16] B. Kruseman, A. Majhi, and G. Gronthoud, On Performance Testing with Path Delay Patterns, in VTS, 2007.

[17] M. Zandrahimi, P. Debaud, A. Castillejo, and Z. Al-Ars, *Using Transition Fault Test Patterns for Cost Effective Offline Performance Estimation*, in DTIS, 2017.

[18] http://www.cad.polito.it/downloads/tools/itc99.html

[19] M. Zandrahimi, P. Debaud, A. Castillejo, and Z. Al-Ars, *Transition Fault Testing for Offline Adaptive Voltage Scaling*, in ITC, 2017.