Tunable Sensors for Process-Aware Voltage Scaling

Tuck-Boon Chan
ECE Department, UC San Diego
La Jolla, CA 92093, USA
tbchan@ucsd.edu

Andrew B. Kahng
CSE and ECE Departments, UC San Diego
La Jolla, CA 92093, USA
abk@ucsd.edu

***Abstract*—VLSI circuits usually allocate excess margin to account for worst-case process variation. Since most chips are fabricated at process conditions better than the worst-case corner, adaptive voltage scaling (AVS) is commonly used to reduce power consumption whenever possible. A typical AVS setup relies on a performance monitor that replicates critical paths of the circuit to guide voltage scaling. However, it is difficult to define appropriate critical paths for an SoC which has multiple operating modes and IPs. In this paper, we propose a different methodology for AVS which matches the voltage scaling characteristics of a circuit rather than the delays of critical paths. This fundamental change in monitoring strategy simplifies the monitoring circuitry as well as the calibration flow of conventional monitoring methods. To enable the proposed methodology, we study voltage scaling characteristics of digital circuits. Based on our analyses, we develop design guidelines as well as design monitoring circuits which have tunable voltage scaling characteristics. Our experimental results show that this methodology can be used for AVS with a simplified calibration flow.**

I. INTRODUCTION

Process variation is a critical aspect of VLSI circuit design because it causes wide performance spread [2] [13]. To recover excess margin allocated for process variation, many adaptive voltage scaling (AVS) techniques have been proposed [5] [9] [14] [16] [17].

AVS techniques can be classified as either open- or closed-loop. A typical *open-loop* AVS system utilizes a pre-characterized lookup table (LUT) to find the corresponding minimum supply voltage for a given chip frequency target [14] [5]. Since the open-loop technique does not have a feedback mechanism, the LUT is heavily guardbanded to ensure reliable system operation. At the same time, characterizing the LUT is a time-consuming and expensive procedure, especially for a system-on-chip (SoC) design which has multiple operating modes and IPs.

A *closed-loop* AVS system adjusts supply voltage by probing actual chip performance, using on-chip monitors instead of using a LUT. To track timing performance of a chip, many critical path replica or *in-situ monitor* approaches have been proposed [7] [9] [19] [15] [16] [10] [6] [18]. However, the "critical paths" in a multiple-IP SoC design are not clearly defined, as chip performance depends on both operating mode and interactions among the IPs. Moreover, there are cases where exact input vectors to exercise worst-case timing paths in an SoC are not known during design time.

In this paper, we propose an approach to design sensors for *process-aware voltage scaling* (PVS). Instead of designing performance monitors to track the timing performance of critical paths, we design ring-oscillators (ROs) which have the worst-case voltage scaling characteristics across the entire process condition (see Section II for the details of voltage scaling characteristics). We design the PVS ROs such that they require a relatively higher supply voltage compared to critical paths of a SoC to compensate process variation-induced frequency drift. Therefore, any SoC manufactured in the process can safely perform a closed-loop AVS by using these ROs as hardware performance monitors. A new analysis of voltage scaling characteristics is a key enabler to our PVS methodology. Design ROs for worst-case voltage scaling characteristics is distinguished from a conventional RO-based monitoring method (e.g., [3]) which uses an arbitrary RO.

IEEE/ACM International Conference on Computer-Aided Design (ICCAD) 2012, Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

IEEE/ACM International Conference on Computer-Aided Design (ICCAD) 2012, November 5-8, 2012, San Jose, California, USA

Copyright (c) 2012 ACM 978-1-4503-1573-9/12/11... $15.00

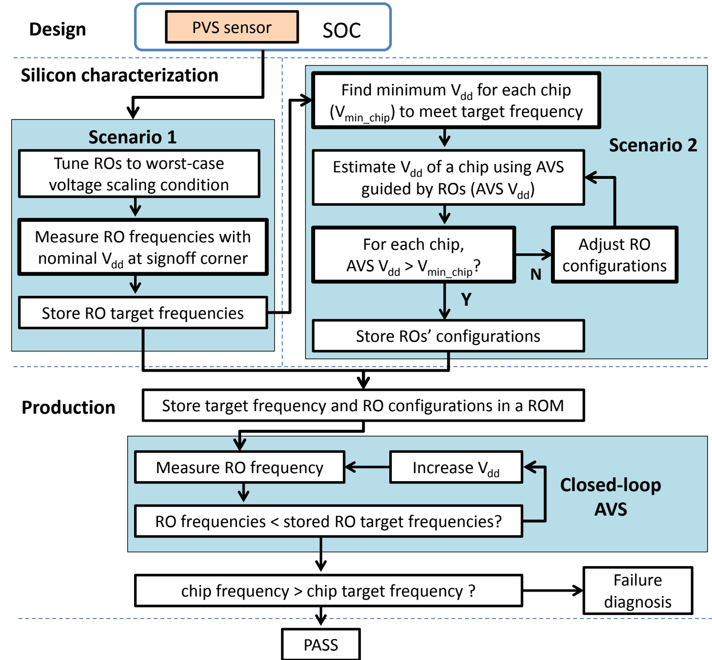


Fig. 1. An application example for the proposed tunable ROs.

Application examples (scenarios) for the proposed ROs are shown in Figure 1. At the design stage, we design the PVS ROs using SPICE models and standard cells. Since there will be some difference between simulation and the silicon data, a silicon characterization step is required to calibrate the error between simulation and silicon data. At the silicon characterization stage, sample test chips at different process corners are provided by the foundry. In this stage, we measure the ROs' frequencies with nominal operating voltage (V_0). The frequencies measured at the signoff corner (e.g., SS corner) will be used as the target frequencies of the ROs during AVS (**Scenario 1**). In this application scenario, our ROs have no information about the design, and they are designed to guardband for the worst-case voltage scaling characteristics. Therefore, the AVS guided by our ROs will always overestimate the supply voltage needed for a chip to meet its operating frequency. The excess supply voltage can be reduced when chip maximum frequency f_{max} is also measured during the silicon characterization stage (**Scenario 2**). In this scenario, we can tune the voltage scaling characteristics of the ROs so that for each chip in the silicon characterization stage, the supply voltage suggested by the AVS (guided by the ROs) is slightly higher than the minimum voltage (V_{min_chip}) needed for a chip to meet its required operating frequency. When all test chips manufactured for silicon characterization can

safely operate at their respective operating frequencies using AVS guided by the PVS ROs, we record the configurations of the ROs. In this characterization step (Scenario 2), the test chips are manufactured at biased process corners. Thus, calibrating the ROs with these test chips will configure the ROs to account for circuit performance variation due to widely spread process variation. Sampling the test chip at different process corners is important because this allows the configurations of the ROs to be applied in the subsequent production stage without additional calibrations.

To capture the within-die systematic process variation, we can place multiple copies of the ROs in a chip (e.g., a set of ROs for every $1mm^2$ area on the chip). However, the effect of within-die random variation cannot be captured by our method due to the nature of the replica-type monitoring approach. Thus, additional timing or voltage margin must be added to ensure reliable circuit operation. Meanwhile, by having multiple copies of the ROs in a chip, the effect of within-die temperature variation on circuit performance can be also captured by the ROs.

During mass production, the previously obtained ROs' configurations will be stored in every production chip. Then, we run AVS tests with the stored ROs' configurations and RO target frequencies. If a chip fails to meet its target frequency with the AVS guided by PVS ROs, this means that either the calibration during silicon characterization is inaccurate or the chip has failed due to other reasons. After studying the root cause of the failure, the silicon characterization step can be modified if necessary (e.g., adjust ROs' configurations so that the AVS is less aggressive in reducing supply voltage).

Note that in Scenario 1, we skip the procedures of Scenario 2, and all ROs are configured to the worst-case voltage scaling condition. Although this approach leads to a more pessimistic AVS, the tunability of the ROs allows the chip customer to recover the pessimism in AVS by calibrating RO configurations. Since the PVS ROs are design-independent, a PVS IP can be embedded in different SoCs to support AVS. For example, PVS ROs can be deployed within a performance monitor block in a power management IP such as [24].

Our method is different from critical path-driven tunable circuits [7] [9]. First, critical path replica techniques design the replica to be flexible to match the timing performance of a set of critical paths. Because of the inherent design intention to match the timing performance, the design of a critical path replica is dependent on the circuit to be matched (e.g., the TRC must have the flexibility to match the total critical delays). By contrast, we design our tunable ROs so that they can be configured to have different *voltage scaling characteristics*. This difference in design intention is important because, as we will show, matching the voltage scaling characteristics of different circuits can be achieved by having a set of tunable ROs which are design-independent. As a result, we can optimize the ROs and reuse them in other designs. Second, our proposed method only calibrates the ROs at the silicon characterization stage. After this calibration step, the settings will be applied to all production chips instead of calibrating the ROs for every production chip. Since per-chip calibration is not required, our method saves testing time during chip production. We summarize our contributions as follows.

- We propose a simplified *process-aware voltage scaling* (PVS) methodology and analyses of the worst-case condition of voltage scaling under process variation.
- We propose circuit techniques to tune the voltage scaling characteristic of the sensor so that it has flexibility to mimic the voltage scaling characteristics of a chip across a range of process variations. With the tunability, we can reduce the supply voltage by up to 30 mV (compared to non-tunable ROs) without causing any timing violation.
- Our tunable sensor is design-independent, and can therefore be embedded in any other IPs.

The rest of this paper is organized as follows. In Section II, we discuss the basic concepts of the proposed PVS methodology. In Section III, we discuss voltage scaling characteristics of CMOS circuits. We present a tunable sensor in Section IV and show experimental results for the proposed tunable sensor in Section V. Finally, we conclude the paper in Section VI.

II. Process-Aware Voltage Scaling

A. Overview of PVS

Figure 2 shows the basic idea of the PVS methodology, wherein we model the frequency of a critical path as a linear function of supply voltage (V)[1]. In this paper, we denote the frequency of a critical path by $f_{path}(j,k,V)$ where j is the index of a critical path, k denotes the process condition, and V is the supply voltage. Similarly, we define the frequency of an RO by $f_{ro}(i,k,V)$, where i is the index of a RO.

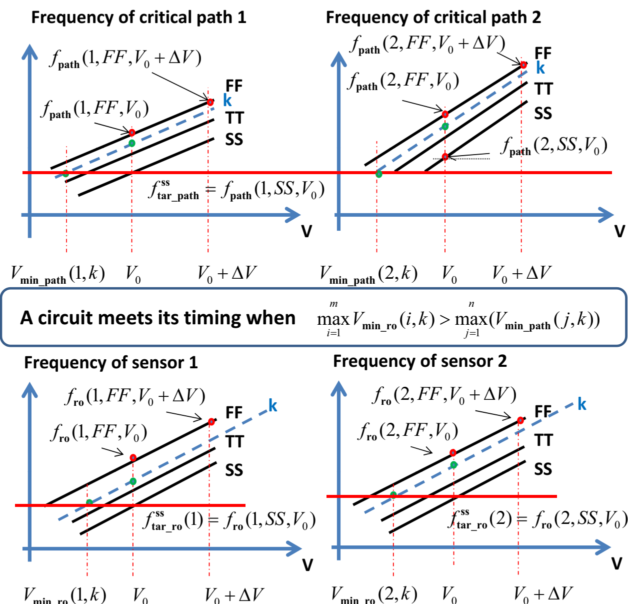


Fig. 2. Illustration of process-aware voltage scaling.

We define the target frequency of the critical paths f_{tar_path} as the minimum frequency of all critical paths at nominal voltage V_0. Note that the target frequency is specific to the signoff corner. Unless otherwise specified, we define the target frequency at the SS corner, i.e,

$$f^{ss}_{tar_path} = \min_{j=1}^{n} f_{path}(j,SS,V_0)$$

where n is the total number of critical paths, V_0 is the nominal voltage and $f^{ss}_{tar_path}$ is the target frequency of the chip at the SS signoff corner.

When a circuit is manufactured at process condition k (dashed line in Figure 2), the frequency of the circuit is significantly higher than $f^{ss}_{tar_path}$. Thus, we can perform voltage scaling to reduce the power of the circuit as long as the circuit meets the targeted frequency. The minimum voltage required for a critical path j to meet its targeted frequency at a process condition k is denoted as $V_{min_path}(j,k)$. When there is more than one critical path, the minimum voltage for a circuit $V_{min_chip}(k)$ is given by

$$V_{min_chip}(k) = \max_{j=1}^{n} V_{min_path}(j,k) \tag{1}$$

[1] This approximation simplifies calculation while introducing small error [9].

As mentioned above, finding the exact critical paths in an SoC to calculate $V_{min_chip}(k)$ is very difficult. Therefore, we propose to adjust the supply voltage of a circuit by measuring the frequencies of on-chip ROs. As shown in the lower part of Figure 2, the frequency of the ith RO is represented as $f_{ro}(i,k,V)$. The target frequency of each on-chip RO ($f^{ss}_{tar_ro}(i)$) is defined at the same signoff corner as the circuit, e.g., $f^{ss}_{tar_ro}(i) = f_{ro}(i,SS,V_0)$, and each RO has a specific target frequency. We denote $V_{min_ro}(i,k)$ as the minimum voltage for the ith RO to meet its targeted frequency, where k represents the process condition of the RO. By measuring the RO frequencies at two or more supply voltages, we can extract each RO's frequency-versus-voltage "slope", and calculate $V_{min_ro}(i,k)$ from the equation

$$V_{min_ro}(i,k) = V_0 - \frac{(f_{ro}(i,k,V_0) - f^{ss}_{tar_ro}(i))\Delta V}{f_{ro}(i,k,V_0+\Delta V) - f_{ro}(i,k,V_0)} \quad (2)$$

where ΔV is the difference between nominal voltage and voltage during RO measurement. After obtaining $V_{min_ro}(i,k)$, we can use it as a reference to scale the supply voltage of the chip. A chip will still meet its performance target as long as $V_{min_ro}(i,k)$ is larger than $V_{min_chip}(k)$. Thus, the "safe voltage scaling condition" for a chip is defined as

$$V_{min_chip}(k) < \max_{i=1}^{m}\{V_{min_ro}(i,k)\}, \forall\ k \quad (3)$$

To ensure that the chip meets its targeted frequency, we scale the supply voltage of the chip to

$$V_{min_est}(k) = \max_{i=1}^{m}\{V_{min_ro}(i,k)\} \quad (4)$$

B. Fundamental Properties of PVS

Equation (2) shows that the minimum scaling voltage of a RO (or a critical path) is determined by two fundamental properties:

(1) **Process distance**: $f_{ro}(i,k,V_0) - f^{ss}_{tar_ro}(i)$
(2) **Scaling rate** : $(f_{ro}(i,k,V_0+\Delta V) - f_{ro}(i,k,V_0))/\Delta V$

Process distance is the process-induced frequency shift relative to target frequency. This property is usually modeled as a random variable due to the randomness in manufacturing processes. However, it is also affected by the design of the circuit. For example, different critical paths have different sensitivities to sources of process variation. Another fundamental aspect of PVS is its formulation based on a *scaling rate* of frequency with respect to supply voltage. Clearly, this is also a circuit-related property which varies depending on the process condition.

Note that *voltage scaling* for a circuit is defined by relative value of the process distance and the scaling rate (i.e., process distance/scaling rate). Based on these properties, we can derive the voltage scaling characteristic of an arbitrary circuit. We are interested in studying the following questions:

(1) Given a process technology, what is the range of voltage scaling defined by process distance and scaling rate?
(2) What circuit techniques can be used to design a monitoring circuit with tunable voltage scaling characteristics?

Answering the first question helps to identify the worst-case voltage scaling condition, which is the design goal of our PVS ROs. Answering the second question gives us feasible design options to design PVS ROs to achieve the goal.

III. Circuit Analysis

A. Voltage Scaling Sensitivity

As mentioned above, the voltage scaling characteristic of a critical path is given by

$$\begin{aligned}\text{voltage scaling} &\equiv \frac{\text{process distance}}{\text{scaling rate}} \\ &\equiv \frac{f_{path}(i,k,V_0) - f^{ss}_{tar_path}(i)}{f_{path}(i,k,V_0+\Delta V) - f_{path}(i,k,V_0)}\end{aligned} \quad (5)$$

To gain intuition about the sensitivity of voltage scaling to circuit parameters, we model $f_s(.)$ using the Elmore delay model [8].

$$\begin{aligned}f_{path}(i,k,V_0) &= \frac{2}{D_n(i,k,V_0) + D_p(i,k,V_0)} \\ D_n(i,k,V_0) &= \frac{R_n(k,V)}{w}(1+\beta)[w(\beta+1)C_g(k)N + l*C_w] \\ &\quad + l^2R_wC_w + lR_w(\beta+1)C_g(k)N \\ D_p(i,k,V_0) &= \frac{R_p(k,V)}{w\beta}(1+\beta)[w(\beta+1)C_g(k)N + l*C_w] \\ &\quad + l^2R_wC_w + lR_w(\beta+1)C_g(k)N\end{aligned} \quad (6)$$

where l is wire length, w is channel width of NMOS, N is the fanout of the driver, R_w is wire resistance per μm, C_w is wire capacitance per μm, β is the beta ratio between PMOS and NMOS channel width, $C_g(k)$ is gate capacitance per μm channel width, and $R_n(k,V)$ and $R_p(k,V)$ are effective drive resistance of NMOS and PMOS, respectively. To study the sensitivity of voltage scaling, we extract parameters in (6) from an inverter of a 65*nm* foundry library. The values of $R_n(k,V)$ and $R_p(k,V)$ are calculated by using effective current approximation [1],

$$\begin{aligned}R_{n,p}(k,V) &= \frac{2V}{I_L + I_H} \\ I_L &= I_{ds} \text{ when } V_{gs} = V/2, V_{ds} = V \\ I_H &= I_{ds} \text{ when } V_{ds} = V/2, V_{gs} = V\end{aligned}$$

where I_L and I_H are the drive currents (I_{ds}) of a MOS transistor at different bias conditions. The parameters and effective currents are summarized in Table I.

TABLE I
TECHNOLOGY PARAMETERS OF A 65*nm* LIBRARY.

	Process corners		
parameters	SS	TT	FF
w (μm)	0.09	0.09	0.09
R_w (Ω/μm)	0.16	0.16	0.16
C_w (fF/μm)	0.00017	0.00017	0.00017
C_g (fF/μm)	1.03	1.09	1.16
I_L NMOS, 1.0V (μA)	52	134	258
I_L NMOS, 0.9V (μA)	29	87	192
I_H NMOS, 1.0V (μA)	459	591	723
I_H NMOS, 0.9V (μA)	348	470	594
I_L PMOS, 1.0V (μA)	29	66	125
I_L PMOS, 0.9V (μA)	16	41	88
I_H PMOS, 1.0V (μA)	232	294	353
I_H PMOS, 0.9V (μA)	172	227	281

Using the parameters in Table I, from Equations (5) and (6) we calculate V_{min} of the inverter for TT corner (i.e., $k = TT$) and its sensitivities. First, we calculate the nominal V_{min} of the inverter with l = 10μm, w = 1μm, β = 1.5, N = 1. Then, we sweep the value of the l, w, β, N, R_n and R_p parameters, one at a time (other parameters remain at their nominal values), from 0.2 to 4 times of their nominal values, to evaluate the effect of each parameter on V_{min}. The results in Figure 3 show that V_{min} is most sensitive to R_n and R_p, followed by β, l, fanout, and w. We also observe that when the value of each parameter is increased, its impact on the value of V_{min} becomes smaller. V_{min} changes rapidly as the (normalized) parameter values scale below 1.0. There is also a practical lower limit for the parameters. For example, the driver size (w), fanout, R_n, etc. cannot scale down to zero. Hence, voltage scaling of a circuit has finite bounds. From our studies, we also observe that V_{min} can be

significantly lower (resp. higher) when we only consider D_n (resp. D_p) in Equation (6).

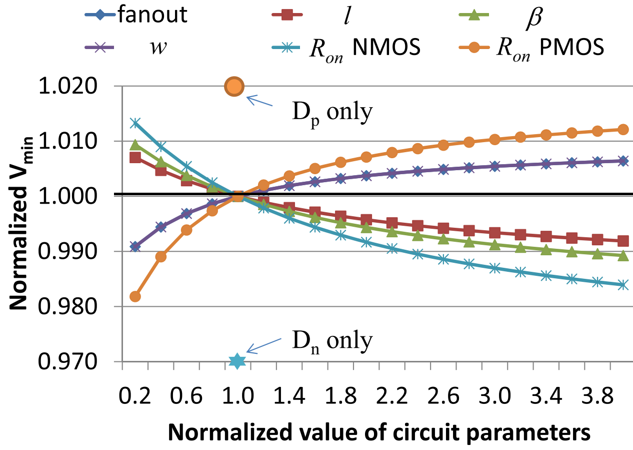
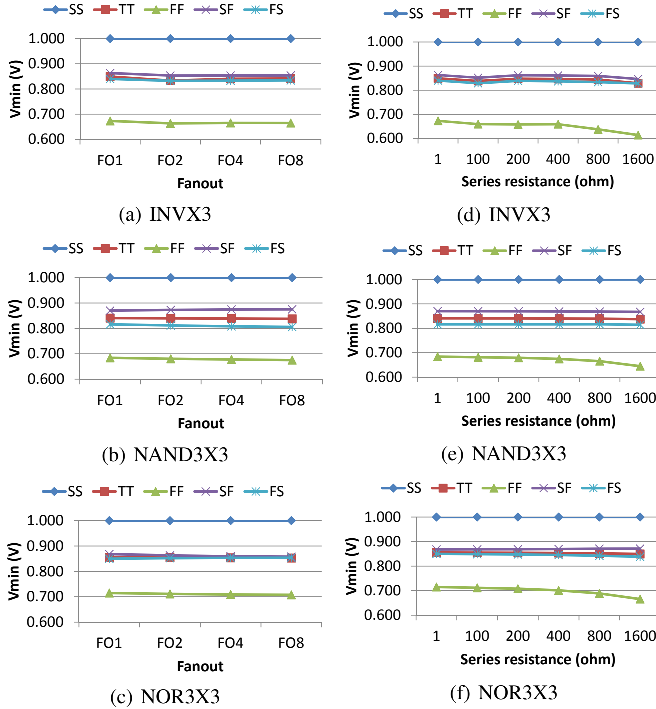


Fig. 3. Sensitivity of V_{min} to circuit parameters.

B. Voltage Scaling Analysis Using SPICE Simulation

Although the previous analysis provides useful information regarding the sensitivities of V_{min} to circuit parameters, many effects are not captured by the simplified equations. To investigate the range of voltage scaling as well as the effect of circuit parameters, we simulate different ring-oscillators with different configurations.



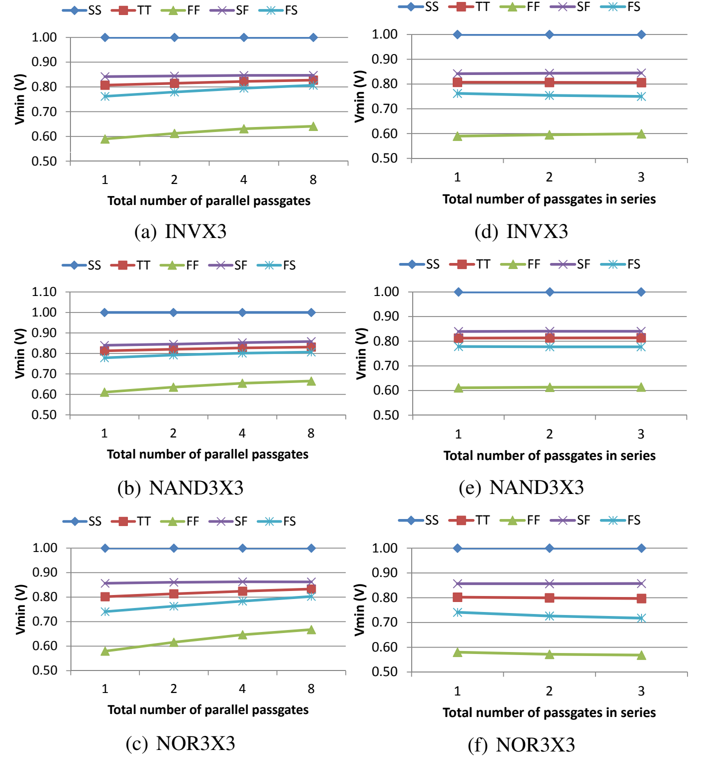
(a) INVX3 (d) INVX3

(b) NAND3X3 (e) NAND3X3

(c) NOR3X3 (f) NOR3X3

Fig. 4. SPICE simulations of ROs implemented with INV, NAND and NOR standard cells. For the fanout experiment, the output of each gate in the RO is connected to multiple dummy gates to achieve different fanout values. Then, using the same ROs, we add series resistance (DC) to the output of every gate. The results show that V_{min} is not sensitive to fanout and series resistance.

First, we evaluate the effect of fanout by adding dummy gates in every stage in the RO. Figure 4 shows that V_{min} extracted from the ROs is not sensitive to fanout for ROs implemented with different standard cells. Second, we increase the series resistance along the signal transition path of the ROs with fanout = 1. Figure 4 shows that series resistance can affect V_{min} when the resistance value is large. For 65*nm* technology, the wire resistance per μm is approximately 0.16Ω. Therefore, V_{min} at 400Ω corresponds to the case where a 2.5mm long wire is connected to the output of a driver. Since reasonable design usually does not permit such a long wire, it is safe to assume that wire resistance will not affect V_{min}. This implies that the voltage scaling characteristic of a chip is not affected by wire parasitics.



(a) INVX3 (d) INVX3

(b) NAND3X3 (e) NAND3X3

(c) NOR3X3 (f) NOR3X3

Fig. 5. V_{min} is increased when the number of passgates in parallel is increased. Adding more passgates in series has little effect on V_{min}.

Third, we add passgates at the output of each driver of the ROs to study their effects on V_{min}. To study different scenarios, we also change the effect of the passgates by adding more passgates in parallel or in series. Results in Figure 5 show that adding passgates in parallel can change the V_{min} significantly. V_{min} increases when the number of parallel passgates is increased. This is because more passgates in parallel reduces the series resistance of the ROs. This result agrees with the estimations obtained in (6), in which increasing l reduces V_{min}. Figure 5 shows that V_{min} changes only slightly when the number of series passgates is increased. This is because the effect of adding series resistances saturates as the sum of series resistance increases.

Equation (6) shows that R_n or R_p has significant impact on V_{min}. To study this, we simulate ROs with different standard cell types. Results in Figure 6 show that V_{min} varies over ROs with different cell types. For example, we see that V_{min} of NOR-based ROs is larger than that of INV-based ROs. This is because the NOR-type standard cell has a stacked pull-up network with a larger R_p compared to the balanced pull-up and pull-down networks of an inverter. On the other hand, V_{min} of NAND-based ROs is smaller than that of INV-based ROs especially at TT and FS process corners. This agrees with the estimations obtained from (6), where V_{min} is smaller for a larger R_n (a NAND gate has a larger R_n compared to an INV gate). However, the trend is not obvious at FF process corner. This may be due to layout

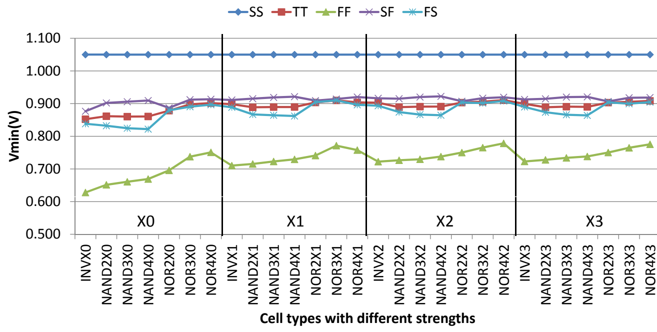


Fig. 6. V_{min} varies across different cell types {INV, NAND2, NAND3, NAND4, NOR2, NOR3, NOR4} and strength {X0, X1, X2, X3}.

parasitics and other second-order effects not modeled in our analysis. Note that V_{min} increases sharply when the driver is increased from minimum size (X0) to larger sizes. This is due to the diffusion height of the minimum-sized cell being significantly less than the row height of the standard cell. Thus, the layout parasitics of cells with minimum driver size are typically different from those of other cells. Note that the maximum value of V_{min} at different corners is determined by the V_{min} of different cell types. For example, the NAND-based RO has the largest V_{min} at SF corner while the NOR-based RO has the largest V_{min} at FS corner. Therefore, we require ROs implemented with different cell types to ensure that we capture the worst-case scenario in voltage scaling.

IV. Design of a Sensor with Tunable Voltage Scaling Characteristics

From the studies in the previous section, we observe that the voltage scaling characteristic of a circuit (RO) is mainly affected by the cell type. Among the circuit parameters, we only see significant changes in V_{min} when we add passgates in parallel to the ROs. Thus, we design our PVS sensor with different cell types and use passgates in parallel to tune the characteristic of the ROs. Our PVS sensor design seeks to achieve two main goals:

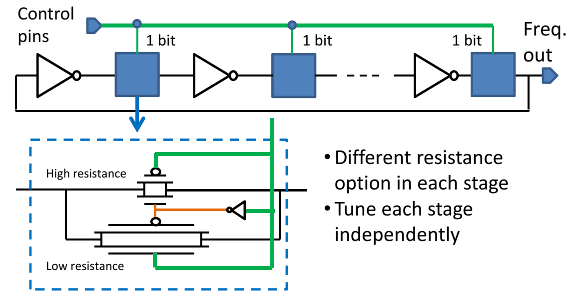
(1) maximize the range of V_{min}; and
(2) ensure that tunability of the sensor (V_{min} versus RO configuration) is consistent across different process corners.

Here, we present two of the circuit approaches that we have investigated to achieve these goals. The circuits are illustrated in Figure 7.

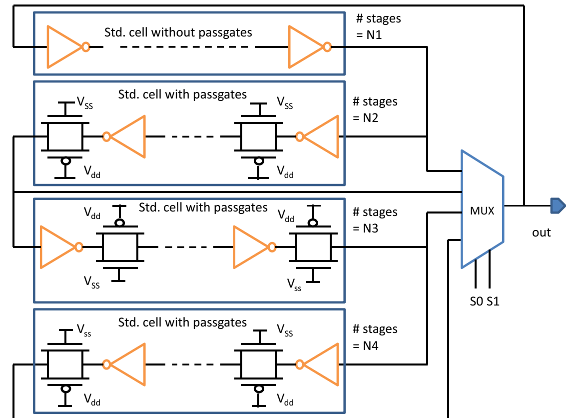
In the first approach, we add a pair of passgates in parallel at every stage of a RO, one with minimum-sized devices and the other with large device sizes. In this design, we can choose to turn on one passgate through a control pin assigned to the passgate. When we choose to turn on the passgate with minimum-sized devices, the high resistance passgate will reduce V_{min} – and vice-versa when we turn on the passgate with larger device sizes. Although we can assign a control pin for each stage of the RO to achieve fine granularity, having a large number of control pins will incur higher design and area overheads. Since the voltage levels in an AVS system are discrete with coarse granularity, there is no need to have very fine granularity for the sensor. In this paper, we divide the 33 stages of the RO into nine sections (the last section has five stages whereas all other sections have four stages), with all passgates in each section sharing a control pin. Thus, only nine control pins are required instead of 33.

In the second approach, we divide an RO into several sections and connect the output of the sections to a MUX such that we can choose which section is included in the oscillation. For example, when we set the MUX select bits to {0,0}, the output of the MUX is connected to "IN 1". As a result, only the first section is included in the oscillation. If we change the select bits to {0,1}, then the first and second sections are included. The advantage of this method is that through the MUX and select bits, we can bypass the cells with passgates, and achieve the maximum V_{min} of the RO (adding passgates will reduce V_{min}). Since the V_{min} of the RO is determined by the ratio between stages with and without passgate cells, always including the first section could limit the tunability. For example, we need a large number of stages with passgates (and area) to increase the ratio of cells with passgates to cells without passgate.

Simulation results in Figure 8 and Figure 9 show that both of these circuit approaches achieve similar ranges of tunability. Since the first approach has lower area overhead, we choose it for use in our simulation experiments. Based on the analysis in Figure 6, we observe that the maximum V_{min} is determined by different gate types, depending on the process conditions. To ensure that the ROs can have the maximum V_{min} across different process conditions, we choose to build the RO in 7(b) with INVX3, NAND3X3 and NOR3X3 instances[2]. As mentioned above, the circuit option in Figure 7(b) has a slightly lower V_{min_ro} due to the passgates in the ROs. To ensure that V_{min_ro} of the ROs includes the worst-case voltage scaling characteristic, we add an additional $5mV$ margin to the V_{min_ro} in our simulation experiments.



(a) We can use a MUX-like structure to control the ratio between different gates. Since V_{min} varies from one gate to another, we can connect different gates in series to achieve tunability of V_{min}.



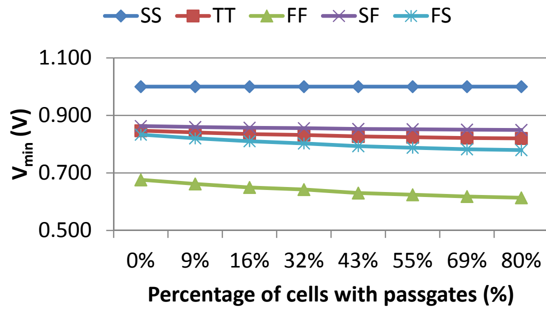
(b) By controlling the select bits, we can change the number of series transistors along the signal transition path of the RO. This changes the effective resistance when the RO charges or discharges a node. As a result, this changes the V_{min} of the sensor.

Fig. 7. Proposed tunable circuits.

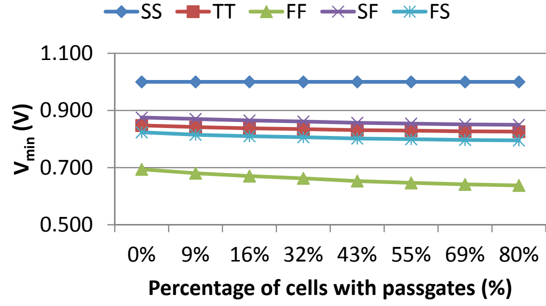
V. Experiment and Results

In our experiments, we use three modules of the *OpenSPARC T1* processor [20] (Table II). Module designs are implemented with a

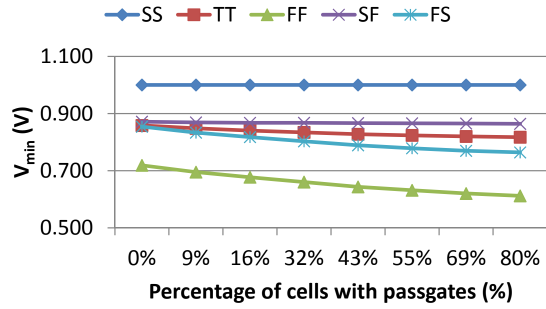
[2]For gates with multiple inputs, we connect the inputs as a single net.



(a) INVX3



(b) NAND3X3



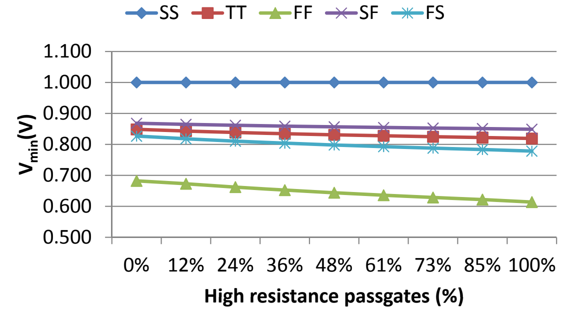
(c) NOR3X3

Fig. 8. V_{min} is minimum when the RO consists of standard cells without passgates. By controlling the values of N_1, N_2 etc., we can control the percentage of cells without passgates, and achieve a linear relationship between V_{min} and the decimal values represented by the select bits of the MUX.

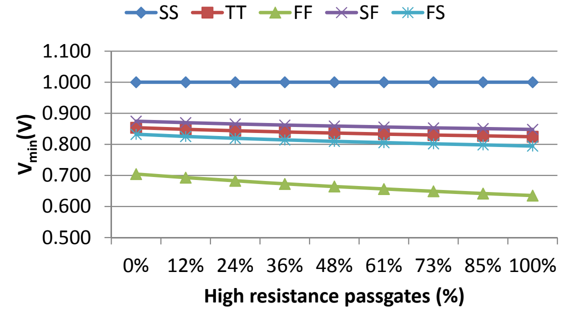
65*nm* foundry library. The netlists are synthesized with *Synopsys Design Compiler vD-2010.03-SP1* [21]. We extract critical paths of the modules in Table II at *SS*, *TT* and *FF* corners with *Synopsys PrimeTime vC-2009.06-SP2* [22]. For each process corner, we extract the top 100 critical paths and their corresponding SPICE netlists. We then simulate all the critical paths with *Synopsys HSpice vE-2010.12* [23] at *SS* corner, V_0 = 1.0V and 125^oC to obtain the $f^{ss}_{tar_path}$ of each module. The $f^{ss}_{tar_path}$, power and area values of the implemented modules are given in Table II.

TABLE II
OPENSPARC T1 MODULES. $V_0 = 1.0$V.

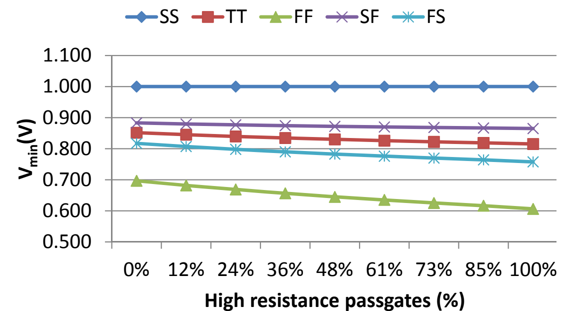
	power (mW)	area (mm^2)	$f^{ss}_{tar_path}$ (MHz)
fpu_div	4.13	0.015	710.2
tlu	438	0.098	506.6
mul_top	19.8	0.050	1042.1



(a) INVX3



(b) NAND3X3



(c) NOR3X3

Fig. 9. V_{min} of the proposed circuit for different standard cells. Through controlling the percentage of cells with higher resistance, we can tune the V_{min} of the RO.

TABLE III
GLOBAL VARIATION PARAMETERS

Variation source	μ	3σ
$\Delta\ V_{thn}$	0	30mV
$\Delta\ V_{thp}$	0	30mV
Δ channel length	0	5.00nm
Δ oxide thickness	0	0.06nm

A. Guardband Voltage Scaling

We perform an experiment to validate that our PVS sensors satisfy the "safe condition" in Equation (3) when the ROs are configured to have maximum V_{min_ro} (i.e., all passgates in the ROs have low resistance). To emulate process variation, we model threshold voltage of NMOS (V_{thn}) and PMOS (V_{thp}), channel length and oxide thickness as independent Gaussian random variables. The 3σ values of these variation sources are extracted from the foundry device

model.[3] The mean (μ) and standard variation (σ) of the random variables are summarized in Table III.

To estimate timing performance of the critical paths and ROs under process variations, we sample the variation sources randomly. We then apply the variations when running an HSPICE simulation, and repeat this 100 times. This Monte Carlo experiment only includes global variation because our simulation setup does not support a local variation model.

Based on the simulated critical paths and RO delays, we calculate V_{min_chip} and V_{min_est} based on their definitions in Equations (1) and (4). Since there are INV-, NAND- and NOR-based ROs, V_{min_est} is the maximum V_{min_ro} of the three ROs. For comparison, we also include the results of non-tunable INVX3-, NAND3X3- and NOR3X3-chained ROs. These ROs are similar to our ROs, but there is no passgate in between consecutive stages.

Figure 10 shows that the voltage difference between V_{min_est} and V_{min_chip} is always positive. This implies that the sensors can be used to guardband the modules without calibration.

B. Optimizing Target Frequency for Margin Reduction

Our next experiment considers a scenario where V_{min_chip} of every chip is available to calibrate the PVS sensors. Hence, we can optimize the configuration (control bits) of the tunable ROs to reduce supply voltage. The problem can be formulated as follows:

$$\begin{aligned} \text{min. :}\ & \sum_k \{V_{min_est}(k) - V_{min_chip}(k)\} \\ \text{s.t. :}\ & V_{min_est}(k) > V_{min_chip}(k), \forall\ k, i \\ & \max_i [V_{min_ro}(i,k)|_{\gamma(i)}] = V_{min_est}(k) \\ & V_{min_ro}(i,k)|_{\gamma(i)} = V_0 + \frac{f_{tar_ro}(i)|_{\gamma(i)} - f_{ro}(i,k,V_0)|_{\gamma(i)}}{\alpha(i,k)|_{\gamma(i)}} \end{aligned} \quad (7)$$

where $\gamma(i)$ denotes the configuration of the ith RO. Note that $f_{tar_ro}(i)$, $f_{ro}(i,k,V_0)$ and $\alpha(i,k)$ are all specific to $\gamma(i)$. This ensures that V_{min_est} guided by our ROs is always less than V_0. This property is a key reason why the tunability in our circuit is different from using f_{tar_ro} as a means to adjust voltage scaling. For example, increasing f_{tar_ro} will cause the chip at SS corner to operate at a voltage higher than V_0, which may cause reliability-related failures. Since each INV, NAND or NOR RO has 9 configurations, we calculate $V_{min_est}(k)$ for all 729 combinations. After that, we compare the $V_{min_est}(k)$ with $V_{min_chip}(k)$, and discard solutions that violate the safe condition in (3). Finally, for each $V_{min_est}(k)$ that satisfies the safe condition, we calculate the average of its resultant $V_{min_est}(k)$ across k process conditions.

The results in Table IV show that the tunable sensor can achieve a lower supply voltage compared to the normal (non-tunable) ROs in all cases. From the experimental data, we see that the benefits of the tunability vary depending on the difference between V_{min_est} and V_{min_chip}. For example, Figure 10 shows that the V_{min_est} values obtained from the non-tunable ROs are very close to the V_{min_chip} values, especially for the fpu and mul_top modules. Thus, there is not much room left in which to reduce V_{dd} without causing a timing violation. When V_{min_est} is larger than V_{min_chip}, we can recover the wasted voltage margin by tuning the configurations of PVS ROs. Figure 11 shows that by tuning the configuration of the PVS ROs, we can obtain a more aggressive AVS configuration for voltage reduction. For the maximum voltage reduction configuration shown in the figure (green color), we can achieve about $13mV$ voltage reduction compared to the non-tunable ROs, on average (mean of 100 Monte Carlo samples). Note that the voltage reduction varies depending on the process variation. For example, the maximum V_{min} reduction compared to the non-tunable ROs is $31.3mV$ for a specific instance.

[3] We assume that process parameters at SS and FF corners define the $\mu \pm 3\sigma$ of the variation sources.

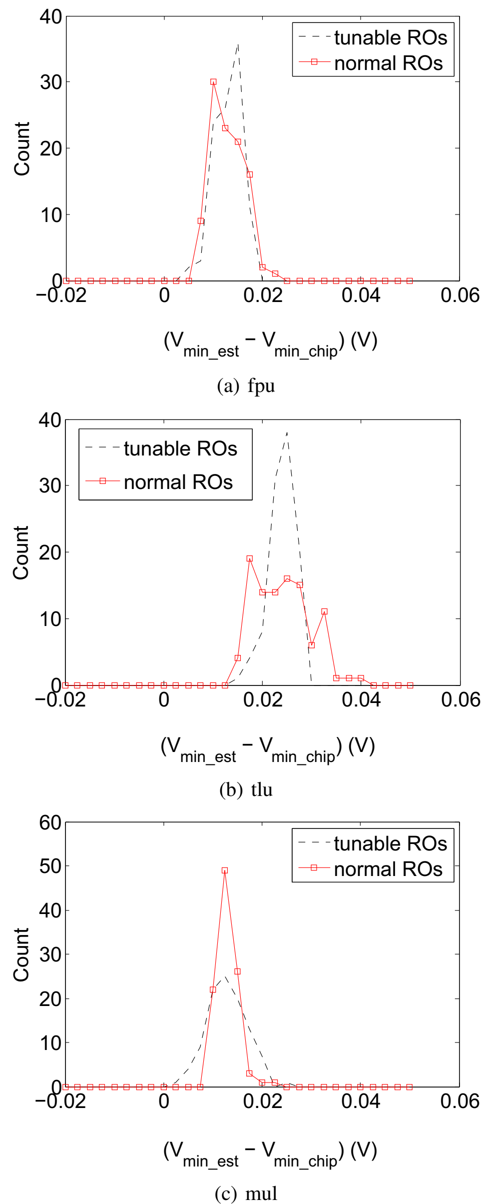


Fig. 10. Distributions of $(V_{min_est} - V_{min_chip})$ for different circuit modules. The results show that $V_{min_est} - V_{min_chip}$ is always positive. This implies that the tunable ROs can be used for voltage scaling without causing any timing violations.

In summary, our experimental results confirm that our methodology allows selection of standard cells to build ROs with worst-case voltage scaling characteristics, which can be used as performance monitors for AVS. The overhead $(V_{min_est} - V_{min_chip})$ of these ROs varies depending on the circuit. Although our study uses single-V_{th} devices, the methodology can be extended to designs with multi-V_{th} devices by having a set of ROs for each V_{th}. Since the V_{min_est} in our methodology is defined by the maximum V_{min_ro} of all ROs, the V_{min_chip} defined by mixed-V_{th} cells will always be less than V_{min_est}.

VI. Conclusion

In this paper, we have presented a different approach to enable process-aware voltage scaling. In contrast to the conventional

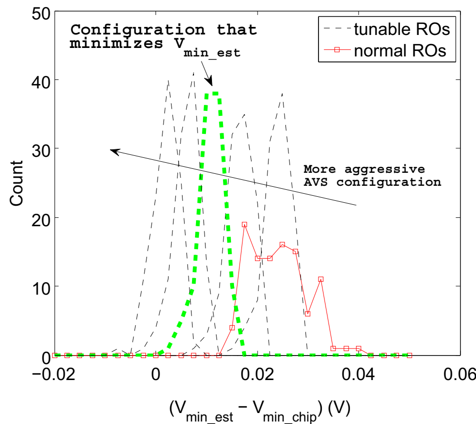


Fig. 11. Distribution of $(V_{min_ro} - V_{min_chip})$ for the tlu testcase with different PVS RO configurations. By tuning the configuration of the ROs, we can change the voltage scaling characteristics (V_{min_est}). An optimized configuration can reduce V_{min_est} by $13mV$ (on average) compared to normal ROs.

TABLE IV
V_{min_est} REDUCTION ENABLED BY THE TUNABILITY OF PVS ROS.

	V_{dd} reduction		mean
	average (mV)	maximum (mV)	$V_{min_chip}(mV)$
fpu_div	2.7	16.8	851
tlu	13.3	31.3	840
mul_top	2.7	16.8	851

monitoring approaches that attempt to track critical paths, we propose to enable process-aware AVS by synthesizing a set of ROs which achieve a worst-case voltage scaling property across different process conditions. Since the ROs always require a relatively higher voltage to meet their target frequencies compared to that required by critical paths, a closed-loop AVS guided by these ROs will always scale voltage to a (safe) value that is higher than what needed by the critical paths. Our experimental results also confirm that through detailed analysis of voltage scaling characteristics, we can design ROs for AVS without any information regarding critical paths or timing performance of a specific design. At the same time, the proposed method could be too pessimistic, and hence we propose circuit design techniques to tune the voltage scaling characteristics of the ROs. We show that the tunability can be used in a scenario where chip frequency is available during ROs characterization. By calibrating the ROs, we can enable up to an additional $30mV$ of supply voltage scaling on a per-instance (per-chip) basis, and up to an average of $13mV$ for a given design. We note that our experiments have been conducted with parameters from a mature ($65nm$) process. The benefit of tunability in the PVS monitors is likely to be larger in less-mature processes which have larger variations around nominal condition. Intuitively, this is because the voltage scaling characteristics vary more in the presence of process variations. Early experiments that we have performed, where we seek monitors to track V_{min_chip} across corners (FF, FS, SF, SS) – that is, extreme process conditions – show that that benefit of tunability increases when process variation is more widely spread.

These ROs can also capture circuit delay degradation due to aging mechanisms (e.g., bias temperature instability and hot carrier injection) if the ROs have the same activity as the circuits being monitored. We can capture the aging effect by connecting the ROs and circuits to the same power rails so that the ROs and the circuits are turned on and off together. Alternatively, more sophisticated aging sensors can be used to quantify the additional voltage margin to guardband for circuit aging [12]. Our ongoing work pursues a proof of concept for application of our proposed methodology to multi-V_{th} design, as well as validation of the methodology across different temperatures. We are also extending the voltage scaling characteristic analysis to include on-chip memory elements.

REFERENCES

[1] K. von Arnim, C. Pacha, K. Hofmann, T. Schulz, K. Schrüfer and J. Berthold, "An Effective Switching Current Methodology to Predict the Performance of Complex Digital Circuits", *Proc. IEDM*, 2007, pp. 483-486.
[2] S. Borkar, T. Karnik, S. Narendra, J. Tschanz, A. Keshavarzi and V. De, "Parameter Variations and Impact on Circuits and Microarchitecture", *Proc. DAC*, 2003, pp. 338-342.
[3] T. D. Burd, T. A. Pering, A. J. Stratakos and R. W. Brodersen, "A Dynamic Voltage Scaled Microprocessor System", *JSSC* 35(11) (2000), pp. 1571-1580.
[4] T.-B. Chan, R. S. Ghaida and P. Gupta, "Electrical Modeling of Lithographic Imperfections", *Intl. Conf. on VLSI Design*, 2010, pp 423-428.
[5] S. Chandra, A. Raghunathan and S. Dey, "Variation-Aware Voltage Level Selection", *IEEE Trans. on VLSI Systems* 20(5) (2012), pp. 925-936.
[6] S. Das, C. Tokunaga, S. Pant, W.-H. Ma, S. Kalaiselvan, K. Lai, D. M. Bull and D. T. Blaauw, "Razor II: In Situ Error Detection and Correlation for PVT and SER Tolerance", *IEEE Journal of Solid-State Circuits* 44(1) (2009), pp. 32-48.
[7] A. Drake, R. Senger, H. Singh, G. Carpenter and N. James, "Dynamic Measurement of Critical-Path Timing", *Proc. IEEE International Conference on Integrated Circuit Design and Technology and Tutorial*, 2008, pp. 249-252.
[8] W. C. Elmore, "The Transient Analysis of Damped Linear Networks with Particular Regard to Wideband Amplifiers", *J. of Applied Physics* 19(1) (1948), pp. 55-63.
[9] M. Elgebaly and M. Sachdev, "Variation-Aware Adaptive Voltage Scaling System", *IEEE Trans. on VLSI Systems* 15(5) (2007), pp. 560-571.
[10] D. Fick, N. Liu, Z. Foo, M. Fojtik, J.-S. Seo, D. Sylvester and D. Blaauw, "In Situ Delay-Slack Monitor for High-Performance Processors Using an All-Digital Self-Calibrating $5ps$ Resolution Time-to-Digital Converter", *Proc. ISSCC*, 2010, pp. 188-189.
[11] K. Kang, S. P. Park, K. Kim and K. Roy, "On-Chip Variability Sensor Using Phase-Locked Loop for Detecting and Correcting Parametric Timing Failures", *IEEE Trans. on VLSI Systems* 18(2) (2010), pp. 270-280.
[12] K. K. Kim, W. Wang and K. Choi, "On-Chip Aging Sensor Circuits for Reliable Nanometer MOSFET Digital Circuits", *IEEE Trans. on Circuits and Systems* 57(10) (2010), pp. 798-802.
[13] K. J. Kuhn, M. D. Giles, D. Becher, P. Kolar, A. Kornfeld, R. Kotlyar, S. T. Ma, A. Maheshwari and S. Mudanai, "Process Technology Variation", *IEEE Trans. on Electron Devices* 58(8) (2011), pp. 2197-2208.
[14] B. Lin, A. Mallik, P. Dinda, G. Memik and R. Dick, "User- and Process-Driven Dynamic Voltage and Frequency Scaling", *Proc. ISPASS*, 2009, pp. 11-22.
[15] Q. Liu and S. S. Sapatnekar, "Capturing Post-Silicon Variations Using a Representative Critical Path", *IEEE Trans. on CAD* 29(2) (2010), pp. 211-222.
[16] N. Mehta and B. Amrutur, "Dynamic Supply and Threshold Voltage Scaling for CMOS Digital Circuits Using In-Situ Power Monitor", *IEEE Trans. on VLSI Systems*, to appear.
[17] L. S. Nielsen, C. Niessen, J. Sparsø and K. V. Berkel, "Low-Power Operation Using Self-Timed Circuits and Adaptive Scaling of the Supply Voltage", *IEEE Trans. on VLSI Systems* 2(4) (1994), pp. 391-397.
[18] A. Raychowdhury, J. Tschanz, K. Bowman, S.-L. Lu, P. Aseron, M. Khellah, B. Geuskens, C. Tokunaga, C. Wilkerson, T. Karnik and V. De, "Error Detection and Correction in Microprocessor Core and Memory Due to Fast Dynamic Voltage Droops", *Journal on Emerging and Selected Topics in Circuits and Systems* 1(3) (2011), pp. 208-217.
[19] K. Shaik, "Implementation of a Critical Path Based Parametric Ring Oscillator", *BSEE Thesis*, Texas Tech University, Texas, 2011.
[20] *Sun OpenSPARC T1 Project.*, http://www.sun.com/processors/opensparc/.
[21] *Synopsys Design Compiler Users Manual.*, http://www.synopsys.com/.
[22] *Synopsys PrimeTime Users Manual.*, http://www.synopsys.com/.
[23] *Synopsys Hspice Users Manual.*, http://www.synopsys.com/.
[24] *Texas Instruments PowerWise.*, http://www.ti.com/ww/en/analog/power_management/powerwise-avs.shtml.