Characterization and Compensation of Performance Variability Using On-Chip Monitors

A.K.M. Mahfuzul Islam, and Hidetoshi Onodera
Graduate School of Informatics, Kyoto University, Kyoto, Japan
email:{mahfuz,onodera}@vlsi.kuee.kyoto-u.ac.jp

Abstract—**Aggressive technology scaling and strong demand for lowering supply voltage impose a serious challenge in achieving robust and energy-efficient circuit operation. This paper first overviews circuit techniques for variability resilience including on-chip circuits for performance and variability monitoring. We then focus on on-chip delay cells for transistor performance estimation and homogeneous and inhomogeneous ring oscillators for Die-to-Die (D2D) and Within-Die (WID) variability extraction. We also explain topology-reconfigurable on-chip monitors for in-situ variability characterization which can be used for D2D and WID variability modeling. The monitor can also be used for monitoring temporal variability such as Random Telegraph Noise (RTN). Compensation of performance variability can be done by a localized body biasing with on-chip monitors. A proof-of-concept circuit fabricated in a 65 nm process will be demonstrated such that a test chip fabricated at the slow process corner can achieve a target performance under the typical process condition by the compensation.**

I. Introduction

With aggressive technology scaling, the impact of variability on circuit performance is increasing [1]. The conventional worst case design (WCD) results in large energy, delay and area overheads. In the state-of-the-art LSI design, design margins are allocated in each layer of the design hierarchies. These margins are often set pessimistically, and therefore, large amount of energy loss occurs. In order to utilize the full potential of a process, dynamic tuning of supply and threshold voltage based on PVT (process, voltage and temperature) condition as well as circuit parameters such as operating frequency, activity, etc., has become a necessity. Transistor body biasing gives the designers an option to tune the threshold voltages. However, without knowing the transistor performance, tuning of threshold voltages may not result in energy reduction. Thus, on-chip monitors giving information on the devices come to play an important role. We discuss various variability resilience techniques first. Then, we discuss variability measurement and characterization techniques based on digital delay cells. A topology-reconfigurable circuit for in-situ characterization of static and dynamic variations will be presented. We also discuss the feasibility of performance compensation by localized body biasing with on-chip monitors.

The paper is organized as follows. Variability resilience design techniques are overviewed in Sec. II. Section III discusses variability modeling and estimation techniques using on-chip monitors. A topology-reconfigurable monitor circuit is presented in Sec. IV. A built-in variability compensation scheme is demonstrated in Sec. V. Sec. VI concludes this paper.

II. Design Techniques for Variability Resilience

Variability resilience can be achieved by a) Post-silicon de-skew method, b) Post-silicon static/dynamic supply voltage tuning, c) Post-silicon static/dynamic threshold voltage tuning, or d) Combination of the above techniques.

The de-skew method is applied by tuning programmable delay elements in the design. De-skew method can be applied to clock tree [2], or critical paths [3]. These techniques require additional circuitry and programmable devices to store the configurations which increases area and power.

Supply and threshold voltage tuning can be applied statically and dynamically. Static voltage tuning is performed during the test to meet the target frequency of the chip and the voltage setting is not changed over the lifetime [4], [5]. Dynamic voltage tuning is performed by run-time monitoring of path delays and adjusting the voltage just above the point where the error starts occurring. Run-time error detection/prediction and recovery techniques [6–8] are required for these techniques. Lower supply operation and the use of multi-threshold devices on the non-critical paths make almost every path a potential for causing timing error which introduces serious limitations to these run-time techniques.

Threshold voltage can be tuned in post-silicon by body biasing [9–11]. Forward bias is applied to slow chips and reverse bias is applied to fast chips to compensate delay and leakage variations. Adaptive body bias can be applied dynamically [10], [11]. nMOSFET and pMOSFET can be tuned independently to balance performance skews between the two devices [10], [11]. Adaptive body bias technique can be integrated with dynamic voltage and frequency scaling (DVFS) architectures to increase energy-efficiency [12].

The above techniques can be applied based on on-chip monitors rather that monitoring actual path delays [13], [14]. This approach has small area and power overhead compared to the detection/prediction circuits. Compensation based on simple critical path monitors (CPM) may not be energy-efficient as the skew between nMOSFET and pMOSFET performances may cause energy increase [11]. Independent monitoring of nMOSFET and pMOSFET performances is thus required.

III. Variability Modeling using On-chip Monitors

A. Variability Modeling Using On-chip Monitors

Typically, variability is modeled as the statitistics of several key transistor parameters such as threshold voltage, gate length,
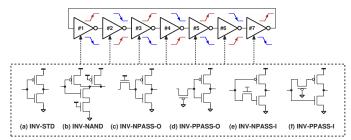
Fig. 1: Conventional ring oscillator circuit schematic. Inverter stages with fixed topology are used.
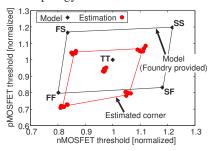

Fig. 2: Threshold voltage comparison between models and estimations.

etc. *I–V* data from large device arrays are analyzed to extract the parameter variations which is costly. In order to reduce measurement and implementation cost, several on-chip monitor circuits are proposed [15]. On-chip monitor circuits can be measured with simple equipments reducing measurement cost. Ring oscillators (RO) can be used as on-chip monitors [16–20].

B. Delay Cells for On-chip Monitors

Figure 1 shows several inverter structures used in an RO [16–18], [20]. The inverter structures shown in Figs. 1(a) and 1(b) are conventionally used to monitor process characteristics. Due to the inability of differentiating pMOSFET and nMOSFET variations, inverter structures of Figs. 1(c), 1(d), 1(e), and 1(f) are proposed [16], [18]. Using ROs consisting of these inverter structures, measurement and estimation of several process parameters are possible [19]. Pass-gate based inverter structures of Figs. 1(e) and 1(f) are reported to be useful for measurement of WID random and dynamic variations [20].

C. Ring Oscillators for Variability Characterization

1) Homogeneous RO: Homogeneous RO consists of inverters of the same structure. As a result, the variability between the stages are averaged out. Thus, this structure is suitable for die-to-die (D2D) variation modeling. Figure 2 shows threshold voltage estimation for several corner chips fabricated in a 65 nm process [19]. Some mismatches have been observed between the estimation and the foundry provided corner models. The use of on-chip monitors can thus give us the opportunity to update the variability models for model-hardware correlation.

2) Inhomogeneous RO: In order to evaluate the statistical properties of variation, large number of the same RO structure is measured [16–18]. Using the delay cells that are highly sensitive
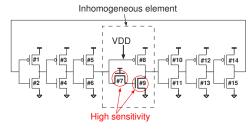

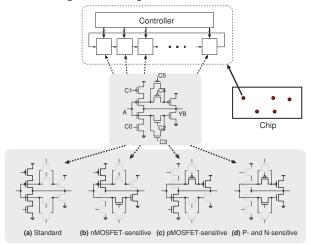Fig. 3: Inhomogeneous RO structure.

Fig. 4: Topology-reconfigurable RO for area-efficient monitoring of static and dynamic variations.

to a particular variability source, the statistical properties can be obtained [16], [19]. However, as the variability of a single transistor is averaged out, detailed evaluation of variations such as random telegraph noise (RTN) is difficult. Inhomogeneous RO structure, as shown in Fig. 3, enables transistor-by-transistor variability evaluation by increasing the visibility to a particular set of transistors [20]. For the structure in Fig. 3, the RO period sensitivities to the two nMOSFETs of the inhomogeneous stage are multiple times larger than the sensitivities to the other transistors. Thus, the delay variation can be approximated with the following equation:

$$\sigma_d^2 \approx k_{n,1}^2 \cdot \sigma_{V_{thn,pg}}^2 + k_{n,2}^2 \cdot \sigma_{V_{thn}}^2 + (k_{L,1}^2 + k_{L,2}^2) \cdot \sigma_L^2. \quad (1)$$

Here, $k_{n,1}$ and $k_{n,2}$ are sensitivity coefficients to threshold voltage variations for the two nMOSFETs of the inhomogeneous stage. $k_{L,1}$ and $k_{L,2}$ are sensitivity coefficients to gate length variations. Target parameters can be extracted utilizing the voltage dependency of the sensitivity coefficients [20].

IV. TOPOLOGY-RECONFIGURABLE ON-CHIP MONITORS FOR IN-SITU VARIABILITY CHARACTERIZATION

A. Topology-reconfigurable RO

The inhomogeneous RO structure enhances the sensitivity to small number of transistors. However, large number of ROs are required causing implementation cost and area overhead. Reconfigurable ROs can be used to measure variability between

the inverter stages [21], [22]. Reconfigurable RO can used to estimate global parameter values [23]. Figure 4 shows our proposed reconfigurable RO structure where reconfigurable inverter cells are used to monitor global and local variations for both of the nMOSFET and pMOSFET [24]. Each cell can be configured to several delay modes. The sensitivity of a particular stage's delay is enhanced by configuring the RO as inhomogeneous [20]. In order to identify the devices, we place two pass-gates in parallel. With this structure, device mismatch can be measured directly by taking the difference between the delays of two different pass-gate configurations. To identify devices with RTN effects, frequency fluctuation is observed for each of the pass-gate configurations. With an N-staged RO, $N \times 2$ types of inhomogeneous configurations for either nMOSFET or pMOSFET can be sampled. Thus, statistical properties can be obtained by choosing N to be sufficiently large. Global variation is measured by configuring the RO as homogeneous [18].

B. Device Level Variability Monitoring Results

Figures 5(a) and 5(b) show the measured nMOSFET and pMOSFET variability of a 65 nm test chip at 0.8 V supply. nMOSFET variability is larger than pMOSFET variability which agrees with device level measurement results [25]. Figure 6(a) shows the frequency fluctuation over time for a particular inhomogeneous configuration with nMOSFET pass-gate of "C2" signal is turned ON. Figure 6(b) shows the frequency fluctuation when "C3" pass-gate is turned ON. RTN induced binary fluctuation is observed for "C3" pass-gate. Thus, device level measurement becomes possible. Figures 7(a) and 7(b) show the comparisons between WID and RTN-induced variability for nMOFSET and pMOSFET. nMOSFET has larger variability.

V. Variability Compensation by Localized Body Biasing

A. Concept and Architecture

Applying body bias to the whole chip may not give us optimum energy-efficiency when large spatial variation exists. Variations also differ between blocks. Therefore, fine-grain compensation by localized body bias is required. Small area body bias generators can be used for local substrate islands where body voltages are applied dynamically based on on-chip monitors. Dynamic control can compensate time-dependent variations such as NBTI [26]. Figure 8 shows the proposed self-adjustment scheme by localized body bias for P/N-performance compensation [11]. The scheme consists of P/N-performance monitors, a controller, and DACs (Digital to Analog Converter) to generate body voltages for pMOSFET and nMOSFET independently. Only the system supply voltage and clock signal are used. Automatic place and route is possible using digital delay cells. Cell-based design of DACs is possible by dividing the circuit into several components compatible to standard cells [27].

B. Demonstration with Corner Chips

A proof-of-concept circuit was fabricated in a 65 nm triple well process targeting grain area of 0.1 mm^2 [11]. Area overhead
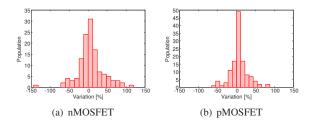

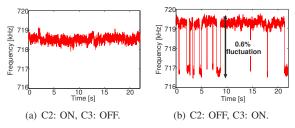Fig. 5: Measured delay variations for pass-gate pairs.

Fig. 6: Frequency fluctuation over time of an inhomogeneous RO topology for two different pass-gate configurations.
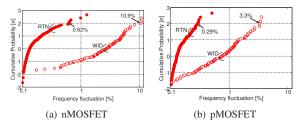

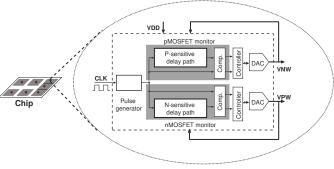Fig. 7: Comparison between WID and RTN-induced variability.

Fig. 8: Fine-grain performance compensation by localized body bias. Body biases are generated based on on-chip monitors for nMOSFET and pMOSFET independently.

is around 3 %. Test chips have been fabricated targeting "TT" condition, as well as four corners of "SS", "FF", "FS" and "SF", where the first letter denotes nMOSFET. Figure 9 plots the pMOSFET-sensitive speeds against the nMOSFET-sensitive speeds. 0.34 V of body bias is generated to compensate speed at "SS" corner to meet "TT" corner speed. Similarly, "SF", "SF" and "FS" chips are also compensated so that the target speeds are achieved.

Speeds of different logic gates are measured through RO frequencies to confirm the correlation to circuit speeds. Figure 10
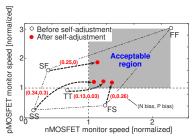
Fig. 9: pMOSFET-sensitive monitor speed against nMOSFET-sensitive monitor speed.
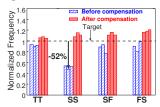

Fig. 10: RO frequencies for INV, NAND and NOR gates.

shows the values of INV, NAND and NOR RO frequencies for all the corner chips. After self-adjustment, all the gates achieve the target speeds. Thus, variability compensation by localized body bias is shown to be feasible.

VI. CONCLUSION

On-chip Monitor circuits having small area and digital nature can be placed into various places of the chip providing useful information on performance variability. Delay cells and monitor topology have been discussed in the paper to facilitate area-efficient variability monitoring. Utilizing localized body biasing, a built-in self-adjustment scheme to compensate local variation is demonstrated. By dividing the chip into several substrate islands, the scheme adjusts MOSFET performance of each island to their target values so that design margins can be minimized.

REFERENCES

[1] S. Borkar et al., "Parameter Variations and Impact on Circuits and Microarchitecture," in *Design Automation Conference*, 2003, pp. 338–342.
[2] S. Tam, R. D. Limaye, and U. N. Desai, "Clock Generation and Distribution for the 130-nm," *IEEE Journal of Solid-State Circuits*, vol. 39, no. 4, pp. 636–642, 2004.
[3] K. Mishra, A. Faraz, A. D. Singh, and A. Chatterjee, "Path Delay Tuning for Performance Gain in the Face of Random Manufacturing Variations," in *Internatioal Conference on VLSI Design*, Jan. 2011, pp. 382–388.
[4] T. Chen and S. Naffziger, "Comparison of Adaptive Body Bias (ABB) and Adaptive Supply Voltage (ASV) for Improving Delay and Leakage Under the Presence of Process Variation," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 11, no. 5, pp. 888–899, Oct. 2003.
[5] J. Tschanz, S. Narendra, R. Nair, and V. De, "Effectiveness of Adaptive Supply Voltage and Body Bias for Reducing Impact of Parameter Variations in Low Power and High Performance Microprocessors," in *Symposium on VLSI Circuits*, 2002, pp. 310–311.
[6] D. Blaauw et al., "Razor II: In Situ Error Detection and Correction for PVT and SER Tolerance," in *IEEE International Solid-State Circuits Conference (ISSCC)*, 2011, pp. 400–402.
[7] T. Sato and Y. Kunitake, "A Simple Flip-Flop Circuit for Typical-Case Designs for DFM," in *International Symposium on Quality Electronic Design*, 2007, pp. 539–544.
[8] B. P. Das and H. Onodera, "Warning Prediction Sequential for Transient Error Prevention," in *International Symposium on Defect and Fault Tolerance in VLSI Systems*, 2010, pp. 382–390.
[9] J. Tschanz et al., "Adaptive Body Bias for Reducing Impacts of Die-to-Die and Within-Die Parameter Variations on Microprocessor Frequency and Leakage," *IEEE Journal of Solid-State Circuits*, vol. 37, no. 11, pp. 1396–1402, Nov. 2002.
[10] G. Ono and M. Miyazaki, "Threshold-Voltage Balance for Minimum Supply Operation," *IEEE Journal of Solid-State Circuits*, vol. 38, no. 5, pp. 830–833, May 2003.
[11] I. Mahfuzul, N. Kamae, T. Ishihara, and H. Onodera, "A Built-in Self-adjustment Scheme with Adaptive Body Bias using P/N-sensitive Digital Monitor Circuits," in *IEEE Asian Solid State Circuits Conference*, 2012, pp. 101–104.
[12] S. M. Martin, T. Mudge, K. Flautner, and D. Blaauw, "Combined Dynamic Voltage Scaling and Adaptive Body Biasing for Lower Power Microprocessors under Dynamic Workloads," in *IEEE International Conference on Computer-Aided Design*, vol. 0, no. 2, 2002, pp. 721–725.
[13] A. Drake et al., "A Distributed Critical-path Timing Monitor for a 65nm High-Performance Microprocessor," in *IEEE International Solid-State Circuits Conference*, 2007, pp. 398–399.
[14] J. Park and J. a. Abraham, "A Fast, Accurate and Simple Critical Path Monitor for Improving Energy-Delay Product in DVS Systems," in *IEEE/ACM International Symposium on Low Power Electronics and Design*, Aug. 2011, pp. 391–396.
[15] S. Mukhopadhyay and K. Kim, "Statistical Characterization and On-chip Measurement Methods for Local Random Variability of a Process Using Sense-Amplifier-Based Test Structure," in *IEEE International Solid-State Circuits Conference*, 2007, pp. 20–22.
[16] M. Bhushan, A. Gattiker, M. B. Ketchen, and K. K. Das, "Ring Oscillators for CMOS Process Tuning and Variability Control," *IEEE Transactions on Semiconductor Manufacturing*, vol. 19, no. 1, pp. 10–18, 2006.
[17] M. Bhushan, M. B. Ketchen, S. Polonsky, and A. Gattiker, "Ring Oscillator Based Technique for Measuring Variability Statistics," in *IEEE International Conference on Microelectronic Test Structures*, Mar. 2006, pp. 87–92.
[18] A. Islam, A. Tsuchiya, K. Kobayashi, and H. Onodera, "Variation-sensitive Monitor Circuits for Estimation of Global Process Parameter Variation," *IEEE Transactions on Semiconductor Manufacturing*, vol. 25, no. 4, pp. 571–580, 2012.
[19] A. M. Islam and H. Onodera, "On-Chip Detection of Process Shift and Process Spread for Post-Silicon Diagnosis and Model-Hardware Correlation," *IEICE Transactions on Information and Systems*, vol. E96-D, no. 9, pp. 1971–1979, 2013.
[20] S. Fujimoto, A. K. M. M. Islam, T. Matsumoto, and H. Onodera, "Inhomogeneous Ring Oscillator for Within-Die Variability and RTN Characterization," *IEEE Transactions on Semiconductor Manufacturing*, vol. 26, no. 3, pp. 296–305, 2013.
[21] B. Zhou and A. Khouas, "Measurement of Delay Mismatch Due to Process Variations by Means of Modified Ring Oscillators," in *IEEE International Symposium on Circuits and Systems*, 2005, pp. 5246–5249.
[22] B. P. Das, B. Amrutur, H. S. Jamadagni, N. V. Arvind, V. Visvanathan, and B. Prasad Das, "Within-Die Gate Delay Variability Measurement Using Reconfigurable Ring Oscillator," *IEEE Transactions on Semiconductor Manufacturing*, vol. 22, no. 2, pp. 256–267, May 2009.
[23] Y. Higuchi, K.-i. Shinkai, M. Hashimoto, R. Rao, and S. Nassif, "Extracting Device-Parameter Variations using a Single Sensitivity-Configurable Ring Oscillator," in *IEEE European Test Symposium*, no. 1, 2013, pp. 106–111.
[24] A. M. Islam, T. Ishihara, and H. Onodera, "Reconfigurable Delay Cell for Area-efficient Implementation of On-chip MOSFET Monitor Schemes," in *IEEE Asian Solid State Circuits Conference*, 2013, pp. 125–128.
[25] T. Tsunomura, A. Nishida, and T. Hiramoto, "Effect of Channel Dopant Profile on Difference in Threshold Voltage Variability Between NFETs and PFETs," *IEEE Transactions on Electron Devices*, vol. 58, no. 2, pp. 364–369, Feb. 2011.
[26] H. Mostafa, M. Anis, and M. Elmasry, "NBTI and Process Variations Compensation Circuits Using Adaptive Body Bias," *IEEE Transactions on Semiconductor Manufacturing*, vol. 25, no. 3, pp. 460–467, 2012.
[27] N. Kamae, A. Tsuchiya, and H. Onodera, "A Body Bias Generator Compatible with Cell-based Design Flow for Within-die Variability Compensation," in *IEEE Asian Solid State Circuits Conference*, 2012, pp. 389–392.