The Impact of BTI Variations on Timing in Digital Logic Circuits

Jianxin Fang and Sachin S. Sapatnekar, *Fellow, IEEE*

***Abstract*—A new framework for analyzing the impact of bias temperature instability (BTI) variations on timing in large-scale digital logic circuits is proposed in this paper. This approach incorporates both the reaction-diffusion model and the charge trapping model for BTI, and embeds these into a temporal statistical static timing analysis (T-SSTA) framework capturing process variations and path correlations. Experimental results on 32nm, 22nm and 16nm technology models, verified through Monte Carlo simulation, confirm that the proposed approach is fast, accurate and scalable, and indicate that BTI variations make a significant contribution to circuit-level timing variations.**

***Index Terms*—Bias Temperature Instability, Circuit Reliability, Process Variation, Timing Analysis**

I. Introduction

Reliability issues in very large scale integrated (VLSI) circuits have been a growing concern as technology trends in semiconductor technologies show progressive downscaling of feature sizes. One of the major reliability issues is bias temperature instability (BTI), which causes the threshold voltage, V_{th}, of CMOS transistors to increase over time under voltage stress, resulting in a temporally-dependent degradation of digital logic circuit delay. Various optimizations have been proposed to cope with this degradation, such as slowing the operating frequency with time, adding delay guardbands, and using adaptive methods to recover from delay degradation.

The reaction-diffusion (R-D) model [1]–[3], based on dissociation of Si–H bonds at the Si/SiO_2 interface, has been the prevailing theory of BTI mechanism and has been widely used in research on circuit optimization and design automation. However, over the years, several limitations in the theory have been exposed. For instance, in R-D theory, the rate of recovery is determined by the diffusion of neutral hydrogen atoms, which is not affected by the gate bias. However the measured device recovery begins faster and lasts longer than the prediction of R-D theory, and shows strong dependence on the applied gate bias. An alternative mechanism for explaining BTI effects is the charge trapping and detrapping model [4], in which the defects in gate dielectrics can capture charged carriers, resulting in V_{th} degradations. The major difference between the two models is the nature of the diffusing species and the medium that facilitates the diffusion. Based on published works, both R-D and charge trapping mechanisms exist in current semiconductor technologies, and the superposition of both models is shown to better match experimental device data [3].

In nanometer-scale technologies, variations in the BTI effect are gaining a great deal of attention under both R-D and charge trapping frameworks, due to the random nature of defect localization in smaller and smaller transistors; together, these result in increased variations in the number of defects in a transistor. While there has been a great deal of research on timing variability due to process variations [5]–[7], and a few previous works have combined random variation effects from process variations with deterministic BTI degradations [8]–[10], the problem of BTI variations has not received much attention.

Most of the published circuit-level works incorporating BTI variations are based on the variability model of ΔN_{IT} randomness within the R-D framework, introduced by [11]. This model was applied to analytically determine the effect of BTI variations on SRAM and logic cells, and on circuit and pipeline performance using Monte Carlo simulations in [12], [13]. However, as explored in our paper, for digital logic circuits, the ΔN_{IT} variation in this R-D based model has a relatively small impact on circuit timing variation, as compared with variations under the charge trapping model and process variations. Another model of the BTI-related variations was considered in [9], as caused by process perturbations. Since these small model perturbations lead to a relatively small change in the BTI-driven delay shift, the impact on circuit timing is a second-order perturbation that is relatively small.

On the other hand, the variations of device-level BTI degradations under the charge trapping model has been discovered to be a significant issue for nanoscale transistors. Charge trapping and detrapping at each defect are random events that are characterized by the capture and emission time constants. This paradigm is intrinsically statistical and it captures not only the variations in the number of defects, but also the variations in ΔV_{th} induced by each defect [14]–[16]. Under this statistical model, the variation of device lifetime increases significantly, especially for devices with a smaller number of defects N, as illustrated in Fig. 1.

However, the impact of BTI variations under the charge trapping model on circuit performance has not received much attention, with only limited works that explore this issue beyond the device level. In [17], models and approaches were proposed for analyzing the impact of BTI variations on circuit performance; however the proposed SPICE-based atomistic approaches are time-consuming and not scalable to

Manuscript received June 2, 2012; revised September 14, 2012. This work was supported in part by the SRC under contract 2012-TJ-2234 and the NSF under award CCF-1017778 and CCF-1162267.

J. Fang and S. S. Sapatnekar are with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455, USA (e-mail: fang0116@umn.edu, sachin@umn.edu).

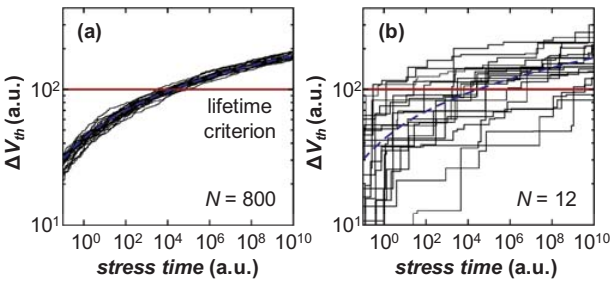


Fig. 1. [17]: (a) Narrow distribution of lifetime in large devices where randomness averages out; (b) Large variation of lifetime in small devices where stochasticity predominates.

large-scale circuits. As we will show, the charge trapping model is the dominant component of BTI intrinsic variations, and makes significant contributions to circuit delay variation. Furthermore its impact grows rapidly as devices scale down, posing increasingly severe reliability issues to digital logic circuits.

In this paper, we first introduce the notion of precharacterized *mean defect occupancy probability* for the charge trapping model to effectively reduce the complexity of circuit-level analysis and to make it possible to handle large-scale circuits. Then we incorporate variations under both the R-D model and the charge trapping model into a novel temporal statistical static timing analysis (T-SSTA) framework, capturing randomness from both process variations and temporal BTI degradations. We exercise this approach on large digital logic circuits and show simulation results for the 32nm, 22nm, and 16nm technology nodes. The correlation of process parameters due to path reconvergence is considered efficiently in modeling and analysis to guarantee both high accuracy and low complexity. To the best of our knowledge, this is the first circuit-level work that incorporates variations in BTI effects into SSTA under a scalable and computationally efficient procedure.

Our experimental results are based on simulations, and show that the proposed analysis approach has an accuracy that lies within 2.2% of Monte Carlo simulation while speeding up the calculation by 15×. Averaging over all benchmarks considered in our work, the fraction of the variance attributable to process variations, BTI R-D effects, and BTI charge trapping effects is, respectively, 81%, 3%, and 16% at the 32nm node, 70%, 4%, and 26% at the 22nm node, and 66%, 5%, and 29% at the 16nm node. Thus, under these models, the relative role of BTI charge trapping to circuit variability is projected to increase significantly in the future, but is less than the contribution of process variations.

II. Modeling Variations

This section introduces the models used to capture the effects of variations that affect BTI-induced aging. We begin by discussing BTI variations under both the R-D and charge trapping models. Next, we overview models for process variations, including spatial correlation effects. As in [3], the total threshold degradation ΔV_{th} of an MOS device is modeled by superposition as

$$\Delta V_{\text{th}} = \Delta V_{\text{th-RD}} + \Delta V_{\text{th-CT}} + \Delta V_{\text{th-RDF}}, \tag{1}$$

in which the BTI terms $\Delta V_{\text{th-RD}}$ and $\Delta V_{\text{th-CT}}$ are independent Gaussian random variables that will be given in (5) and (15), and $\Delta V_{\text{th-RDF}}$ is the variation component due to random dopant fluctuation (RDF) [18], [19], which also follows Gaussian, and have no spatial correlations [20]. For each transistor, the sum, ΔV_{th}, of Gaussian variables is still a Gaussian, and this sum is an independent random variable for different MOS transistors.

A. BTI Variability under the R-D Model

Under the R-D framework, the mechanism of BTI in a MOS transistor is explained through the dissociation of Si–H bonds at the Si/SiO_2 interface and the diffusion of hydrogen into dielectric and gate. The number of generated interface traps, Si^+, is denoted as ΔN_{IT}, and absolute value of the induced threshold voltage shift, ΔV_{th}, is

$$\Delta V_{\text{th}} = \frac{q\Delta N_{\text{IT}}}{C_{\text{ox}}} \tag{2}$$

Under the R-D model, the long term threshold voltage shift ΔV_{th} under AC BTI stress is modeled in [12], [21] as

$$\Delta V_{\text{th}}^{(\text{nom})} = f_{\text{AC}}(SP) \cdot K_{\text{DC}} \cdot t^n \tag{3}$$

in which K_{DC} is a technology dependent constant for DC BTI degradation, and $f_{\text{AC}}(SP)$ is the coefficient that captures the AC degradation with signal probability SP (the probability of effective BTI stress). The function $f_{\text{AC}}(SP)$ can be precomputed numerically using method proposed in [2].

For deeply scaled technologies, the device size is small enough that ΔN_{IT} is a random variable, modeled as a Poisson distribution [11]:

$$\begin{aligned} \Delta N_{\text{IT}} &\sim \text{Poisson}(\lambda), \\ \text{where } \lambda &= \Delta N_{\text{IT}}^{(\text{nom})} = \Delta V_{\text{th}}^{(\text{nom})} \cdot C_{\text{ox}}/q \end{aligned} \tag{4}$$

Our reliability analysis focuses on late lifetime behavior, when the average numbers of interface traps λ in MOS transistors have relatively large values. For instance, the value of λ corresponding to $\Delta V_{\text{th}} = 0.1$V for a device with $\frac{W}{L} = 2$ is about 49 for 32nm PTM [22] model, or 15 for 16nm PTM model, and it increases proportionally with the device size. It is well-known that for $\lambda > 10$, a Gaussian approximates the Poisson distribution well [23]. Therefore, to simplify our analysis without significant loss of accuracy, this Poisson distribution is approximated as a Gaussian distribution with the same mean and variance $\mu = \sigma^2 = \lambda$, hence $\Delta N_{\text{IT}} \sim N(\lambda, \lambda)$. From (2), the threshold voltage degradation under R-D model has the distribution

$$\Delta V_{\text{th-RD}} \sim N\left(\frac{q\lambda}{C_{\text{ox}}}, \frac{q^2\lambda}{C_{\text{ox}}^2}\right) \tag{5}$$

As will be shown in Sec IV, this distribution approximation does not induce significant errors to the circuit level results.

B. BTI Variability under the Charge Trapping Model

Recent work [4] on the BTI effect of small-area devices reveals that the degradation and recovery of ΔV_{th} proceed in discrete steps, with variable heights, which could not be explained by the R-D model, but are fully consistent with charge trapping, which is also observed in random telegraph noise (RTN) and $1/f^2$ noise.

Based on these observations, a newer charge-trapping model was proposed for the BTI effect, in which each defect is characterized by parameters of the capture time τ_c and emission time τ_e, and each defect's contribution to the device threshold change, ΔV_t. These parameters are characterized using the time-dependent defect spectroscopy (TDDS) method [4], [14], as a distribution shown in the form of a density map as Fig. 2, in which defects with similar time constants are binned together, and the total ΔV_t is shown in each grid.

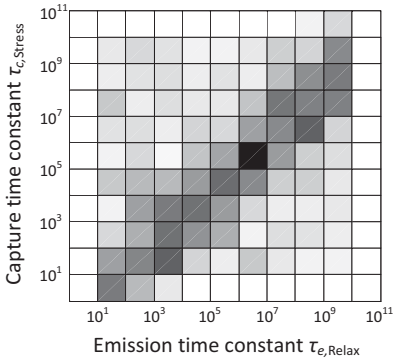


Fig. 2. The distribution of defects according to their capture time constant τ_c and emission time constant τ_e.

If this characterization is performed on a large enough device, with the assumption that ΔV_t of all defects are independent and identically distributed (i.i.d.), the density map could be interpreted as the distribution of defects, in which each grid's value represents the probability of defects falling into that grid. The generation of this distribution is part of technology process characterization and independent of circuit structure.

Charge trapping (capture) and detrapping (emission) is a stochastic process. Following the models in [16], the capture time, τ_c, and the emission time, τ_e, are strongly dependent on bias voltage and temperature. In digital circuits there are only two nontransient voltage stages, logic 1 and logic 0, hence the bias condition can be simplified to two static modes of stress and relaxation. We capture the temperature dependence effect by the use of a standard corner-based approach where the worst-case temperature corner is assumed. In this way each defect can be described by four time constants, denoted by the vector $\vec{\tau}$ as

$$\vec{\tau} = (\tau_{c,\text{Stress}}, \tau_{c,\text{Relax}}, \tau_{e,\text{Stress}}, \tau_{e,\text{Relax}}). \tag{6}$$

The defect occupancy probability (i.e., the probability of charge trapping) of a single defect with time constants $\vec{\tau}$ under AC stress of duty factor DF and time span t is derived in [16] to be:

$$P_c(DF, t, \vec{\tau}) = \frac{\tau_e^*}{\tau_c^* + \tau_e^*}\left(1 - \exp\left(-\left(\frac{1}{\tau_c^*} + \frac{1}{\tau_e^*}\right)t\right)\right), \tag{7}$$

Here the duty factor DF of a device under AC stress is defined as the probability of the transistor in accumulation mode that is effective for BTI stress (in some papers, DF is also referred to as the signal probability SP). The parameters τ_c^* and τ_e^* are defined as the effective capture and emission time constants under AC stress, which account for the duty factor effect:

$$\frac{1}{\tau_c^*} = \frac{DF}{\tau_{c,\text{Stress}}} + \frac{1-DF}{\tau_{c,\text{Relax}}} \tag{8}$$

$$\frac{1}{\tau_e^*} = \frac{DF}{\tau_{e,\text{Stress}}} + \frac{1-DF}{\tau_{e,\text{Relax}}} \tag{9}$$

Fig. 3 shows an example plot of the occupancy probability function, $P_c(DF, t, \vec{\tau})$, of a single defect as defined in (7), with the values of the time constants shown in the figure. The plot indicates that the occupancy probability P_c increases gradually with DF, but rises rapidly with time at the range of 10^5 to 10^6 a.u..

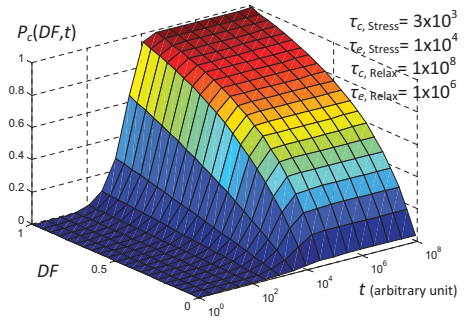


Fig. 3. An example plot of defect occupancy probability function $P_c(DF, t, \vec{\tau})$ of a single defect.

Since the defect precursors (Si−Si bond in the SiO_2 dielectric according to [24]) are created in the fabrication process and uniformly distributed in the dielectric layer, the statistical distribution of capture/emission time constants associated with each defect is i.i.d. For each defect, the four components of $\vec{\tau}$ are correlated [4], and their joint distribution can be characterized for a specific technology. In this paper, we follow the assumptions in [17] to generate the distributions of time constants. Fig. 2 shows an example 2-D histogram of the joint distribution of $\tau_{c,\text{Stress}}$ and $\tau_{e,\text{Relax}}$, which are the dominant components of $\vec{\tau}$. The proposed approaches in this paper are independent of the distribution of $\vec{\tau}$.

We introduce the concept of the *mean defect occupancy probability*, $\bar{P}_c(DF, t)$, which captures the expected value of the probability of a defect charged with carriers (i.e., captured), based on the single defect model of (7), and $f(\vec{\tau})$, the joint pdf of $\vec{\tau}$:

$$\bar{P}_c(DF, t) = \int P_c(DF, t, \vec{\tau}) f(\vec{\tau}) d\vec{\tau} \tag{10}$$

Fig. 4 shows an example of $\bar{P}_c(DF, t)$ function corresponding to the assumed $f(\vec{\tau})$ plotted in Fig. 2. This plot indicates that the mean occupancy probability is a monotonically increasing

function of both DF and time. Due to averaging effects over large number of defects with different $\vec{\tau}$, $\bar{P}_c(DF,t)$ changes more gradually with time, compared with P_c of a single defect in Fig. 3.

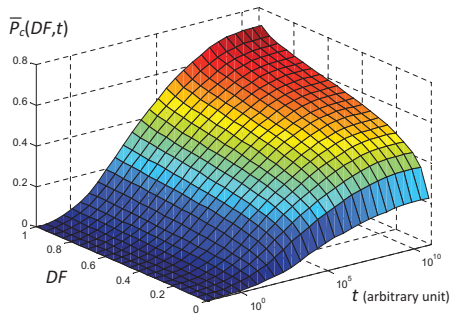


Fig. 4. The plot of mean defect occupancy probability function $\bar{P}_c(DF,t)$.

Since $\bar{P}_c(DF,t)$ is only determined by the distribution $f(\vec{\tau})$ and is independent of the circuit structure, it can be pre-characterized numerically using (10) and stored in a look-up table (LUT) for use in the circuit analysis.

For small-geometry devices, the number of defects in a MOS transistor is a relatively small number with relatively large variation [14]. For a transistor of length L and width W, the total number of oxide defects n is empirically modeled as a Poisson distribution [11]:

$$\begin{aligned} n &\sim \text{Poisson}(N), \\ \text{where } N &= N_{ot}WL. \end{aligned} \tag{11}$$

Here N_{ot} is the density of defects in the dielectric, and N is the total number of defects in the MOS transistor. Note that the Poisson distribution in (11) has similar form as the R-D model (4), but they are from different underlying mechanisms: R-D is modeled with interface traps (Si−H bond), while T-D is modeled with bulk oxide traps (missing oxygen atom in Si−O−Si bond [24]). Both kinds of traps are modeled as Poisson distributions due to the random location of the traps in small devices, however these two distributions are not correlated in nature.

Similarly, the number of *occupied* defects, n_c, in a transistor also has a Poisson distribution[1], with its mean value N_c calculated as follows.

$$\begin{aligned} n_c &\sim \text{Poisson}(N_c), \\ \text{where } N_c &= N \cdot \bar{P}_c(DF,t) \end{aligned} \tag{12}$$

Observed in [15], the BTI-induced threshold degradation corresponding to each single defect follows an exponential distribution. Each defect $k = 1, ..., n$, contributes a threshold degradation of:

$$\begin{aligned} \Delta V_{\text{th}}^{(k)} &\sim \text{Exp}(\eta), \\ \text{where } \eta &= \eta_0/(WL). \end{aligned} \tag{13}$$

[1]The number of occupied defects in a device follows a Poisson distribution by definition because (a) each occupied defect has the same occurrence rate $N_c/(WL)$ within the device area of W by L, and (b) the occurrence of all occupied defects are independent with each other. This is similar to the number of all defects which follows $n \sim \text{Poisson}(N)$, and is verified by experimental results in Sec IV.

Like N_{ot}, η_0 is a technology-specific constant.

The total threshold voltage degradation, ΔV_{th}, of a transistor is the sum of contributions $\Delta V_{\text{th}}^{(k)}$ from all *occupied* defects k in the transistor, i.e.,

$$\Delta V_{\text{th}} = \sum_{k=1}^{n_c} \Delta V_{\text{th}}^{(k)}. \tag{14}$$

A closed form of this sum is derived in [15], and the PDF of ΔV_{th} turns out be to a complex distribution with mean $\mu = N_c\eta$ and variance $\sigma^2 = 2N_c\eta^2$. The mean value corresponds to the nominal case (i.e., each of N_c defects results in a threshold degradation of η). In [15], the probit plot of ΔV_{th} indicates that for an adequate number of defects (e.g., $N_c \geq 10$), the transistor ΔV_{th} distribution can be approximated as a Gaussian by matching the mean and variance, resulting in the distribution:

$$\Delta V_{\text{th-CT}} \sim N(N_c\eta, 2N_c\eta^2) \tag{15}$$

When the number of occupied defects, N_c, is sufficiently large, this Gaussian approximation is justified by central limit theorem (CLT), using the fact that the total threshold degradation is the sum of ΔV_{th} from each defect, which are i.i.d. exponential. For smaller devices with lower values of N_c, this Gaussian approximation is not necessarily accurate for individual devices, but the circuit level timing analysis results still have good accuracy compared with Monte Carlo simulation, which can be justified by the central limit theorem (CLT) because the circuit delay is the sum of the cell delays along the critical paths and approaches a Gaussian distribution. A more detailed discussion about this Gaussian approximation model is available in Sec IV.

C. Process Variations and Spatial Correlation

Variations in the process parameters also contribute to BTI variability. Process variations are typically classified as lot-to-lot, die-to-die (D2D), and within-die (WID) variations, according to their scope; they can also be categorized, based on their causes and predictability, as systematic or random variations. Some (but not all) WID variations exhibit spatial dependence knows as spatial correlation, which must be considered for accurate circuit analysis.

We employ a widely-used variational paradigm, where a process parameter X is modeled as a random variable about its mean, X_0, as [25]:

$$\begin{aligned} X &= X_0 + \Delta X \\ \Delta X &= X_g + X_s + X_r \\ \sigma_X^2 &= \sigma_{X_g}^2 + \sigma_{X_s}^2 + \sigma_{X_r}^2 \end{aligned} \tag{16}$$

Here, X_g, X_s, and X_r stand for, respectively, the global component (from lot-to-lot or D2D variations), the spatially correlated component (from WID variation), and the residual random component of process variations. Under this model, all devices on the same die have the same global part X_g. The spatially correlated part is modeled using a widely-used grid-based method [5] for the parameters that exhibit this property, and is zero for those that are spatially uncorrelated. Under

the spatial correlation model, the entire chip is divided into grids. All devices within the same grid have the same spatially correlated part X_s; the X_s parameters for devices in different grids are correlated, with the correlation falling off with the distance. The random part X_r is unique to each device in the system.

In this paper we consider the variations in the transistor width (W), the channel length (L), the oxide thickness (T_{ox}), as well as shifts in the threshold voltage V_{th} due to random dopant fluctuations (RDFs). In other words, for each device, X represents elements of the set $\{W, L, T_{\text{ox}}, V_{\text{th}}\}$. As in the large body of work on SSTA, we assume Gaussian-distributed parameters for each of these process parameters, with W and L exhibiting spatial correlation, and T_{ox} and V_{th} being uncorrelated from one device to the next. The spatial correlation structure is extracted as a correlation matrix [26], and processed using principal components analysis (PCA) to facilitate fast timing analysis [5]. The process parameter value in each grid is expressed as a linear combination of the independent principal components, with potentially reduced dimension.

Notationally, we express each process parameter X as a vector in a random space, with basis $\mathbf{e} = [\mathbf{e}_g, \mathbf{e}_s, \mathbf{e}_r, \epsilon]^{\mathbf{T}}$, as

$$\begin{aligned} X &= X_0 + \Delta X = X_0 + \mathbf{k}_X^{\mathbf{T}}\mathbf{e} \\ &= X_0 + \mathbf{k}_{Xg}^{\mathbf{T}}\mathbf{e}_g + \mathbf{k}_{Xs}^{\mathbf{T}}\mathbf{e}_s + \mathbf{k}_r^{\mathbf{T}}\mathbf{e}_r + k_\epsilon\epsilon \qquad (17) \\ \sigma_X^2 &= \mathbf{k}_X^{\mathbf{T}}\mathbf{k}_X, \quad \text{cov}(X_i, X_j) = \mathbf{k}_{Xi}^{\mathbf{T}}\mathbf{k}_{Xj} - k_{\epsilon_i}k_{\epsilon_j} \qquad (18) \end{aligned}$$

Here, $\mathbf{e}_g = [e_{Wg}, e_{Lg}]^{\mathbf{T}}$ is the basis for the global part (T_{ox} variation and RDF effect are purely random hence do not have a global part), $\mathbf{e}_s = [e_1, ..., e_t]^{\mathbf{T}}$ is the basis of principal components for the spatially correlated part, in which t is the number of dimensions after the PCA processing of correlated part, and $\mathbf{e}_r = [\epsilon_1, ..., \epsilon_m]^{\mathbf{T}}$ is the basis of random part. The dimension of random part, m, will depend on the implementation of the SSTA algorithm, and can vary from constant to linear (of circuit size), as will be shown later in this paper. The random basis $\mathbf{e}_r$ and its coefficient vector $\mathbf{k}_r$ are implemented using a sparse data structure. The Gaussian variable $\epsilon \sim N(0, 1)$ is a separate independent random part for use in circuit-level timing analysis.

D. Consideration of Process Variations and BTI Interaction

Process variations are created at manufacture time, while BTI degradation occurs during the circuit operation. Therefore the effect of process variations is independent from BTI, but the BTI effect will be impacted by process variations, i.e., the BTI degradation is dependent on the actual W, L and T_{ox} of a transistor. This paper assumes the process variations and BTI effects (both R-D and charge trapping model including variabilities) to be independent and uses a superposition model to calculate the total effect on circuit-level degradations. This is based on the following facts and considerations.

- The impact of process variations on BTI degradation is a second order effect that is relatively small in nature.
- For W and L variations, [27] indicates the NBTI effect is more pronounced in narrow and long transistors. However the transistors on the critical paths are normally sized larger (wider) for timing performance, hence less affected by the W and L variations.
- For T_{ox} variation, a smaller T_{ox} causes elevated BTI degradation speed, but also gives smaller initial V_{th} and propagation delay. Therefore the interaction effect actually cancels out with each other to some degree, and ignoring it yields pessimistic and safe approximations.
- The independent assumption simplifies the modeling and analysis and helps achieve linear computational complexity and good scalability (Section III-C).

III. Timing Analysis under Variations

This section introduces the logic cell delay model under BTI variations and process variations. Based on this model, a scalable approach for statistical timing analysis of large digital logic circuits is outlined.

A. Cell Timing Model and Characterization

We use a cell delay degradation model that is similar to [28]. The delay d_i of cell i is modeled using a first-order Taylor approximation, as a linear function of process parameters W_j, L_j and $T_{\text{ox-}j}$ of each transistor j in cell i, and BTI degradation $\Delta V_{\text{th}}^{(j)}$ of each transistor j:

$$d_i = d_{i0} + \Delta d_i = d_{i0} + \sum_{X \in \mathbf{P}_i} \frac{\partial d_i}{\partial X} \Delta X$$

Here $\mathbf{P}_i = \{W_j, L_j, T_{\text{ox-}j}, V_{\text{th}}^{(j)}\}, j \in$ cell i is the set of variational parameters. The nominal propagation delay d_{i0} and its sensitivity $\partial d_i / \partial X$ to parameter $X \in \mathbf{P}_i$ are computed using standard techniques through SPICE simulations. This part of the calculation is circuit-independent and performed as part of library characterization.

Since all variational parameters $X \in \mathbf{P}_i$ are expressed as vectors in the random variable space $\mathbf{e}$ in Section II-C, d_i, which is a linear combination of these parameters, is also a vector in space $\mathbf{e}$:

$$\begin{aligned} d_i &= d_{i0} + \left(\sum_{X \in \mathbf{P}_i} \frac{\partial d_i}{\partial X} \mathbf{k}_X \right)^{\mathbf{T}} \mathbf{e} \\ &= d_{i0} + \mathbf{k}_{d_g}^{\mathbf{T}}\mathbf{e}_g + \mathbf{k}_{d_s}^{\mathbf{T}}\mathbf{e}_s + \mathbf{k}_{d_r}^{\mathbf{T}}\mathbf{e}_r \qquad (19) \end{aligned}$$

Here the random part $\mathbf{e}_r = \{\epsilon_X\}_{X \in \mathbf{P}_i}$ is extended to include the random parts from all variational parameters $X \in \mathbf{P}_i$ in cell i.

B. Circuit Level Timing Analysis

At the circuit level, timing analysis is performed using a PERT-like traversal [5] at a fixed time point, where the contributions of the temporal BTI variations can be characterized using the models described in Sections II-A and II-B. The V_{th} degradation due to these two models are uncorrelated, and are found to substantially affect the circuit level delay.

In our initial implementation of algorithm, as in [7], the random part $\mathbf{k}_r^T \mathbf{e}_r$ of arrival time is merged into the separate independent term $k_\epsilon \epsilon$ that is the product of scalars to reduce

the computational complexity. The temporal statistical static timing analysis (T-SSTA) result of this method is denoted as T-SSTA1. Table I shows the mean and standard deviation of circuit delay degradation on benchmark c3540 under a 16nm technology model at t=2000 (a.u.), splitting the contribution of the mean and variance into those attributable to BTI R-D, BTI charge trapping (CT), process variations (PV), and finally presenting the combined values (ALL). The results of mean and standard deviation calculated by Monte Carlo (MC) simulation are listed as reference, and the results indicate that the mean value of delay degradation is mainly contributed by BTI RD and BTI CT, while the standard deviation is mainly contributed by BTI CT and PV effects. By comparing results for "ALL" from T-SSTA1 method with MC, we can see T-SSTA1 overestimates the mean value and underestimates the standard deviation, where the errors are mainly coming from the BTI CT part.

TABLE I
T-SSTA RESULTS UNDER VARIATIONS (TIME UNIT: PS)

c3540 16nm D_0=582.3	MC		T-SSTA1		T-SSTA2		T-SSTA3	
	$\mu_{\Delta D}$	$\sigma_{\Delta D}$	$\mu_{\Delta D}$	$\sigma_{\Delta D}$	$\mu_{\Delta D}$	$\sigma_{\Delta D}$	$\mu_{\Delta D}$	$\sigma_{\Delta D}$
BTI RD	23.8	3.6	23.8	3.7	23.8	3.7	23.8	3.7
BTI CT	29.8	8.9	29.7	8.8	29.6	8.8	29.6	8.8
PV	0.5	14.8	4.2	14.1	0.3	14.6	1.3	14.6
ALL	53.9	17.2	57.1	17.0	54.0	17.5	54.8	17.4

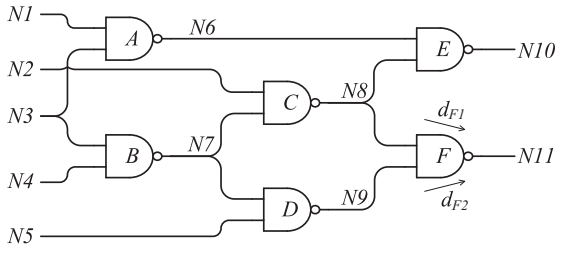


Fig. 5. An example circuit showing path reconvergence.

Investigating this further, we find that the error between conventional method (T-SSTA1) and Monte Carlo (MC) simulation can be attributed to the correlations that arise due to path reconvergence, which are neglected in T-SSTA1. The BTI CT part of V_{th} degradation contains a significant amount of independent random component in the form (17), hence generates large errors due to path reconvergence. We illustrate the path reconvergence effect through Fig. 5, which shows an example circuit where the arrival time AT of node $N11$, denoted as AT_{N11}, is calculated as follows

$$AT_{N11} = \max\left(AT_{N8} + d_{F1}, AT_{N9} + d_{F2}\right) \tag{20}$$

where AT_{N8} and AT_{N9} stand for the arrival times of node $N8$ and $N9$, while d_{F1} and d_{F2} stand for the delays from the first and second input to the output of cell F. The arrival times and cell delays are modeled as vectors in random variable space $\mathbf{e}$. Note that since cell C and D have a common fanin of cell B, AT_{N8} and AT_{N9} are both dependent on the random component of the parameters of cell B, corresponding to the impact of X_r in Equation (16). As a result, the independent components in the expression for AT_{N8} and AT_{N9} are not independent of each other, but are correlated. However the conventional SSTA method, using an separate independent term $k_\epsilon \epsilon$ to replace $\mathbf{k}_r \mathbf{e}_r$, does not capture this path correlation and introduces errors. The same situation occurs when calculating the total delay using the maximum of AT_{N10} and AT_{N11}, which are correlated because the paths from node $N8$ reconverge.

One natural way to resolve this problem is to preserve the entire random part $\mathbf{k}_r \mathbf{e}_r$ when calculating the arrival times, by which the path correlation is captured. This method is denoted as T-SSTA2 in Table I, and the results indicate this method is much more accurate than T-SSTA1. However, the cost paid for this accuracy is in the increased computation time associated with the growing size of the random part (e.g., AT_{N11} in the example contains components from cells B, C, D, and F). The computational complexity is discussed with more details in Section III-C and the experimental runtime and storage comparison will be given in Section IV.

We employ a third method, denoted as T-SSTA3 in Table I, taken from [29], to provide a trade off between the accuracy and complexity. This removes only the smaller elements in the random vector $\mathbf{k}_r$ using preset threshold and merges them into the separate term k_ϵ. Results in Table I show that this method achieves good accuracy (within 2% error compared with T-SSTA2 and Monte Carlo) with low computation. We will expand on this in Section IV.

C. Computational Complexity

To calculate the circuit delay, the SSTA algorithm does a topological traversal through the digital circuit. For each node (logic cell), the timing analysis performs k sum-of-two and $k-1$ max-of-two operations, where k is the number of fan-in of the cell. In random space $\mathbf{e}$ with dimension d, the numbers of total sum and max operations for SSTA are

$$N_{sum} = n \cdot k \cdot d \tag{21}$$

$$N_{max} = n \cdot (k-1) \cdot d \tag{22}$$

Here $d = d_g + d_s + d_r$, in which d_g, d_s and d_r are the dimension of global component $\mathbf{e_g}$, spatial component $\mathbf{e_s}$ and random component $\mathbf{e_r}$, respectively. The values of d_g and d_s are well bounded by PCA algorithm therefore can be regarded as constant. The values of d_r depends on how the random part is handled as discussed in Section III-B. For methods T-SSTA1 and T-SSTA3 d_r is bounded by a constant, while for T-SSTA2 d_r can grow significantly depending on the circuit size and structure. For simplicity it can be roughly approximated as $d_r \propto \sqrt{n}$, which corresponds to the depth of the circuit (number of cells on the critical paths). Therefore the computational complexity is $O(n)$ for T-SSTA1 and T-SSTA3, and $O(n^{1.5})$ for T-SSTA2. This result indicates the proposed T-SSTA3 method has good scalability to handle large scale circuits.

IV. RESULTS

Our approach for timing analysis under BTI variations and process variations is applied to ISCAS85 and ITC99

benchmarks. The benchmark circuits are mapped to a subset of the Nangate cell library [30] using ABC [31], with placement carried out using a simulated annealing algorithm. The benchmark circuits are scaled down to 32nm, 22nm and 16nm for comparisons under different technology models. The characterization of cell delay and of its sensitivities to variational parameters is performed using HSPICE simulation under PTM models [22]. Both the proposed analytical method and the Monte Carlo method (for verification) are implemented in C++ and run on a Linux PC (3GHz CPU, 2GB RAM).

The process variations in W, L, and T_{ox} are set to $3\sigma = 4\%$ of their mean values [32]. The V_{th} variation due to RDF is dependent on the device size [20]. It has a Gaussian distribution with mean value $\mu = 0$, and standard variation

$$\sigma_{V_{\text{th}}} = \sigma_{V_{\text{th0}}} \sqrt{\frac{W_0 L_0}{WL}} \tag{23}$$

in which $\sigma_{V_{\text{th0}}}$ is the RDF-induce threshold standard deviation of a minimum-sized device (W_0 by L_0). The value of $\sigma_{V_{\text{th0}}}$ is dependent on process parameters T_{ox} and N_a, as well as the doping profile of the channel [20]. Here we assume $3\sigma_{V_{\text{th0}}} = 5\%$ of the nominal V_{th}. The parameter variations of W and L are split into 20% of global variation, 20% of spatially correlated variation and 60% of random variation, while the variations of T_{ox} and V_{th} are fully random. The grid-based spatial correlation matrix is generated using the distance based method in [26], with the number of grids growing with circuit size, as shown in the Table III.

The Monte Carlo simulation framework for verification of the proposed approach is set up as follows: the simulation program randomly generates 5000 circuit instances (we found it a good trade-off of accuracy and runtime). For each circuit instance, the ΔV_{th} of each MOS transistor is calculated as the sum of the following three components:

(a) $\Delta V_{\text{th-RD}}$, which is set by (5) and is randomly generated based on the distribution of ΔN_{IT} as specified in (4),
(b) $\Delta V_{\text{th-CT}}$, which is set as the sum of ΔV_{th} of all defects that are randomly generated using distributions (11) and (13), and
(c) $\Delta V_{\text{th-RDF}}$, which is due to RDF effects and set by (23).

The contributions of $\Delta V_{\text{th-RD}}$ and $\Delta V_{\text{th-CT}}$ vary with different technologies [3]. In the experiments it is assumed that these two have comparable mean values, so that their contributions to circuit-level variations can be easily visualized. The process parameters W, L, and T_{ox} of each MOS transistor are also generated according to their distributions and correlation models. Then the propagation delay of each cell is calculated using (19) and pre-characterized cell delay and sensitivity data. Based on these values and a PERT-like traversal, the total delay of the circuit instances is evaluated using statistical static timing analysis (SSTA).

For each benchmark circuit, the mean and standard deviation of the circuit delay are calculated at time t=2000 (a.u.), using the proposed analytical method and Monte Carlo (MC) simulation. The three methods of handling random parts discussed in Section III-B are implemented separately. As before,

- T-SSTA1 merges the random part into one variable,
- T-SSTA2 preserves full random part, and
- T-SSTA3 partially lumps the random part.

Table II shows the nominal delay D_0 of each benchmark circuit, as well as the mean μ and standard deviation (SD), σ, of the circuit delay using three analytical methods and the MC simulation, at 32nm, 22nm, and 16nm. The last row shows the relative error of $\mu_{\Delta D}$ and σ_D of each analytical method, compared with MC.

The mean and SD of the ΔV_{th} contributions (averaged over all devices in the circuit) from the R-D model and from the charge trapping model are also listed in Table II. Note that the simulation is based on the assumption that the ΔV_{th} contributions (mean value) from R-D model and charge trapping model are comparable. This assumption is made to give a general insight that the charge trapping model predicts significantly larger BTI variability than R-D model. The proposed approaches for circuit degradation analysis is actually independent with this assumption and can handle different cases of the BTI degradation model. In general cases of application, both R-D and charge trapping model of BTI effects can be characterized for given technology and used for analyzing the circuit timing degradations.

It is also worth noting that under certain cases (especially at 16nm, under the charge trapping model), the value of 3σ can be larger than μ, indicating Gaussian distribution may not be an accurate approximation of ΔV_{th} since ΔV_{th} from BTI effects should always be positive. However this inaccuracy of the Gaussian approximation is averaged out by the sum of delay along the critical path, and the circuit level delay, calculated by sum and max operations in SSTA, and approaches a Gaussian distribution according to the central limit theorem (CLT), which does not require the transistor ΔV_{th} to be Gaussian. Therefore the proposed method appears to be robust even under this model inaccuracy, as verified by the good accuracy indicated in Table II, and the visually-verified match between distribution functions plotted in Fig. 6, which shows an example of the circuit delay distribution for c3540 at 16nm at t=2000 (a.u.). The T-SSTA3 and MC methods match well, verifying the validity of our assumptions; T-SSTA1 is significantly different, due to the omission of path correlations.

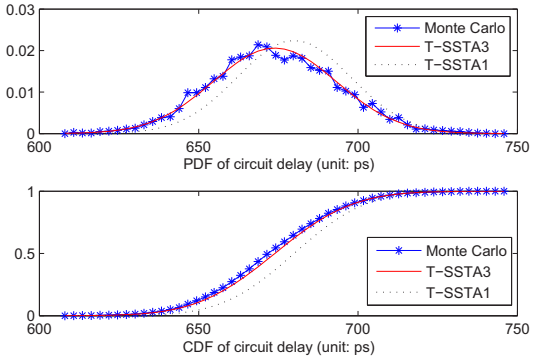


Fig. 6. The delay PDF and CDF of c3540 with 16nm model.

Table III compares the runtime and storage complexity (in

TABLE II
MEAN AND SD OF CIRCUIT DELAY USING DIFFERENT METHODS (TIME UNIT: PS, AVERAGE ERROR SHOWN FOR $\mu_{\Delta D}$ AND σ_D)

Circuit & Technology		Initial	T-SSTA1		T-SSTA2		T-SSTA3		MC		$V_{\text{th-RD}}$(mV)		$V_{\text{th-CT}}$(mV)	
		D_0	μ_D	σ_D	μ_D	σ_D	μ_D	σ_D	μ_D	σ_D	μ	σ	μ	σ
32 nm	c2670	734	802	13.3	795	14.6	796	14.4	794	14.7	17.8	3.3	16.9	6.1
	c3540	812	910	15.6	903	16.7	904	16.7	903	16.5	17.7	3.1	16.6	5.7
	c5315	666	736	12.9	735	13.1	735	13.1	735	12.9	18.1	3.2	17.5	6.0
	c6288	1416	1580	23.5	1574	24.1	1576	24.0	1574	23.6	17.3	2.9	16.5	5.3
	c7552	650	714	11.6	709	12.5	710	12.5	710	12.4	18.2	3.3	17.6	6.2
	b15	1416	1580	24.2	1571	26.2	1574	25.7	1572	25.9	17.0	3.1	16.5	5.6
	b17	1634	1770	27.0	1750	29.1	1757	28.5	1752	28.7	16.6	3.1	16.1	5.5
	b20	1432	1566	24.1	1554	26.8	1555	26.7	1555	26.3	17.8	3.2	17.3	6.0
	b21	1598	1765	27.0	1757	29.9	1758	29.7	1758	30.5	17.8	3.3	17.4	6.1
	b22	1520	1655	23.4	1645	25.3	1646	25.1	1645	25.3	17.4	3.2	17.0	6.0
Avg Err %			7.24	6.19	0.46	1.36	1.22	1.36						
22 nm	c2670	617	669	12.6	663	13.8	664	13.6	663	13.9	18.1	4.6	17.4	9.0
	c3540	671	760	15.1	754	16.5	755	16.4	754	16.7	17.6	4.3	16.7	8.1
	c5315	557	625	11.9	624	12.2	624	12.1	624	12.3	17.7	4.4	17.0	8.4
	c6288	1211	1366	22.8	1359	23.5	1362	23.3	1359	23.6	17.3	4.0	16.6	7.7
	c7552	549	607	10.6	601	12.3	602	12.2	601	12.3	17.9	4.5	17.3	8.8
	b15	1151	1287	23.9	1274	27.4	1275	27.0	1275	26.4	16.8	4.3	16.2	8.0
	b17	1312	1444	23.7	1437	24.4	1443	23.4	1440	25.4	16.5	4.3	16.0	7.9
	b20	1144	1293	23.4	1273	27.8	1274	27.6	1274	28.0	18.1	4.5	17.6	8.9
	b21	1251	1392	25.0	1388	26.9	1388	26.8	1388	26.6	18.0	4.5	17.6	8.9
	b22	1252	1379	23.0	1371	24.7	1372	24.6	1372	24.9	17.7	4.5	17.4	8.8
Avg Err %			7.48	8.45	0.75	1.50	1.30	2.16						
16 nm	c2670	537	592	13.9	582	15.7	584	15.4	582	15.9	18.5	6.2	17.9	12.7
	c3540	582	657	17.5	652	18.1	653	18.0	652	18.9	17.0	5.4	16.1	10.7
	c5315	489	570	14.2	567	14.6	568	14.5	568	15.1	18.2	5.8	17.7	12.0
	c6288	1092	1253	26.2	1247	26.7	1251	26.5	1247	26.6	17.3	5.3	16.6	10.5
	c7552	480	559	12.9	545	16.8	548	16.6	546	16.9	18.2	6.0	17.6	12.4
	b15	986	1143	26.0	1121	31.9	1124	31.2	1125	33.1	16.8	5.6	16.3	11.0
	b17	1100	1318	28.6	1293	33.2	1299	32.8	1293	33.6	16.7	5.6	16.2	11.0
	b20	941	1083	23.6	1075	23.5	1080	23.0	1078	25.2	17.9	5.9	17.4	12.1
	b21	1038	1158	26.7	1148	30.5	1149	29.8	1149	31.1	17.2	5.8	16.7	11.6
	b22	1072	1219	25.8	1206	29.0	1207	28.7	1208	29.1	17.9	6.0	17.5	12.2
Avg Err %			9.97	11.95	0.97	2.39	1.53	3.64						
Total Avg Err			8.23	8.86	0.73	1.75	1.35	2.39						

TABLE III
COMPARISON OF COMPUTATIONAL COMPLEXITIES, WHERE T_{exe}= RUNTIME, [CELLS] = AVERAGE NUMBER OF CORRELATED CELLS.

Circuit Name	Size		T-SSTA1	T-SSTA2		T-SSTA3		MC
	#cells	#grids	T_{exe}	T_{exe}	[Cells]	T_{exe}	[Cells]	T_{exe}
c2670	759	16	3.4s	5.8s	26.2	6.7s	3.0	108s
c3540	1033	16	5.7s	13.5s	109.0	12.8s	2.7	201s
c5315	1699	16	7.2s	14.1s	40.8	15.2s	2.9	261s
c6288	3560	64	17.1s	137.8s	473.6	38.9s	2.8	627s
c7552	2361	36	9.8s	21.1s	53.7	20.3s	3.2	352s
b15	6548	100	34.8s	352.4s	512.1	89.6s	3.0	1181s
b17	20407	361	109.3s	1513s	421.6	306.0s	3.5	3645s
b20	11033	169	55.2s	482.1s	362.3	139.0s	3.4	1926s
b21	10873	169	52.9s	439.1s	351.4	133.5s	2.8	1845s
b22	14794	225	72.4s	671.2s	304.9	188.1s	3.1	2507s

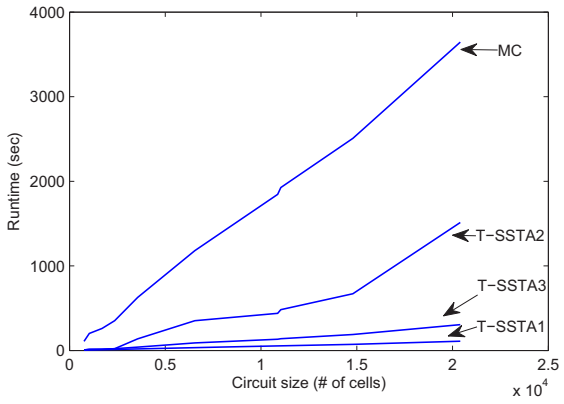


Fig. 7. Runtime vs. circuit size of different methods.

terms of the average number of correlated cells, denoted as [Cells]) of the analytical methods and MC. Fig. 7 shows the runtime vs. circuit size (number of logic cells) for the different methods.

The results indicate that the runtime of partially lumping random part (T-SSTA3) method grows linearly with circuit size increasing, indicating good scalability. It has an overall error of about 2% to MC, and is 15× faster on average. Furthermore, it reduces runtime by 60% and storage by 98% on average compared with T-SSTA2, with similar accuracy. The conventional method (T-SSTA1) has the shortest runtime, but has nearly 9% errors with respect to MC. The results also verify that the Gaussian approximations for ΔV_{th} in BTI R-D and charge trapping models are valid; the method is accurate, efficient, and scalable. Moreover, the standard deviation of circuit delay σ_D increases with technology downscaling, indicating that random timing variation attributable to BTI is a growing issue.

Fig. 8 shows the variance of circuit delay that originates from process variations, R-D BTI variations, and charge trapping BTI variations separately, for different benchmarks under the 32nm, 22nm, and 16nm technology models. For better presentation of data, the variances are normalized to the total variance of 32nm model for each benchmark. These results indicate that the charge trapping model is the dominant component of BTI variations, and makes a significant contribution to circuit delay variation. In contrast, the BTI variations under the R-D model only introduce a relatively small portion

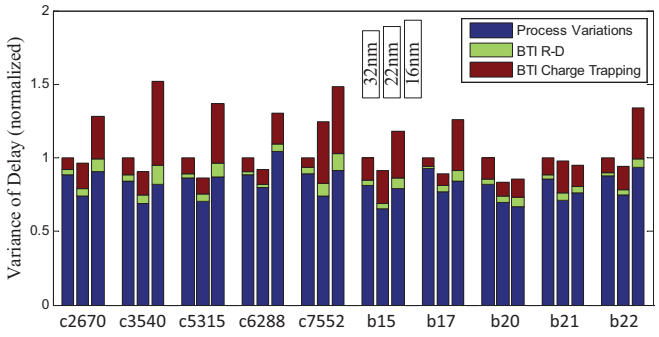


Fig. 8. Relative contribution of BTI charge trapping, BTI R-D, and process variation to circuit delay variation.

of delay variations. Unlike process variations which have nearly constant influence on delay variation, the impact of BTI variations grows with scaling, becoming increasingly severe in future.

Further, according to the results in Table II and Fig. 8, the circuit level delay variation that can be attributed to BTI variations is not as significant as the single-device ΔV_{th} variation due to BTI effect of a small transistor shown in Fig. 1 (b). This is mainly due to the facts that (a) transistors on the critical paths usually have larger than minimum sizes to help with timing, and (b) the average out of randomness of the transistor ΔV_{th} on the critical paths due the sum of delay.

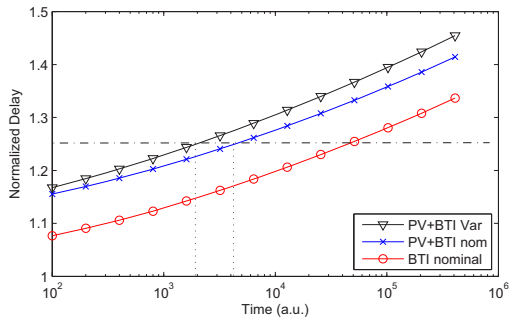


Fig. 9. Delay degradation vs. time of c5315

Fig. 9 presents the circuit delay degradation vs. time curves of benchmark c5315 at 16nm. Three curves are shown for the normalized delay of (1) nominal BTI degradation, without any variation model; (2) $\mu + 3\sigma$ of process variation (PV) and nominal BTI degradation; and (3) $\mu + 3\sigma$ of PV and BTI with variabilities (under both R-D and charge trapping models). The results indicate that BTI degradation and variability, which grow with time, make up the dominant part of total delay degradation, especially at the later point of circuit lifetime. Furthermore, BTI variations has a significant impact on circuit reliability. In this case, the circuit lifetime will be overestimated by over $2\times$ if BTI variations is not considered (lifetime defined as 25% increase of delay from time zero).

V. Conclusion

This paper incorporates both the R-D and charge trapping models of BTI variations into a T-SSTA framework, capturing process variations and path correlations. Experimental results show that the proposed analysis method is fast and accurate. Our results indicate that the charge trapping mechanism, which has been neglected by the EDA field so far, is the dominant source of BTI variations, with significant and growing contributions to circuit timing variations.

References

[1] M. A. Alam, "A critical examination of the mechanics of dynamic NBTI for PMOSFETs," in *Proceedings of the IEEE International Electronic Devices Meeting*, Dec. 2003, pp. 14.4.1–14.4.4.

[2] S. V. Kumar, C. H. Kim, and S. S. Sapatnekar, "A finite-oxide thickness-based analytical model for negative bias temperature instability," *IEEE Transactions on Devices and Materials Reliability*, vol. 9, no. 4, pp. 537–556, Dec. 2009.

[3] S. Mahapatra, A. E. Islam, S. Deora, V. D. Maheta, K. Joshi, A. Jain, and M. A. Alam, "A critical re-evaluation of the usefulness of R-D framework in predicting NBTI stress and recovery," in *Proceedings of the IEEE International Reliability Physics Symposium*, Apr. 2011, pp. 6A.3.1–6A.3.10.

[4] T. Grasser, B. Kaczer, W. Goes, H. Reisinger, T. Aichinger, P. Hehenberger, P.-J. Wagner, F. Schanovsky, J. Franco, P. Roussel, and M. Nelhiebel, "Recent advances in understanding the bias temperature instability," in *Proceedings of the IEEE International Electronic Devices Meeting*, Dec. 2010, pp. 4.4.1–4.4.4.

[5] H. Chang and S. S. Sapatnekar, "Statistical timing analysis considering spatial correlations using a single PERT-like traversal," in *Proceedings of the IEEE/ACM International Conference on Computer-Aided Design*, Nov. 2003, pp. 621–625.

[6] A. Agarwal, D. Blaauw, and V. Zolotov, "Statistical timing analysis for intra-die process variations with spatial correlations," in *Proceedings of the IEEE/ACM International Conference on Computer-Aided Design*, Nov. 2003, pp. 900–907.

[7] C. Visweswariah, K. Ravindran, K. Kalafala, S. G. Walker, S. Narayan, D. K. Beece, J. Piaget, N. Venkateswaran, and J. G. Hemmett, "First-order incremental block-based statistical timing analysis," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 25, no. 10, pp. 2170–2180, Oct. 2006.

[8] W. Wang, V. Balakrishnan, B. Yang, and Y. Cao, "Statistical prediction of NBTI-induced circuit aging," in *Proceedings of the International Conference on Solid State and Integrated Circuits Technology*, Oct. 2008, pp. 416–419.

[9] Y. Lu, L. Shang, H. Zhou, H. Zhu, F. Yang, and X. Zeng, "Statistical reliability analysis under process variation and aging effects," in *Proceedings of the ACM/IEEE Design Automation Conference*, Jul. 2009, pp. 514–519.

[10] S. Han and J. Kim, "NBTI-aware statistical timing analysis framework," in *Proceedings of the IEEE International SoC Conference*, Sep. 2010, pp. 158–163.

[11] S. E. Rauch, "The statistics of NBTI-induced V_T and β mismatch shifts in pMOSFETs," *IEEE Transactions on Devices and Materials Reliability*, vol. 2, no. 4, pp. 89–93, Dec. 2002.

[12] K. Kang, S. P. Park, K. Roy, and M. A. Alam, "Estimation of statistical variation in temporal NBTI degradation and its impact on lifetime circuit performance," in *Proceedings of the IEEE/ACM International Conference on Computer-Aided Design*, Nov. 2007, pp. 730–734.

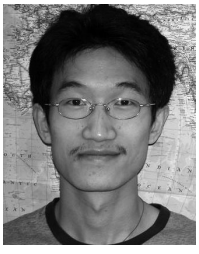
[13] B. Vaidyanathan, A. Oates, and Y. Xie, "Intrinsic NBTI-variability aware statistical pipeline performance assessment and tuning," in *Proceedings of the IEEE/ACM International Conference on Computer-Aided Design*, Nov. 2009, pp. 164–171.

[14] H. Reisinger, T. Grasser, W. Gustin, and C. Schlünder, "The statistical analysis of individual defects constituting NBTI and its implications for modeling DC- and AC-stress," in *Proceedings of the IEEE International Reliability Physics Symposium*, May 2010, pp. 7–15.

[15] B. Kaczer, T. Grasser, P. J. Roussel, J. Franco, R. Degraeve, L. Ragnarsson, E. Simoen, G. Groeseneken, and H. Reisinger, "Origin of NBTI variability in deeply scaled pFETs," in *Proceedings of the IEEE International Reliability Physics Symposium*, May 2010, pp. 26–32.

[16] M. Toledano-Luque, B. Kaczer, P. J. Roussel, T. Grasser, G. I. Wirth, J. Franco, C. Vrancken, N. Horiguchi, and G. Groeseneken, "Response of a single trap to AC negative bias temperature stress," in *Proceedings of the IEEE International Reliability Physics Symposium*, Apr. 2011, pp. 4A.2.1–4A.2.8.

[17] B. Kaczer, S. Mahato, V. V. de Almeida Camargo, M. Toledano-Luque, P. J. Roussel, T. Grasser, F. Catthoor, P. Dobrovolny, P. Zuber, G. Wirth, and G. Groeseneken, "Atomistic approach to variability of bias-temperature instability in circuit simulations," in *Proceedings of the IEEE International Reliability Physics Symposium*, Apr. 2011, pp. XT.3.1–XT.3.5.
[18] M. J. M. Pelgrom, A. C. J. Duinmaijer, and A. P. G. Welbers, "Matching properties of MOS transistors," *IEEE Journal of Solid-State Circuits*, vol. 24, no. 5, pp. 1433–1439, Oct. 1989.
[19] A. J. Bhavnagarwala, X. Tang, and J. D. Meindl, "The impact of intrinsic device fluctuations on CMOS SRAM cell stability," *IEEE Journal of Solid-State Circuits*, vol. 36, no. 4, pp. 658–665, Apr. 2001.
[20] Y. Taur and T. H. Ning, *Fundamentals of modern VLSI devices*. Cambridge University Press, 1998.
[21] W. Wang, V. Reddy, A. T. Krishnan, R. Vattikonda, S. Krishnan, and Y. Cao, "Compact modeling and simulation of circuit reliability for 65-nm CMOS technology," *IEEE Transactions on Devices and Materials Reliability*, vol. 7, no. 4, pp. 509–517, Dec. 2007.
[22] Predictive Technology Model, http://www.eas.asu.edu/~ptm/.
[23] J. H. Pollard, *A Handbook of Numerical and Statistical Techniques: With examples mainly from the Life Sciences*. Cambridge University Press, 1977.
[24] T. Grasser, B. Kaczer, W. Goes, T. Aichinger, P. Hehenberger, and M. Nelhiebel, "A two-stage model for negative bias temperature instability," in *Proceedings of the IEEE International Reliability Physics Symposium*, Apr. 2009, pp. 33–44.
[25] S. Nassif, "Delay variability: Sources, impact and trends," in *Proceedings of the IEEE International Solid-State Circuits Conference*, Feb. 2000, pp. 368–369.
[26] J. Xiong, V. Zolotov, and L. He, "Robust extraction of spatial correlation," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 26, no. 4, pp. 619–631, Apr. 2007.
[27] G. Math, C. Benard, J. L. Ogier, and D. Goguenheim, "Geometry effects on the NBTI degradation of PMOS transistors," in *IEEE Integrated Reliability Workshop (IRW) Final Report*, Oct. 2008, pp. 1–15.
[28] B. C. Paul, K. Kang, H. Kufluoglu, M. A. Alam, and K. Roy, "Impact of NBTI on the temporal performance degradation of digital circuits," *IEEE Electron Device Letters*, vol. 26, no. 8, pp. 560–562, Aug. 2005.
[29] L. Zhang, W. Chen, Y. Hu, and C. C. Chen, "Statistical timing analysis with extended pseudo-canonical timing model," in *Proceedings of the Design, Automation & Test in Europe*, Mar. 2005, pp. 952–957.
[30] Nangate 45nm Open Cell Library, http://www.nangate.com/.
[31] Berkeley Logic Synthesis and Verification Group, ABC: A System for Sequential Synthesis and Verification, Release 70930. http://www.eecs.berkeley.edu/~alanmi/abc/.
[32] C. Zhuo, D. Blaauw, and D. Sylvester, "Post-fabrication measurement-driven oxide breakdown reliability prediction and management," in *Proceedings of the IEEE/ACM International Conference on Computer-Aided Design*, Nov. 2009, pp. 441–448.



Jianxin Fang received the B.S. and M.S. degree in electronic engineering from Tsinghua University, Beijing, China, in 2004 and 2007, and the Ph.D. degree in electrical engineering from University of Minnesota in 2012. He is currently a senior engineer in the ASIC design automation department with Qualcomm Technology Inc., San Diego, CA.

His research interests are reliability- and variation-aware CAD techniques, particularly related to the effects of oxide breakdown, hot carrier effects and bias temperature instability.

Jianxin was a recipient of the ISQED 2010 Best Paper Award and the IRPS 2012 Best Poster Award.

Sachin S. Sapatnekar (S'86, M'93, F'03) received the B. Tech. degree from the Indian Institute of Technology, Bombay, the M.S. degree from Syracuse University, and the Ph.D. degree from the University of Illinois at Urbana-Champaign. From 1992 to 1997, he was on the faculty of the Department of Electrical and Computer Engineering at Iowa State University. Since 1997, he has been at the University of Minnesota, where he currently holds the Distinguished McKnight University Professorship and the Robert and Marjorie Henle Chair in Electrical and Computer Engineering.

He is an author/editor of eight books, and has published widely in the area of computer-aided design of VLSI circuits. He has served as General Chair and Technical Program Chair of the ACM/EDAC/IEEE Design Automation Conference, the ACM International Symposium on Physical Design, and the IEEE/ACM International Workshop on the Specification and Synthesis of Digital Systems (Tau). He has served on the editorial boards of several publications, including the IEEE Transactions on Computer-Aided Design (currently as Editor-in-Chief), IEEE Design & Test of Computers, the IEEE Transactions on VLSI Systems, and the IEEE Transactions on Circuit and Systems II. He is a recipient of the NSF CAREER Award, six conference Best Paper awards and a Best Poster Award, and the Semiconductor Research Corporation (SRC) Technical Excellence award.