Challenges of Using On-Chip Performance Monitors for Process and Environmental Variation Compensation

Mahroo Zandrahimi*, Zaid Al-Ars*, Philippe Debaud†, Armand Castillejo†

*Delft University of Technology, The Netherlands

{m.zandrahimi, z.al-ars}@tudelft.nl

†STMicroelectronics, Grenoble, France

{philippe.debaud, armand.castillejo}@st.com

Abstract— Circuit monitoring techniques have been adopted widely to compensate for process, voltage, and temperature variations as well as power optimization of integrated circuits. For cost and complexity reasons, these techniques are usually implemented by means of performance monitors allowing fast performance evaluation during production. In this paper, we demonstrate the limitations of performance monitoring methodologies in terms of accuracy and effectiveness. Silicon measurements of a nanometric FD-SOI device show that the required design margin is above 10% of the clock cycle, which leads to unacceptable waste of power.

I. INTRODUCTION

As technology scales, circuit performance becomes extremely sensitive to process, voltage, and temperature variations (PVT). Furthermore, over time, circuit performance degrades due to different wear out mechanisms, such as NBTI, HCI, etc. One possible solution to make sure that the circuit works properly during its lifetime despite these sources of variation is to adapt operation parameters, e.g., supply voltage, exclusively to each chip. Measurement of operation parameters is done using various circuit monitoring techniques, which embed one or more performance monitors on chip. Thus, operation parameters can be measured through correlating frequency responses of performance monitors to the circuit frequency.

Various performance monitoring structures have been proposed from simple generic ring oscillators to more complicated design dependent critical path replicas. The technique presented in [1] implements replica-paths, representing the critical paths of the circuit. Alternatively, the critical path replica can be replaced by fan-out of 4 (FO4) ring oscillator [2] or a delay line [3]. The method presented in [4] synthesizes a single representative critical path (RCP) for post-silicon delay prediction. The paper suggests that the RCP is designed such that it is highly correlated to all critical paths for some expected process variations. However, as technology scaling enters the nanometer regime, specially from 45 nm onwards, finding one unique critical path has become impossible. Depending to the process corner, voltage and temperature variations, and also workload many different timing paths might become critical, therefore, for real circuits the concept of finding one critical path and create a critical path replica as a performance monitor is too simplistic.

In this paper, we evaluate the accuracy and effectiveness of using performance monitors for operation parameter estimation. The rest of this paper is organized as follows. Section II
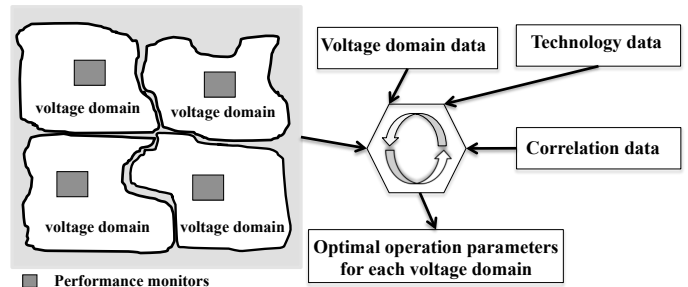

Fig. 1. Operation parameter estimation using monitoring approaches

discusses how circuit monitoring techniques work. Section III shows the limitations of performance monitors, first in terms of accuracy using static timing analysis of a nanometric FD-SOI device, and then in terms of power using silicon measurements of the same FD-SOI device. Section IV concludes the paper.

II. CIRCUIT MONITORING TECHNIQUES

Circuit monitoring techniques are widely used for adapting operation parameters exclusively to each chip for PVT variation compensation as well as power optimization [5]. These techniques embed one or various performance monitors on chip. Using the responses from these performance monitors, operation parameters are measured.

Fig. 1 shows an example of a chip with multiple voltage islands, among which performance monitors are distributed. There is no interaction between performance monitors and the circuit. To be able to estimate the circuit frequency based on performance monitor responses during production, the correlation between performance monitors and circuit frequency should be measured during characterization, which is an earlier stage of manufacturing. This correlation procedure is done for a number of test chips representative of the process window. During production, based on the frequency responses from these monitors, the circuit frequency is estimated so that operation parameters can be adapted to each voltage domain of the chip.

III. EVALUATION OF PERFORMANCE MONITORS

A. Accuracy evaluation

To investigate the accuracy of circuit monitoring techniques, we present some industrial experiences regarding critical path variability of a nanometric FD-SOI device through static timing analysis in sixteen corners having different process and environmental conditions. For each of the sixteen functional corners, we have extracted the 5000 most critical paths of the device. The path lists are sorted from the most
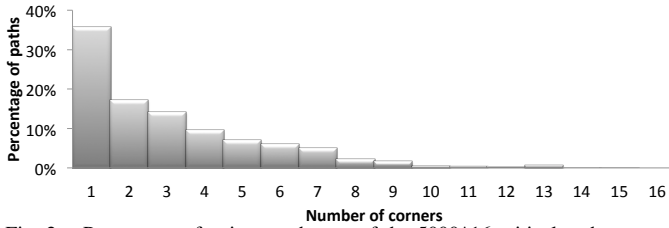
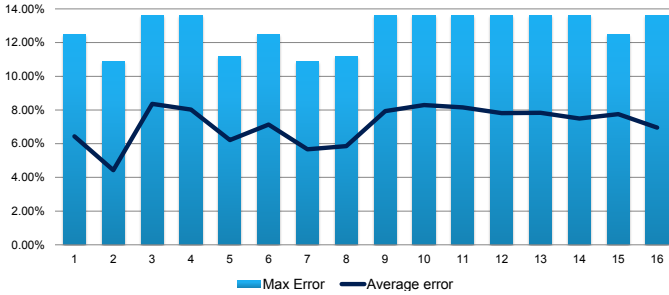Fig. 2. Percentage of unique paths out of the 5000*16 critical paths present in 1 to 16 corners

Fig. 3. Performance estimation error using the critical path of another corner

Fig. 4. Voltage discrepancy and power loss using different performance monitors

critical path to the least critical.

From the sixteen lists of 5000 critical paths, we have extracted the total number of unique paths. We have found 25936 unique paths out of 5000*16. Fig. 2 shows the percentage of the 25936 paths present in 1 or more corners. In this case, only 35.8% of paths are present in 1 corner, and only 53% are present in one or two corners. Two third of the paths are present in maximum 3 corners. None of the paths are present in the list of critical paths of all 16 corners, which means it does not matter which critical path we choose, it does not stay critical even within 5000 most critical paths of all corners.

These results show that identifying a critical path that covers all the corners is not possible. Therefore, when a path is the most critical in a corner, it is important to know how this path is changing across various process, voltage and temperature conditions. Suppose that P_x is the critical path of corner X, P_y is the critical path of corner Y. First, we have computed the distance of the P_x from P_y for all 16 corners against each other in terms of delay. Then, we measured the maximum as well as the average error for each corner if we assume that the critical paths of other corners are the most critical in that corner. Fig. 3 presents average and maximum error measured when the critical path of corner X is used to evaluate performance in corner Y. Results are presented in % of clock period and have been clamped to the value of the 5000th path of the corner Y list. Based on these results, whatever the critical path and the corner we take, the maximum error is above 10% of the clock cycle. As a result, regardless of using generic ring oscillators or design dependent replica paths, the characterization phase should be done to find the correlation between monitoring responses and the actual performance of the circuit.

B. Power evaluation

The process monitors, which are widely used today for many products, are ring oscillators designed based on the most used cells extracted from the potential critical paths of the design, reported by static timing analysis. So, based on the design, some standard logic cells are put in an oscillator to form performance monitors, which will be distributed among the chip to capture all kind of variations. During characteri-
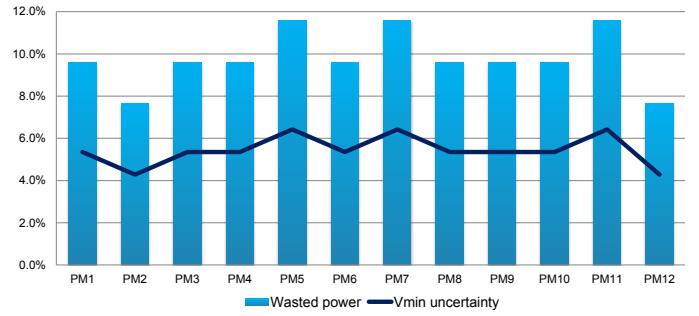
zation, performance monitors are tuned to the design so that during production, according to the frequency responses of performance monitors, the operation parameters are adapted to each chip.

We have done silicon measurement on 625 devices manufactured using nanometric FD-SOI technology on the same circuit as in SectionIII-A. 12 performance monitors (PMs) are embedded in each device. First, we have measured the real value of optimal voltage (Vmin) for each chip using test patterns. Then, we set an arbitrary voltage for each chip and collected frequency responses from all 12 performance monitors. Finally, we mapped each frequency response of a PM to the Vmin of the chip in which that PM is located. Results show variations of a PM frequency. We take the maximum amount of this variation as the Vmin discrepancy for that PM. We measured the amount of Vmin discrepancy for all 12 monitors, the result of which is presented in Fig. 4. This figure also presents the wasted power as a results of inaccuracy in Vmin estimation using performance monitors. Results show that minimum voltage estimation based on performance monitors lead to nearly 10% of wasted power on average and 7.6% in the best case, when a single PM is used for performance estimation.

IV. CONCLUSIONS AND FUTURE WORK

In deep sub-micron technologies, circuit monitoring approaches are showing limitations to accurately estimate silicon performance, which leads to unnecessary power loss. Based on static timing analysis of a nanometric FD-SOI device, we showed that depending on the design, a critical path can change dramatically as a result of PVT variations. Silicon measurements of the same device show that the required design margin is above 10% of the clock cycle leading to unacceptable waste of power. Thus, new power optimization methods are needed.

REFERENCES

[1] A. Drake, et al., *A Distributed Critical-Path Timing Monitor for a 65nm High-Performance Microprocessor*, in ISSCC, pp. 398-399, 2007.
[2] TD. Burd, et al., *A dynamic voltage scaled microprocessor system*, in ISSCC, pp. 294-295, 2000.
[3] 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.
[4] 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.
[5] M. Zandrahimi and Z. Al-Ars, *A Survey on Low-power Techniques for Single and Multicore Systems*, in ICCASA, pp. 69-74, 2014.