

Variation-Sensitive Monitor Circuits for Estimation of Global Process Parameter Variation

Islam A. K. M. Mahfuzul, *Student Member, IEEE*, Akira Tsuchiya, *Member, IEEE*, Kazutoshi Kobayashi, *Member, IEEE*, and Hidetoshi Onodera, *Member, IEEE*

Abstract—This paper proposes a set of monitor circuits to estimate global process variations in post-silicon. Ring oscillators (ROs) are chosen as monitor circuits where ROs are designed to have enhanced sensitivities to process variations. The proposed technique extracts process parameter variations from RO outputs. An iterative estimation method is also developed to estimate variations correctly under the presence of nonlinearity in RO outputs to process variations. Simulation results show that the proposed circuits are robust against uncertainties such as measurement error. A test chip in a 65-nm process has been fabricated to validate the circuits. Process parameter variations are successfully estimated and verified by applying body bias to the chip. The proposed technique can be used for post-silicon compensation techniques and model-to-hardware correlation.

Index Terms—Monitor circuit, MOSFET, process variation.

I. INTRODUCTION

VARIATION in transistor performance has become a major problem in deep submicron CMOS circuits. In the 65 nm process and beyond, this variability plays a major role in chip performance. In order to improve yields, worst case design methodology is being followed that results in suboptimal chip performance [1]. Several post-silicon techniques have been proposed to compensate process variation [2], [3]. In order to apply post-silicon techniques effectively, monitoring of process variation is needed.

Process variation can be divided into die-to-die (D2D) and within-die (WID) variations. As the technology scaling continues, WID variation is becoming more significant [4]. WID variation has two components: random and systematic. For the random component, its effect gets reduced when the number of stages is large. In the case of large chips, the location-correlated systematic component can be as significant as D2D variation as reported in [5]. WID systematic and D2D variations affect the performances of all transistors in a chip in

the same direction, and therefore these variations play a major role in determining chip performance. As D2D and systematic variations are global for a particular die or an area inside a die, it is possible to detect and compensate these global variations on the runtime by means of body bias and supply voltage. Majority of the global variation in MOSFET performance is contributed by MOSFET gate length and threshold voltage variation [6]. For fine tuning of chip performance using post-silicon techniques, on-chip monitoring of global parameter variations is needed.

Various types of monitor circuits have been proposed so far to monitor process variation. Circuit delay and leakage current are the most common targets for monitoring [7]. However, leakage current monitoring needs analog circuits, thus measurement and calibration of analog signals increase design complexity. Digital circuits with small areas are preferable so that they can be embedded anywhere in the chip. On the other hand, delay of a circuit does not provide information on individual MOSFET variation. Monitoring of individual process parameter variation will provide a broader scope for fine-tuning of chip performance. However, monitoring of single process parameter variation requires device arrays or complex circuits that are not suitable for on-chip implementation.

Some approaches are proposed to extract process parameter variations from digital circuits, such as ring oscillators (ROs) [8]–[13]. In [8], the slew rate of the inverter output is used to monitor rise time and fall time variations separately. In [9], the theory of pulse shrinking across a buffer ring is used to monitor rise time and fall time of the inverter cell. Although rise time and fall time variations will give us information on pMOSFET and nMOSFET on-currents, direct monitoring of key process parameters is preferable for fine tuning. In [10], simple inverter structures with different PN ratios are used to extract variations in pMOSFET and nMOSFET on-currents. This paper introduces monitor circuits by which variations in process parameters can be estimated.

In [11], it is shown that extraction of different process parameters is possible with modified inverter structures and proper data processing. Extraction of threshold voltage variation from different path delays is proposed in [12]. In this approach, sensitivities of the monitor circuits are used to extract threshold voltage variations. However, the effect of nonlinearity of monitor circuit outputs on variations is not considered here, which affects the accuracy of estimation. Furthermore, gate length variation is ignored, which is critical

Manuscript received October 29, 2011; revised March 3, 2012 and April 11, 2012; accepted April 16, 2012. Date of publication May 10, 2012; date of current version October 25, 2012. This work was supported by the VLSI Design and Education Center, University of Tokyo.

I. A. K. M. Mahfuzul, A. Tsuchiya, and H. Onodera are with the Department of Communications and Computer Engineering, Kyoto University, Kyoto 606-8501, Japan (e-mail: mahfuz@vlsi.kuee.kyoto-u.ac.jp; tsuchiya@vlsi.kuee.kyoto-u.ac.jp; onodera@vlsi.kuee.kyoto-u.ac.jp).

K. Kobayashi is with the Graduate School of Science and Technology, Kyoto Institute of Technology, Kyoto 606-8585, Japan (e-mail: kazutoshi.kobayashi@kit.ac.jp).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TSM.2012.2198677

for global variation. In [13], a set of ROs consisting of simple inverter structures is proposed as monitor circuits to estimate global variations of threshold voltage and gate length. An iterative estimation technique is used to cope with the error occurring from nonlinearity in circuit outputs.

This paper is an extension over [13]. As the process parameter variations in real chips are unknown, it is extremely difficult to show the validity of the monitor circuits. Body bias has been applied to the chip and process variations are estimated to show the validity of the circuits. High correlation has been found in the estimation results corresponding to the body bias values; thus, the validity is confirmed. Next, the local effect of random variation needs to be canceled out for monitoring of global variations. This can be done by increasing the number of stages, but increasing the number has area overhead that is undesirable. This paper presents a methodology to calculate the adequate number of stages for a given tolerable range of errors in the estimation.

Key enhancements of this paper over [13] are summarized as follows.

- 1) A methodology is presented to calculate the number of stages for the monitor circuits.
- 2) Validity of the monitor circuits is confirmed by applying different body bias values to the chip.

The key contribution of this paper is that it establishes a systematic technique to estimate threshold voltages and gate length variations from on-chip monitor circuit outputs.

The remainder of this paper is organized as follows. In Section II, a variation model and an iterative estimation technique based on the model are described. In Section III, design techniques to realize variation-sensitive ROs are demonstrated. A methodology to choose the most suitable set of monitor circuits and a methodology to calculate the number of stages are described here. In Section IV, experimental results on the effectiveness of the iterative estimation technique and the robustness of the monitor circuits are demonstrated. In Section V, the test chip structure for a 65 nm process is described. Estimation results and their validation are also discussed here. Finally, Section VI concludes this paper.

II. PROPOSED ESTIMATION TECHNIQUE OF PROCESS PARAMETER VARIATIONS

In this section, first the basic concept of the total parameter estimation process is described. Then, a set of parameters is chosen for model global variation. Finally, the estimation procedure is described. Monitor circuits suitable for this technique will be discussed in Section III.

A. Basic Concept

The transistor model plays an important role in very large scale integration design. Circuit designers see the process through the transistor model provided by the foundry. Expressing the total variation in terms of the key model parameters will help the designers to tune their design. The sensitivities of a circuit output to process parameter variations can be calculated by circuit simulation. These sensitivities give us useful information about the silicon.

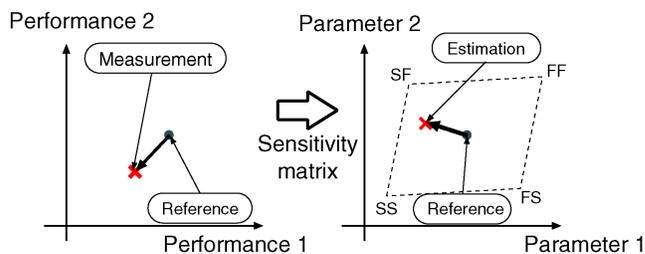


Fig. 1. Extraction of process parameters from variation-sensitive monitor circuits. Sensitivity matrix relates variations in circuit performances to variations in process parameters. Monitor circuits having different sensitivities to different process parameters are needed.

If we have several circuits that have different sensitivities to different process parameters, it is possible to extract the amount of variation for each process parameter using the circuit outputs and sensitivity coefficients. This concept is illustrated in Fig. 1. For simplicity, Fig. 1 shows an example of estimation of two parameters from two circuit performances. The transistor model provided to the circuit designer can be used as an interface between the circuit performances and the process parameters in silicon. Deviations of process parameter values from those defined in the transistor model can be estimated from circuit performances. A suitable set of circuits is needed for the estimation.

The values of the process parameters defined in the model are considered to be the reference point. In design, circuit performances are predicted using this transistor model. The idea is to compare the measured performances with the predicted values and estimate the amount of deviation for each process parameter so that predictions get closer to the measured values. So, the differences between measurements and predictions are observable here, which are shown in the left graph of Fig. 1. Because of D2D variation, the measurement point will vary from chip to chip. The circuits should be designed such that different measurement points give different estimations of process parameters. Design of the sensitivity matrix is most important here to establish an accurate and robust estimation framework.

B. Variation Model

In this estimation technique, a linear model is used to express the relationship between the circuit outputs and process parameter variations. A set of process parameters needs to be defined to express the global variation effect first. Equation (1) shows the α -power law model of transistor on-current [14]

$$I_{on} = \beta \cdot (V_{DD} - V_{th})^\alpha. \quad (1)$$

Here, β is the current factor and equals $\frac{\mu C_{ox} W}{L}$, where μ denotes the effective mobility, C_{ox} is the gate-to-channel capacitance per unit area, W is the channel width, and L is the channel length. When variation exists, (1) can be written as (2) where β_0 and V_{th0} are the values defined by the transistor model

$$I_{on0} + \Delta I_{on} = (\beta_0 + \Delta\beta) \cdot (V_{DD} - (V_{th0} + \Delta V_{th}))^\alpha. \quad (2)$$

I_{on0} is the default on-current calculated from the transistor model and ΔI_{on} is the variation in chip. Values of $\Delta\beta$ and ΔV_{th} are variations that differ from chip to chip. From (2), at least two parameters are needed to model on-current variation of a single type of MOSFET, and four parameters to model variations for pMOSFET and nMOSFET separately. For simplification of the model, L is chosen to be a common parameter to model variations in current factors of both MOSFETs because L variation has large contribution to D2D variation. Furthermore, L variation is common for both MOSFETs in standard cells.

In this paper, we therefore focus on the estimation of total global variation in three key parameters of pMOSFET, threshold voltage (V_{THP}), nMOSFET threshold voltage (V_{THN}), and gate length (L). Suppose ΔV_{THP} , ΔV_{THN} , and ΔL are global variations of those parameters to be estimated and Δf is the corresponding frequency shift that we can measure. If ΔV_{THP} , ΔV_{THN} , and ΔL are small, then those variations can be related in a linear equation as shown in (3), where k_P , k_N , and k_L are sensitivity coefficients

$$\begin{aligned}\Delta f &= f_M - f_{Ref} \\ &= k_P \Delta V_{THP} + k_N \Delta V_{THN} + k_L \Delta L.\end{aligned}\quad (3)$$

Here, f_M is the measured frequency and f_{Ref} is the reference frequency. The value f_{Ref} can be obtained by circuit simulation using the RC extracted netlist from layout. The difference Δf here represents the deviation of frequency in measurement from circuit simulation. Sensitivity coefficients can be calculated by circuit simulation.

C. Estimation Procedure

In (3), there are three unknown parameters. So, at least three equations are needed to extract these three unknown values. The three equations can be derived from three monitor circuits whose sensitivity vectors form a nonsingular matrix. The amount of variation of each process parameter will be estimated by solving

$$\vec{V} = \mathbf{S}^{-1} \vec{F}\quad (4)$$

where

$$\vec{V} = \begin{pmatrix} \Delta V_{THP} \\ \Delta V_{THN} \\ \Delta L \end{pmatrix}, \mathbf{S} = \begin{pmatrix} k_{P1} & k_{N1} & k_{L1} \\ k_{P2} & k_{N2} & k_{L2} \\ k_{P3} & k_{N3} & k_{L3} \end{pmatrix}, \vec{F} = \begin{pmatrix} \Delta f_1 \\ \Delta f_2 \\ \Delta f_3 \end{pmatrix}.$$

Here, vector \vec{V} is the vector for ΔV_{THP} , ΔV_{THN} , and ΔL . Matrix \mathbf{S} is the sensitivity matrix and vector \vec{F} is the vector for the frequency shift from the reference value. Vectors (k_{P1}, k_{N1}, k_{L1}) , (k_{P2}, k_{N2}, k_{L2}) , and (k_{P3}, k_{N3}, k_{L3}) are sensitivity coefficient vectors for three circuits.

Estimation based on the linear model in (3) has two potential problems. First, nonlinearity in circuit output will affect the estimation accuracy. Second, process variations affect the sensitivity values; thus, sensitivity coefficients calculated using the provided transistor model will not reflect the actual coefficient in the chip. In order to overcome these two problems, this paper proposes an iterative estimation technique where

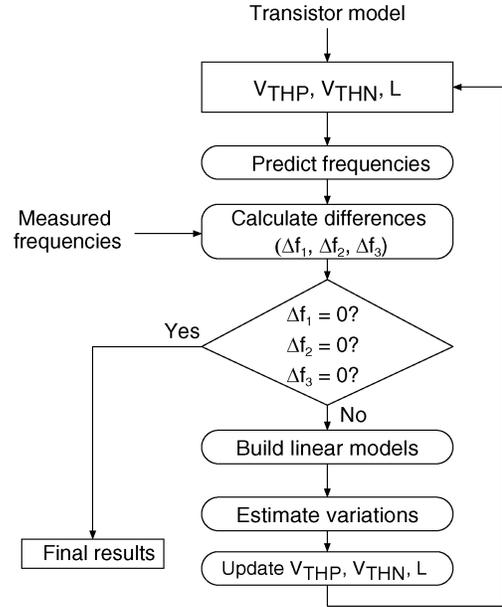


Fig. 2. Proposed iterative estimation procedure of process parameters.

sensitivity coefficients are updated at each iteration; thus, correlation between the model and hardware can be achieved and the nonlinearity problem can be overcome.

Fig. 2 shows the proposed iterative estimation procedure. First, frequencies of the monitor circuits are predicted by circuit simulation using a transistor model. Measured values are obtained from the chip and then compared with the predicted values. Zero difference refers that no variation from the values in process parameters defined in the model exists in the chip. If the difference is not zero, then some variations exist in the chip. Linear models of (3) are built by calculating the sensitivity coefficients. Variations of the target parameters are estimated by solving (4). Process parameter values are updated in the transistor model with the estimated amounts of variations and new predictions are made for the frequencies. If the differences between measurements and predictions are not zero, new linear models are built and variations are estimated again. Thus, a new set of parameter values will be obtained after each iteration. This way, the whole process is iterated until the differences between measurements and predictions are zero.

Selection methodology of monitor circuits suitable for this technique will be discussed in Section III.

III. SET OF MONITOR CIRCUITS FOR ESTIMATION OF PROCESS VARIATION

A suitable set of monitor circuits is needed to realize the proposed estimation technique described in Section II. The monitor circuits should have different sensitivities to the process parameters. In this section, some design options to realize variation-sensitive monitor circuits from simple inverter cell structures will be demonstrated. Sensitivities are calculated by circuit simulation. Commercial 65 nm process technology is assumed in our simulation. Based on the simulation results, a

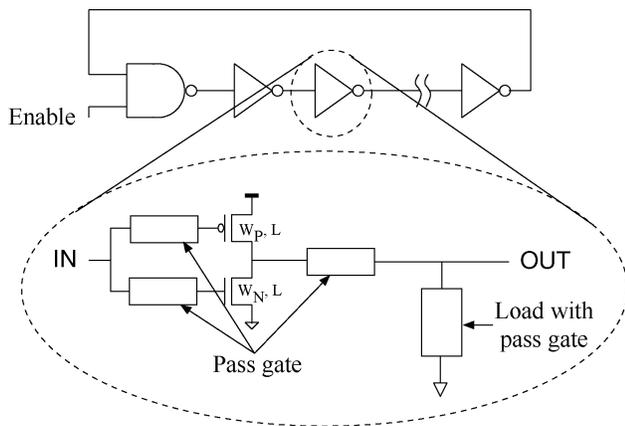


Fig. 3. RO as a monitor circuit. The inverter cell structure can be modified to get enhanced sensitivities.

methodology to choose the best suitable monitor circuits will be described.

A. Design Methodology of Monitor Circuits

For the estimation technique proposed in Section II, monitor circuits with the following characteristics are needed:

- 1) regularity and low design complexity;
- 2) high sensitivity;
- 3) digital in nature;
- 4) small area.

Regularity of poly pitch should be maintained in the monitor circuit as it affects gate length variation. Design complexity should be low so that monitor circuits can be ported to different process technologies. Next, monitor circuits should have high sensitivities to process parameter variations. Digital nature of the monitor circuits is important for on-chip measurement and processing. Finally, area of the monitor circuits should be small enough so that implementing them does not cause large area overhead.

An RO is a good candidate to be used as a monitor circuit. A simple RO fulfils all the requirements mentioned above except the high sensitivity. In this paper, we therefore have modified the inverter structure in the RO to get enhanced sensitivities. The following techniques are used to modify the sensitivities of an RO frequency to process parameters:

- 1) change gate width;
- 2) use pass-gates;
- 3) use gate capacitance and pass-gate in series.

Fig. 3 shows our proposed monitor circuit where the inverter structure can be modified to get enhanced sensitivities. Effects on the sensitivities are described below.

B. Simulation Results of Sensitivity

1) *RO With Parallel MOS*: Increasing the gate width of pMOSFET in the inverter structure will make the RO frequency more sensitive to nMOSFET parameters. We can increase gate width of pMOSFET or we can place multiple pMOSFETs in parallel. In order to maintain regularity, we have designed inverters with parallel MOSFETs. Fig. 4 shows an inverter where pMOSFET is four times larger than that

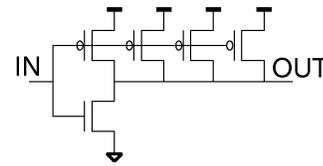


Fig. 4. Inverter cell with parallel pMOSFETs ("PRICH").

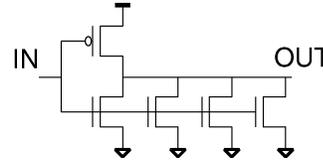


Fig. 5. Inverter cell with parallel nMOSFETs ("NRICH").

of the standard cell. Similarly, the inverter structure shown in Fig. 5 will be more sensitive to pMOSFET parameters. We call these cells as "PRICH" and "NRICH," respectively, whereas the standard inverter cell is called as "STD." From simulation results for a "PRICH" RO, 21% increase in V_{THN} sensitivity and 20% decrease in V_{THP} sensitivity is calculated compared to that of the "STD" RO.

2) *RO With Pass-Gate*: RO with single pass-gate becomes highly sensitive to threshold voltage variation. The operation of RO with pass-gate is demonstrated in [11]. Figs. 6 and 7 show inverters with a pMOSFET pass-gate and an nMOSFET pass-gate at the output. We call these inverter cells as "PPASS_O" and "NPASS_O," respectively. For a "PPASS_O" RO, ΔV_{THP} sensitivity increases by five times than that of "STD" RO. For "NPASS_O" RO, ΔV_{THN} sensitivity increases by seven times than that of "STD" RO.

Same gate sizes are used for pass-gates in the simulations as those in the standard inverter cell MOSFETs. Next, the effects of gate width of pass-gates on the sensitivities are studied. For "NPASS_O" RO, decreasing pass-gate size to half increases the sensitivity to ΔV_{THN} by 8%, which is very small compared to the 500% increase in the sensitivity against the "STD" RO. Considering design and layout complexity, pass-gates with same sizes of MOSFETs as in the standard inverter cell are preferable.

For "PPASS_O" and "NPASS_O" inverters, voltage drop occurs across the pass-gates. For example, in the case of "NPASS_O" inverter, output voltage of the inverter does not rise to a high level during the loading of the next inverter. This voltage drop turns the pMOSFET of the next inverter partially on. In order to avoid this, inverter structures shown in Figs. 8 and 9 are proposed. We call these cells as "PPASS_I" and "NPASS_I," respectively. For "NPASS_I" inverter, nMOSFET pass-gate contributes to the fall time only, thus the pMOSFET of the next stage is turned off fully. Input and output voltages have full swing during the oscillation similar to the behaviors of the standard cells.

3) *RO With Extra Load*: Figs. 10 and 11 are ROs with an extra load in the output. These cells will be called as "PLOAD" and "NLOAD," respectively, where "PLOAD" cell's load is controlled by a pMOSFET pass-gate and "NLOAD" cell's load is controlled by an nMOSFET pass-gate. The loads are

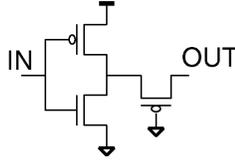


Fig. 6. Inverter cell with pMOSFET pass-gate at output (“PPASS_O”).

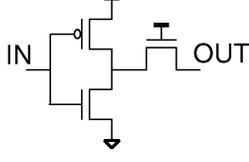


Fig. 7. Inverter cell with nMOSFET pass-gate at output (“NPASS_O”).

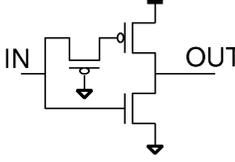


Fig. 8. Inverter cell with pMOSFET pass-gate at input of pMOSFET gate (“PPASS_I”).

realized by MOSFET gate capacitance. For Fig. 10, when V_{THP} increases, resistance for the pMOS pass-gate increases. As a result, the inverter sees smaller load and hence delay decreases. Thus, the effect of V_{THP} variation gets reduced. Sizing of the load determines the sensitivity for this structure. For “PLOAD” RO where the extra load is equivalent to FO4 of the “STD” cell, sensitivity to V_{THP} decreases by 45% than that of an “STD” cell RO.

Table I summarizes sensitivity coefficients for these ROs. Sensitivity coefficients are calculated by $k_P = \frac{\Delta f/f_0}{\Delta V_{THP}/\Delta V_{THP0}}$, $k_N = \frac{\Delta f/f_0}{\Delta V_{THN}/\Delta V_{THN0}}$, and $k_L = \frac{\Delta f/f_0}{\Delta L/\Delta L_0}$.

C. Set of Monitor Circuits for Process Parameter Estimation

A set of monitor circuits is needed to extract ΔV_{THP} , ΔV_{THN} , and ΔL . The question is how to choose the most suitable set of ROs. The sensitivity matrix plays a major role in defining the robustness of estimation. Angles between the sensitivity vectors are good indicators on how good the sensitivity matrix is for accurate estimation. Fig. 12 shows the sensitivity vectors for ROs with “STD,” “PPASS_O,” “NPASS_O,” “PRICH,” and “NRICH” inverter cells. In Fig. 12, “PPASS_O” and “NPASS_O” ROs have large angle between their sensitivity vectors compared to that of “PRICH” and “NRICH” ROs because of their high sensitivities. A quantitative evaluation can be performed by calculating the condition number of the selected ROs.

Condition number is a good indicator on how robust estimation result will be against the uncertainties in sensitivity coefficients or in measurement values. If a matrix has a small condition number, the matrix is called a well-conditioned matrix. The condition number of a matrix can be calculated using the infinite norm of the matrix as follows:

$$\text{cond}(\mathbf{A}) = \|\mathbf{A}\|_{\infty} \cdot \|\mathbf{A}\|_{\infty}^{-1}. \quad (5)$$

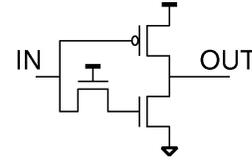


Fig. 9. Inverter cell with nMOSFET pass-gate at input of nMOSFET gate (“NPASS_I”).

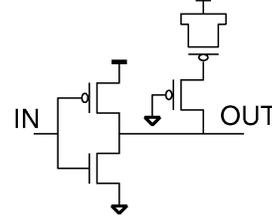


Fig. 10. Inverter cell with extra load and pMOSFET pass-gate. Time for charging and discharging of the extra load depends on pMOSFET pass-gate threshold voltage (“PLOAD”).

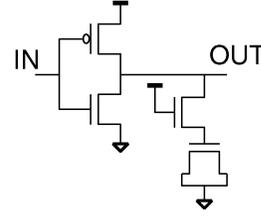


Fig. 11. Inverter cell with extra load and nMOSFET pass-gate. Time for charging and discharging of the extra load depends on nMOSFET pass-gate threshold voltage (“NLOAD”).

Condition number of 1 means that the sensitivity vectors are orthogonal to each other. The bigger the condition number, the smaller the angles between the sensitivity vectors. From the design options presented in Section II, the set of ROs having the smallest condition number is most suitable for this estimation technique. Table II shows condition numbers of the sensitivity matrices for different RO sets. In Table II, “STD,” “PPASS_I,” and “NPASS_I” ROs have the smallest condition. Thus, “PPASS_I,” “NPASS_I,” and “STD” ROs are most suitable for the proposed estimation technique.

D. Number of Stages for Monitor Circuits

For monitoring of global variations of process parameters, the effect of random variation needs to be canceled out. Increasing the number of stages for ROs will not only average out the random effect but also consume a large area. Thus, a tradeoff has to be made between estimation accuracy and monitor circuit area. Equation (6) shows the probability distribution function of Δf , which is the difference between measurement and prediction

$$\phi(\Delta f) = a \exp \frac{\mu_{\Delta f} - \Delta f}{2\sigma_{\Delta f}^2}. \quad (6)$$

Here, $\mu_{\Delta f}$ is the mean value and $\sigma_{\Delta f}$ is the standard deviation of Δf . In this paper, we are concerned on the mean value $\mu_{\Delta f}$. Because of random variations, the monitored value may

TABLE I
SENSITIVITY COEFFICIENTS OF ROs

| RO Type | k_P | k_N | k_L |
|---------|--------|--------|--------|
| STD | -0.038 | -0.035 | -0.026 |
| PPASS_I | -0.18 | -0.033 | -0.063 |
| NPASS_I | -0.028 | -0.2 | -0.034 |
| PPASS_O | -0.24 | 0.052 | -0.085 |
| NPASS_O | 0.054 | -0.34 | -0.029 |
| CPASS | -0.039 | -0.036 | -0.026 |
| PRICH | -0.031 | -0.041 | -0.026 |
| NRICH | -0.046 | -0.034 | -0.027 |
| PLOAD | -0.020 | -0.048 | -0.023 |
| NLOAD | -0.044 | -0.022 | -0.027 |

TABLE II
CONDITION NUMBERS OF SENSITIVITY MATRICES
FOR DIFFERENT SETS OF ROs

| No. | RO Set | | | Condition Number |
|-----|--------|---------|---------|------------------|
| | RO #1 | RO #2 | RO #3 | |
| 1 | STD | PPASS_O | NPASS_O | 39 |
| 2 | CPASS | PPASS_O | NPASS_O | 50 |
| 3 | STD | PPASS_I | NPASS_I | 26 |
| 4 | STD | PLOAD | NLOAD | 34 |
| 5 | STD | PRICH | NRICH | 78 |
| 6 | PRICH | PPASS_I | NPASS_I | 28 |

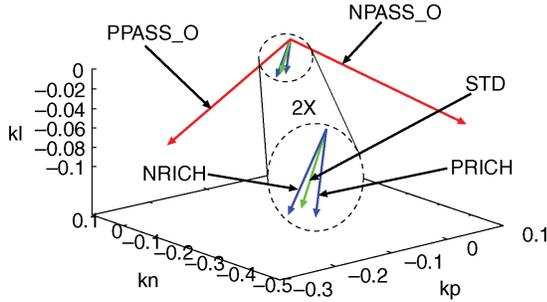


Fig. 12. Sensitivity vectors of various types of ROs. Sensitivity vectors of “PPASS” and “NPASS” ROs forms are near orthogonal referring their robustness in estimation.

deviate from the mean value. The monitored frequency will fall within the range of $\mu_{\Delta f} \pm 3\sigma_{\Delta f}$ with 99.9% probability.

From (4), the estimation value v_i of a particular parameter can be expressed by

$$v_i = z_{i1}\Delta f_1 + z_{i2}\Delta f_2 + z_{i3}\Delta f_3. \quad (7)$$

Here, parameter v_i is the variation to be estimated and parameter z_{ij} is the element of the matrix \mathbf{S}^{-1} of (4). Index i refers to the row number of the matrix. In (6), Δf_1 , Δf_2 , and Δf_3 follow the probability distribution function of (6). Using the method of moment, distribution σ_{v_i} in the estimated value can be calculated as follows [15]:

$$\sigma_{v_i}^2 = \sum_j^3 (z_{ij}\sigma_{\Delta f_i})^2. \quad (8)$$

Equation (8) gives us the tradeoff relationship between estimation accuracy and the number of stages. By calculating

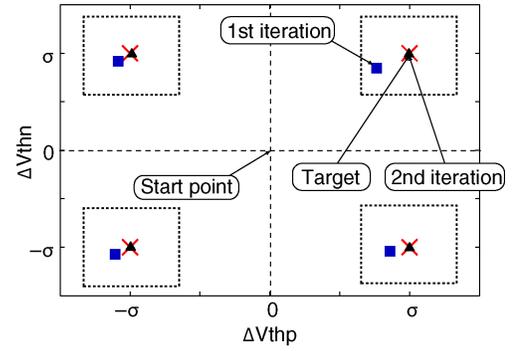


Fig. 13. Effect of iteration on estimation. Estimation results converge to the target point after several iterations.

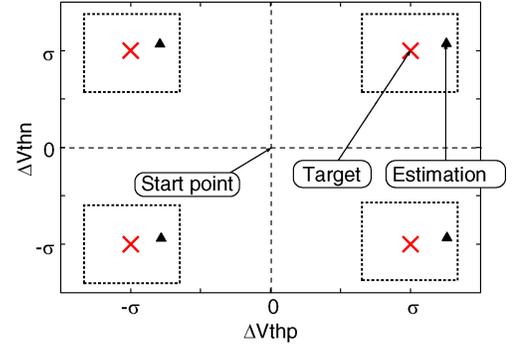


Fig. 14. Effect of uncertainty such as measurement error in frequency on estimation. Despite +1% error in each frequency estimation results converge near the target point.

the distributions of each RO frequency, a sufficient number of stages can be calculated for a value for σ_{v_i} .

IV. SIMULATION RESULTS

We propose “STD,” “PPASS_I,” and “NPASS_I” ROs as monitor circuits for process parameter estimation. Simulation-based experiments have been performed to verify the validity and the robustness of the proposed monitor circuits. A real chip scenario is emulated in our simulation.

A. Simulation Setup

In the experiments, the real chip scenario is emulated in the following way. First, we take a transistor model to predict the circuit performances. The values of V_{THP} , V_{THN} , and L defined in the model are our reference point or start point. RO frequencies will be predicted from this point. Next, we apply some known amounts of ΔV_{THP} , ΔV_{THN} , and ΔL to the transistor model. We call this model as “chip” model because simulation results using this model will be considered as measurement results obtained from the chip. Then, RO frequencies are simulated using the “chip” model. These frequencies are considered to be the values we can obtain from the chip. Finally, ΔV_{THP} , ΔV_{THN} , and ΔL are estimated using the estimation technique described in Section II. ΔV_{THP} , ΔV_{THN} , and ΔL are the amount of deviations from the start point. If the estimated ΔV_{THP} , ΔV_{THN} , and ΔL match with those in the “chip” model, the estimation becomes correct. In

order to emulate measurement errors, some amounts of errors are added to the simulation results obtained from the “chip” model.

B. Validity of the Iterative Estimation Technique

The proposed technique has been verified for different values of ΔV_{THP} , ΔV_{THN} , and ΔL in the “chip” model. For example, Fig. 13 shows the estimation results of ΔV_{THP} and ΔV_{THN} when these values are set to be $\pm\sigma$ in the “chip” model. The σ value for ΔL is set in these cases. In Fig. 13, the x -axis and y -axis refer to ΔV_{THP} and ΔV_{THN} variations, respectively. Cross points refer to the applied ΔV_{THP} and ΔV_{THN} values in the “chip” model. Closed rectangular points refer to the estimation results obtained after the first iteration and closed triangular points refer to the results obtained after the second iteration. Regions enclosed by dotted rectangles are used for separating the corresponding estimation results from each other. After the first iteration, the estimated values of ΔV_{THP} and ΔV_{THN} locate near to the target values but with some amounts of errors. These errors occur from the nonlinearity. However, after the second iteration, the estimated values move closer to the target points. Thus, the iterative technique converges and accurate estimations are obtained.

C. Robustness of the Monitor Circuits

The robustness of the monitor circuits has been verified by applying different error patterns in the frequencies. For example, Fig. 14 shows the estimation results when +1% error exists in each of the measured frequencies. Simulation setup and the meanings of the symbols are the same as in Fig. 13. Closed triangular points are estimation results after the iteration technique has converged. Although some errors are there in the estimations, the important point to note here is that in spite of +1% error in each frequency, estimation results have been converged near the target values. This proves that the proposed circuits are robust for process parameter estimation.

V. ESTIMATION RESULTS FROM TEST CHIP

A test chip has been fabricated in a 65-nm process to confirm the validity of our proposed monitor circuits. In this section, the test structure to evaluate the monitor circuits is described. Measurement results and estimation results are discussed next.

A. Chip Design

A test chip in a 65-nm process technology has been fabricated. The process features one poly layer, 12 metal layers, copper wiring, and low-K insulating material techniques. The physical gate oxide thickness is 1.7 nm. ROs of Table I are implemented in the test chip. In order to evaluate the validity of the monitor circuits, the effect of random variation needs to be evaluated as well. Therefore, an array-based test structure methodology proposed in [16] is used to get both D2D and WID variations. Fig. 15 shows the chip micrograph where 270 sections are integrated into a 15×18 array on chip. Each section contains an instance of a particular type of RO.

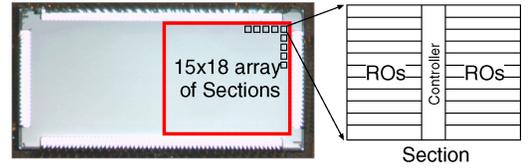


Fig. 15. Test chip in 65-nm process.

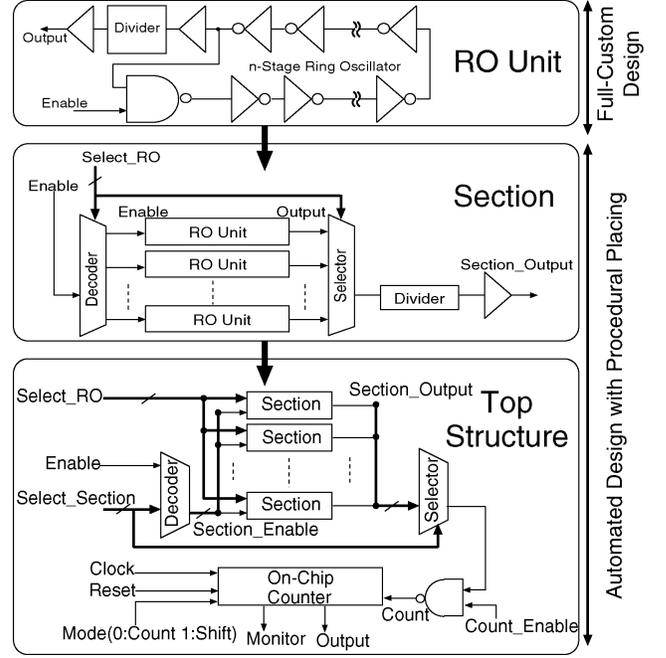


Fig. 16. Block diagram of test structure.

Therefore, 270 ROs of the same type are integrated in a single die. Fig. 16 shows the block diagram of our test structure. Selectors and decoders are used to select an RO to oscillate and capture the waveform outside the chip. Local divider and on-chip counter are used to reduce the frequency below 1 MHz so that the waveform does not get distorted outside the chip. Enable signals are generated locally inside the chip to avoid harmonic oscillation [17]. The number of stages for each RO is chosen to be a prime number 19 to minimize the probability of harmonic oscillation. We get 270 frequency measurements for a single RO; thus, WID variation can be obtained. From this WID variation, we can calculate the number of stages required for a tolerable range of error in the estimation. Global variation is obtained by averaging the 270 measured frequencies. We have 30 chips. So, 30 global values of frequencies are obtained for each RO.

B. Measurement Procedure

The overall procedure for RO frequency measurement is as follows. First, an RO instance is enabled using the selectors. Then, the total time for a fixed number of oscillations is measured with a resolution of 12.5 ns using an 80 MHz clock signal. The number of oscillations is set to 1024 in our procedure. As the frequency outside the chip is around 1 MHz, the maximum error for this procedure is $\pm 1/(1024 * 80) = \pm 0.001\%$. Then, the next RO instance is selected and

TABLE III
AVERAGE VALUES OF MEASURED FREQUENCIES FROM 30 CHIPS

| RO Type | Measurement Ave. [MHz] | WID Variation σ/μ [%] | Prediction [MHz] | Deviation [%] |
|---------|------------------------|--------------------------------|------------------|---------------|
| STD | 2151 | 1.42 | 1907 | 12.8 |
| PPASS_I | 774 | 2.84 | 555 | 39.5 |
| NPASS_I | 769 | 3.78 | 692 | 11.1 |
| PPASS_O | 443 | 3.73 | 244 | 81.8 |
| NPASS_O | 393 | 6.66 | 400 | -1.78 |
| CPASS | 907 | 1.2 | 819 | 10.7 |
| LOAD | 1192 | 1.47 | 1114 | 7.02 |
| NLOAD | 1146 | 1.32 | 1006 | 13.9 |
| PRICH | 1219 | 1.29 | 1141 | 6.81 |
| NRICH | 1199 | 1.2 | 1046 | 14.6 |

Maximum WID variation is shown here. Predicted values for the frequencies and deviation in measurement from the prediction are also shown.

measured. In order to check measurement precision, frequency of the same RO instance is measured 100 times. Standard deviation for the 100 frequencies is 0.022%.

C. Measurement Results

Table III shows the measured frequencies from our test chip. Frequencies shown in Table III are the average values of all frequency measurements from 30 chips. Predictions for RO frequencies using a typical-typical (TT) transistor model are also shown Table III. In this paper, the TT model is used as the reference for estimation. Predicted values are compared with the measured values. Positive value of difference refers that the measured value is higher than the predicted value. Large differences between measurements and predictions are observed for “PPASS_I” and “PPASS_O” ROs. These ROs are highly sensitive to ΔV_{THP} ; thus, ΔV_{THP} is expected to be larger. Maximum amount of WID variation among 30 chips for each RO is also shown in Table III. ROs with higher sensitivities have larger WID variations. From the WID variation, a sufficient number of stages for the monitor circuits will be calculated using (8).

D. Estimation Results

Global variations of ΔV_{THP} , ΔV_{THN} , and ΔL are estimated for 30 chips. Fig. 17 shows the estimation results of ΔV_{THP} , ΔV_{THN} , and ΔL . The TT transistor model is used as the reference, which is located at the center of the graph. Process corners defined in the transistor model are also shown Fig. 17. The estimated results of ΔV_{THP} and ΔV_{THN} are located within the corner boundary. A larger ΔV_{THP} is estimated than ΔV_{THN} . In order to verify the validity of the estimation results, ΔV_{THP} and ΔV_{THN} are compared with that provided by the PCM data. The estimation results are within the PCM data range. Fig. 18 shows the estimation results of ΔL . ΔL values span from -6.0 nm to -3.5 nm.

E. Number of Stages

The adequate number of stages for the monitor circuits are calculated for a fixed tolerable range of estimation errors due to WID variation using (8). From Table III, we get the WID variations for each RO. Using these variations, the number of

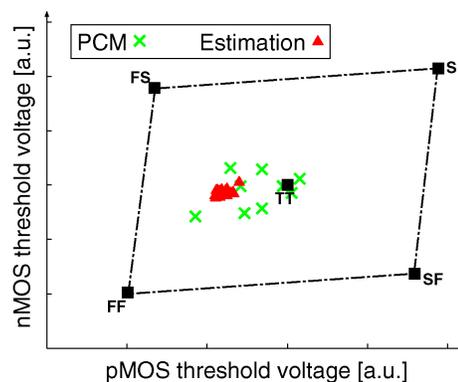


Fig. 17. Estimation results of ΔV_{THP} and ΔV_{THN} for 30 chips. Estimated results are compared with that in PCM data and process corner models.

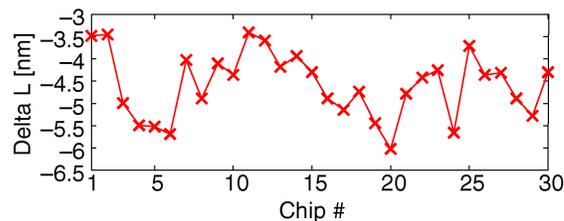


Fig. 18. Estimation results of ΔL for 30 chips.

stages are calculated to be 171 when the standard deviation for threshold voltage estimation is set to 2 mV.

F. Validation

1) *Predictability of Circuit Performance*: Predictions of circuit performances can be made using the estimation results for each chip. Close match between predictions and measurements for all circuits will confirm the validity of the estimation results. Predictions are made for our RO frequencies using estimated ΔV_{THP} , ΔV_{THN} , and ΔL for 30 chips. Table IV shows the mismatch between predictions and measurements for a particular chip. For the top three ROs in the table, no mismatch is found because these ROs are used for the estimation. The key point is whether the predictions for other circuits match closely with the measurements. In Table IV, predictions match with the measurements within maximum mismatch of 6%, which is small compared to the differences in Table III. Thus, the circuit performances can be predicted with high accuracy using the estimated values. Therefore, the proposed technique can be used for post-silicon tuning.

2) *Different Body Bias Condition*: Threshold voltages can be changed by applying body biases to the chip. So, the monitor circuits can be validated by estimating process variations in different bias conditions. If the estimated values correlate to the applied body bias values, then the monitor circuits will be proved to be valid for correct monitoring of process variations.

Fig. 19 plots the values of ΔV_{THP} and ΔV_{THN} estimated in different body bias conditions for a particular chip. In Fig. 19, the x-axis refers to ΔV_{THP} estimation and the y-axis refers to ΔV_{THN} estimation. Rectangular points are estimated values of ΔV_{THP} and ΔV_{THN} when only pMOSFET is biased. Triangular

TABLE IV

COMPARISON BETWEEN MEASUREMENTS AND PREDICTIONS FOR RO FREQUENCIES FOR A CHIP

| RO | Measurement [MHz] | Prediction [MHz] | Difference [%] |
|---------|-------------------|------------------|----------------|
| STD | 2145 | 2145 | 0.0 |
| PPASS_I | 753 | 753 | 0.0 |
| NPASS_I | 766 | 766 | 0.0 |
| PPASS_O | 421 | 395 | -6.0 |
| NPASS_O | 398 | 399 | 0.3 |
| PLOAD | 1189 | 1224 | 2.9 |
| NLOAD | 1135 | 1158 | 2.0 |
| CPASS | 901 | 915 | 1.6 |
| PRICH | 1216 | 1270 | 4.4 |
| NRICH | 1187 | 1181 | -0.5 |

Predictions are made using the estimated ΔV_{THP} , ΔV_{THN} , and ΔL .

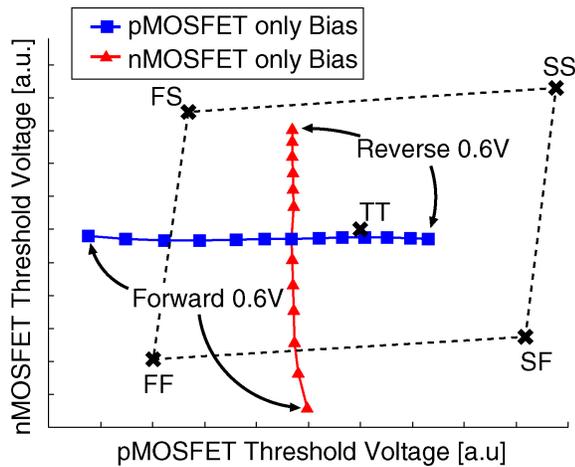


Fig. 19. Estimation of V_{THP} and V_{THN} in different bias conditions. Threshold change is detected properly with the proposed monitor circuits.

points refer to estimated values of ΔV_{THP} and ΔV_{THN} when only nMOSFET is biased. When only pMOSFET is biased, the estimated point moves in the horizontal direction referring only ΔV_{THP} is changed in the estimation. When only nMOSFET is biased, the estimated point moves in the vertical direction referring only ΔV_{THN} is changed in the estimation. Thus, it is proved that any change in the threshold voltage can be detected correctly by the proposed monitor circuits.

Standard deviation of estimated ΔL values in different body bias conditions is calculated as 0.7%, which is small. So, ΔL estimation remains the same in different bias conditions referring to the validness of the monitor circuits.

VI. CONCLUSION

In this paper, a systematic estimation technique of global variations for V_{THP} , V_{THN} , and L was proposed. A set of variation-sensitive ROs as monitor circuits suitable for the estimation technique was proposed. The proposed technique used an iterative method based on the simple linear model to extract process parameter variations from these RO outputs. Experimental results showed that our proposed circuits are robust in the presence of uncertainties. The test chip in a 65-nm process was fabricated to verify our circuits. ΔV_{THP} , ΔV_{THN} ,

and ΔL variations were successfully estimated for each chip. ΔV_{THP} and ΔV_{THN} variation ranges in our estimated result fit within the variation ranges provided by PCM data. Predictions of performances were made for various types of circuits using our estimated amount of variations. Predicted values match closely with the measured values referring to the validness of the estimation technique. The monitor circuits are also verified in different body bias conditions; thus, the validity of the monitor circuits for process parameter monitoring is confirmed. The proposed monitor circuits can be used in post-silicon compensation techniques and model-to-hardware correlation.

ACKNOWLEDGMENT

The authors are grateful to STARC, Yokohama, Japan, E-Shuttle, Inc., Kanagawa, Japan, and Fujitsu, Ltd.

REFERENCES

- [1] S. Nassif, "Process variability at the 65 nm node and beyond," in *Proc. IEEE Custom Integr. Circuits Conf.*, Sep. 2008, pp. 1–8.
- [2] 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," *IEEE J. Solid-State Circuits*, vol. 38, no. 5, pp. 826–829, May 2003.
- [3] S. Martin, K. Flautner, T. Mudge, and D. Blaauw, "Combined dynamic voltage scaling and adaptive body biasing for lower power microprocessors under dynamic workloads," in *Proc. IEEE/ACM Int. Conf. Comput.-Aided Des.*, Nov. 2002, pp. 721–725.
- [4] H. Onodera, "Variability: Modeling and its impact on design," *IEICE Trans. Electron.*, vol. E89-C, no. 3, pp. 342–348, Mar. 2006.
- [5] S. Digha, S. Vangal, P. Aseron, S. Kumar, T. Jacob, K. Bowman, J. Howard, J. Tschanz, V. Erraguntla, N. Borkar, V. K. De, and S. Borkar, "Within-die variation-aware dynamic-voltage-frequency-scaling with optimal core allocation and thread hopping for the 80-core teraflops processor," *IEEE J. Solid-State Circuits*, vol. 46, no. 1, pp. 184–193, Jan. 2011.
- [6] S. Borkar, T. Karnik, S. Narendra, J. Tschanz, A. Keshavarzi, and V. De, "Parameter variations and impact on circuits and microarchitecture," in *Proc. 40th Annu. Des. Autom. Conf.*, 2003, pp. 338–342.
- [7] C. Kim, K. Roy, S. Hsu, R. Krishnamurthy, and S. Borkar, "A process variation compensating technique with an on-die leakage current sensor for nanometer scale dynamic circuits," *IEEE Trans. Very Large Scale Integr. Syst.*, vol. 14, no. 6, pp. 646–649, Jun. 2006.
- [8] A. Ghosh, R. Rao, J. J. Kim, C.-T. Chuang, and R. Brown, "On-chip process variation detection using slew-rate monitoring circuit," in *Proc. 21st Int. Conf. Very Large Scale Integr. Des.*, Jan. 2008, pp. 143–149.
- [9] T. Iizuka, J. Jeong, T. Nakura, M. Ikeda, and K. Asada, "All-digital on-chip monitor for PMOS and NMOS process variability measurement utilizing buffer ring with pulse counter," in *Proc. ESSCIRC*, Sep. 2010, pp. 182–185.
- [10] H. Notani, M. Fujii, H. Suzuki, H. Makino, and H. Shinohara, "On-chip digital I_{dn} and I_{dp} measurement by 65 nm CMOS speed monitor circuit," in *Proc. IEEE Asian Solid-State Circuits Conf.*, Nov. 2008, pp. 405–408.
- [11] M. Bhushan, A. Gattiker, M. Ketchen, and K. Das, "Ring oscillators for CMOS process tuning and variability control," *IEEE Trans. Semicond. Manuf.*, vol. 19, no. 1, pp. 10–18, Feb. 2006.
- [12] T. Takahashi, T. Uezono, M. Shintani, K. Masu, and T. Sato, "On-die parameter extraction from path-delay measurements," in *Proc. IEEE Asian Solid-State Circuits Conf.*, Nov. 2009, pp. 101–104.
- [13] I. Mahfuzul, A. Tsuchiya, K. Kobayashi, and H. Onodera, "Variation-sensitive monitor circuits for estimation of die-to-die process variation," in *Proc. IEEE Int. Conf. Microelectron. Test Structures*, Apr. 2011, pp. 153–157.
- [14] T. Sakurai and A. Newton, "Alpha-power law MOSFET model and its applications to CMOS inverter delay and other formulas," *IEEE J. Solid-State Circuits*, vol. 25, no. 2, pp. 584–594, Apr. 1990.
- [15] R. Spence and R. S. Sooin, *Tolerance Design of Electronic Circuits*. Reading, MA: Addison-Wesley, 1988.

- [16] H. Onodera and H. Terada, "Characterization of WID delay variability using ro-array test structures," in *Proc. 8th IEEE Int. Conf. ASIC*, Oct. 2009, pp. 658–661.
- [17] M. Bhushan and M. Ketchen, "Generation, elimination and utilization of harmonics in ring oscillators," in *Proc. IEEE Int. Conf. Microelectron. Test Structures*, Mar. 2010, pp. 108–113.



Islam A. K. M. Mahfuzul (S'11) received the B.E. degree in electrical and electronics engineering in 2009 and the M.E. degree in communications and computer engineering in 2011, both from Kyoto University, Kyoto, Japan. He is currently pursuing the Ph.D. degree with Kyoto University.

His current research interests include process variation monitoring and variation-aware design techniques for low-power large-scale integration.



Akira Tsuchiya (M'05) received the B.E., M.E., and Ph.D. degrees in communications and computer engineering from Kyoto University, Kyoto, Japan, in 2001, 2003, and 2005, respectively.

Since 2005, he has been an Assistant Professor with the Department of Communications and Computer Engineering, Graduate School of Informatics, Kyoto University. His current research interests include modeling and design of on-chip passive components of high-frequency CMOS, and high-speed analog circuit design.

Dr. Tsuchiya is a member of the IEICE and IPSJ.



Kazutoshi Kobayashi (M'98) received the B.E., M.E., and Ph.D. degrees in electronics engineering from Kyoto University, Kyoto, Japan, in 1991, 1993, and 1999, respectively.

He joined the Graduate School of Informatics, Kyoto University, as an Assistant Professor in 1993. He was promoted to an Associate Professor and stayed in that position until 2009. For two years during this time, he was an Associate Professor with the VLSI Design and Education Center, University of Tokyo, Tokyo, Japan. Since 2009, he has been

a Professor with the Graduate School of Science and Technology, Kyoto Institute of Technology, Kyoto, and led his research group to give presentations in ASP-DAC, ISQED, IRPS, NSREC, and SSDM in 2011. While in the past he focused on reconfigurable architectures utilizing device variations, his current research interests include improving the reliability (soft errors and bias temperature instability) of current and future very large scale integration.

Dr. Kobayashi received the IEICE Best Paper Award in 2009.



Hidetoshi Onodera (M'87) received the B.E., M.E., and Dr.Eng. degrees in electronics engineering from Kyoto University, Kyoto, Japan, in 1978, 1980, and 1984, respectively.

He joined the Department of Electronics, Kyoto University, in 1983. He is currently a Professor with the Department of Communications and Computer Engineering, Graduate School of Informatics, Kyoto University. His current research interests include design technologies for digital, analog, and radio frequency large scale integration, with a particular

emphasis on low-power design, design for manufacturability, and design for dependability.

Dr. Onodera has served as the Program Chair and the General Chair of ICCAD and ASP-DAC. He has been the Chairman of the IPSJ SIG-System LSI Design Methodology (SLDM), the IEICE Technical Group on VLSI Design Technologies, the IEEE SCS Kansai Chapter, and the IEEE CASS Kansai Chapter. He has served as the Editor-in-Chief of the *IEICE Transactions on Electronics* and the *IPSJ Transactions on System LSI Design Methodology*.