

Accounting for Inherent Circuit Resilience and Process Variations in Analyzing Gate Oxide Reliability

Jianxin Fang and Sachin S. Sapatnekar
 Department of ECE, University of Minnesota
 {fang0116, sachin}@umn.edu

Abstract—Gate oxide breakdown is a major cause of reliability failures in future nanometer-scale CMOS designs. This paper develops an analysis technique that can predict the probability of a functional failure in a large digital circuit due to this phenomenon. Novel features of the method include its ability to account for the inherent resilience in a circuit to a breakdown event, while simultaneously considering the impact of process variations. Based on standard process variation models, at a specified time instant, this procedure determines the circuit failure probability as a lognormal distribution. Experimental results demonstrate this approach is accurate compared with Monte Carlo simulation, and gives $4.7\text{--}5.9\times$ better lifetime prediction over existing methods that are based on pessimistic area-scaling models.

I. INTRODUCTION

Gate oxide breakdown is widely recognized to be a major reliability issue in CMOS circuits in current and future technologies, and its significance is increasing with each technology node. A majority of past work has focused on device-level analysis [1]–[3], but there has been growing interest in analysis at higher levels of abstraction.

Oxide breakdown creates an additional path for current to flow from the gate to the channel. The severity of the breakdown is generally described by the terms *hard breakdown* (HBD) and *soft breakdown* (SBD), and these can result either in parametric failures that alter circuit behavior but not its functionality, or in catastrophic functional failures that result in incorrect logic evaluation. In this paper, “circuit failure” is defined as a catastrophic functional failure. It has been observed that SBDs only cause parametric variations but not functional failures [4], [5], and therefore they are not considered in this work. Only HBDs are capable of resulting in functional failures [4], but not every HBD causes a functional failure [6].

The probability of circuit failure is significantly affected by on-chip process variations. At the circuit level, recent work [7] proposed a statistical approach for full-chip oxide reliability analysis considering process variation of T_{ox} ; however, this work did not present a path to determining the full distribution of the reliability function or statistics such as its variance. Subsequent work in [8] improved upon this by presenting a post-silicon analysis and mitigation method involving on-chip sensors and voltage tuning.

A major drawback of traditional approaches for circuit-level oxide reliability analysis, including [7], [8], is that they are all based on the simple notion of area-scaling, extrapolating the circuit-level failure rate from the characterized failure rate for a single device [4]. The idea of area-scaling is that if the total device area in a circuit is A , then the failure rate equals that of an isolated device of size A .

The drawback of this model lies in the fact that it assumes that any device breakdown will cause a functional failure in the entire circuit. There are two reasons why this is pessimistic: first, transistors are not always under stresses that cause HBDs (as shown in [2], only the inversion mode for NMOS can result in HBD generation), and second, not all HBDs lead to failure (as described in [6], digital

circuits can survive several HBDs without losing its functionality).

These drawbacks were addressed in recent work [9] that captured the effective stressing of NMOS inversion mode for HBD generation and utilized circuit simulation-based characterization to calculate the probability of circuit failure after a HBD event. A closed-form expression for the circuit failure probability was derived as a Weibull distribution with the same parameter β as single devices. This work showed that area-scaling based approaches are pessimistic in their calculations of circuit lifetime by $6 - 11\times$ at the nominal point. While this work makes a start towards more realistic failure modeling, it performs all of its analyses at the nominal point and neglects the effects of process variations. This significantly limits the practicability of the work. Moreover, as we will show, the actual distribution under variations is not a Weibull function.

Fig. 1 shows a comparison of failure probability vs. time for benchmark c7552 using (a) area scaling with worst-case T_{ox} , (b) area scaling with the T_{ox} variation model in [7], (c) area scaling with nominal T_{ox} , (d) the approach proposed in this paper, (e) the analysis method using nominal process parameters [9]. The $\mu+3\sigma$ value is used for (b) and (d). The figure shows significant differences between area-scaling based methods and the proposed method, and indicates that [9] is too optimistic. Therefore, to accurately predict circuit reliability, it is essential to account for the inherent circuit resilience and process variations simultaneously.

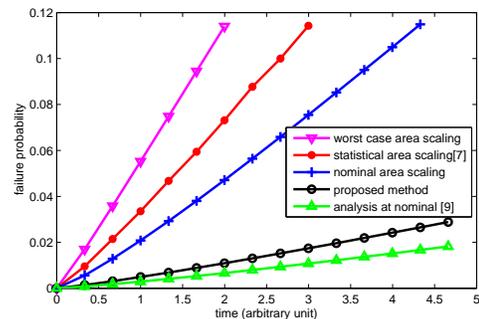


Fig. 1. Comparison of circuit failure, as predicted by various methods.

The goal of this paper is to predict the probability of oxide-breakdown-induced circuit failure for large digital circuits. Like [9], we leverage the fact that the dominant mode of circuit failure is due to NMOS HBDs, and we correct the limitations of the conventional area-scaling model. Over and above this, we capture the effect of process variations on circuit failure. To the best of our knowledge, this is the first work on oxide reliability considering both inherent circuit resilience and process variations. We demonstrate that the circuit failure probability at a specified time instant has a lognormal distribution due to process variations, and this distribution expands as the process variations and spatial correlation increase.

II. MODELING VARIATIONS

A. Process Variations

It is widely accepted that process parameter variations can be classified as lot-to-lot, die-to-die (D2D), and within-die (WID) variations, according to their scope; they can also be categorized as systematic and random variations by their causes and predictability. WID variations exhibit spatial dependence known as spatial correlation, which must be considered for accurate circuit analysis.

We employ a widely-used variational model: a process parameter X is modeled as a random variable about its mean, X_0 , as

$$\begin{aligned} X &= X_0 + X_g + X_s + X_r \\ \sigma_X^2 &= \sigma_{X_g}^2 + \sigma_{X_s}^2 + \sigma_{X_r}^2 \end{aligned} \quad (1)$$

Here, X_g , X_s , and X_r stand for the global part (from lot-to-lot or D2D variations), the spatially correlated part (from WID variation), and the residual random part, respectively. Under this model, all devices on the same die have the same global part X_g . The spatially correlated part is modeled using a method similar as [10], where the entire chip is divided into grids. All devices within the same grid have the same spatially correlated part X_s , and 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), and the oxide thickness (T_{ox}), and assume Gaussian-distributed parameters. The spatial correlation can be extracted as a correlation matrix [11], and processed using principal components analysis (PCA). The process parameter value in each grid is expressed as a linear combination of the independent principal components, with potentially reduced dimension. For a circuit with n transistors, with the three global parts for W , L and T_{ox} , the spatially correlated part and the n random parts, all the process parameters and their linear functions can be expressed in the random space with basis $\mathbf{e} = [\mathbf{e}_g, \mathbf{e}_s, \boldsymbol{\epsilon}]^T$ as

$$\begin{aligned} X &= X_0 + \Delta X = X_0 + \mathbf{k}_X^T \mathbf{e} \\ &= X_0 + \mathbf{k}_{X_g}^T \mathbf{e}_g + \mathbf{k}_{X_s}^T \mathbf{e}_s + k_\epsilon \epsilon \\ \sigma_X^2 &= \mathbf{k}_X^T \mathbf{k}_X, \quad \text{cov}(X_i, X_j) = \mathbf{k}_{X_i}^T \mathbf{k}_{X_j} - k_{\epsilon_i} k_{\epsilon_j} \end{aligned} \quad (2)$$

Here, $\mathbf{e}_g = [e_{Wg}, e_{Lg}, e_{Tg}]^T$ is the basis for global part, $\mathbf{e}_s = [e_1, \dots, e_t]^T$ is the basis of principal components for the spatial part, and $\epsilon \sim N(0, 1)$ is the independent random part for each parameter.

B. Variations in the Breakdown Location and Resistance

Fig. 2(a) includes a scatter plot that shows experimental results from [3], based on real measurements, illustrating the spread of the breakdown location in the channel (as measured from the source) and the breakdown resistance. Also shown is a MEDICI simulation that plots the predicted breakdown resistance at the nominal point. Several conclusions can be drawn from these plots, and we use all of these facts in our variational analysis.

First, the physical location of the breakdown location, x_{BD} , is uniformly distributed over the length of the channel.

Second, the breakdown event can be represented using a widely-used model that includes two linear resistors, R_s and R_d , from the gate to the source and the drain, respectively. As in [9], their nominal values may be modeled using the following piecewise linear/log-

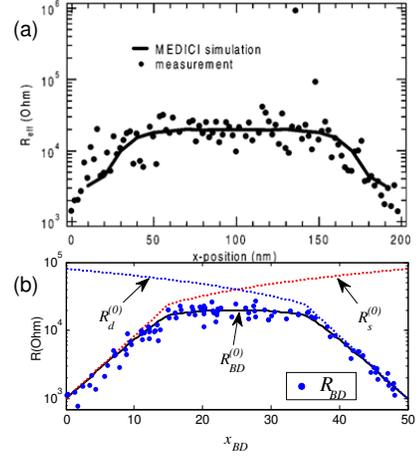


Fig. 2. Modeling the randomness of R_{BD} .

linear fit, illustrated in Fig. 2(b):

$$\begin{aligned} R_s^{(0)}(x) &= \begin{cases} ae^{bx}, & 0 \leq x \leq L_{ext} \\ kx, & L_{ext} \leq x \leq L \end{cases} \\ R_d^{(0)}(x) &= R_s^{(0)}(L - x) \\ R_{BD}^{(0)}(x) &= R_s^{(0)}(x) \parallel R_d^{(0)}(x) \end{aligned} \quad (3)$$

Here, L is the channel length, L_{ext} is the length of the source/drain overlap with channel, and a, b, k are the parameters characterized from experimental data. The effectiveness of this approach is also shown in Fig. 2(b), where the parallel combination of the resistors shows the same variational nature as the experimental data.

Third, due to variations, the data are not all on the nominal curve. We model process variations using a Gaussian distribution as:

$$R_{BD} = R_{BD}^{(0)}(1 + \lambda_r \epsilon_r) \text{ where } \epsilon_r \sim N(0, 1) \quad (4)$$

The random component ϵ_r for each NMOS transistor is assumed to be independent of that for the others, and independent of the process variations of W , L , and T_{ox} .¹

With the modeling of R_{BD} variation in Equation (4), we can update the nominal value of breakdown resistors $R_s^{(0)}$ and $R_d^{(0)}$ in Equation (3) to include the variation as follows,

$$\begin{aligned} R_s(x) &= \begin{cases} ae^{bx}(1 + \lambda_r \epsilon_r), & 0 \leq x \leq L_{ext} \\ kx(1 + \lambda_r \epsilon_r), & L_{ext} \leq x \leq L \end{cases} \\ R_d(x) &= R_s(L - x) \\ R_{BD}(x) &= R_s(x) \parallel R_d(x) \\ \epsilon_r &\sim N(0, 1) \end{aligned} \quad (5)$$

Fig. 2(b) plots this model, and the blue dots representing the Monte Carlo samples show a good modeling for the R_{BD} randomness.

III. FAILURE MODEL AND ANALYSIS UNDER VARIATIONS

In this section, we propose a circuit failure analysis considering process variations. Our nominal model is largely based on the work of [9], which completely neglects all variations. The contribution of this work is in building a framework for statistical analysis over this.

¹Published results do not indicate a correlation with these parameters, but such an effect could easily be included in our formulation.

A. Transistor-Level Models

The time-to-breakdown T_{BD} for device i is expressed as a Weibull distribution in [1]. Based on the observation in [2] that HBDs can only occur in NMOS inversion mode, a stressing coefficient, γ_i , can be introduced to capture the proportion of time that a NMOS device is in inversion mode: this coefficient can be computed from the signal probabilities in a circuit. For a device of area a_i , the breakdown probability as a function of time is the same as the Weibull distribution function for T_{BD} , which is given by

$$\Pr_{BD}^{(i)}(t) = 1 - \exp\left(-\left(\frac{\gamma_i t}{\alpha}\right)^\beta a_i\right) \quad (6)$$

Here, α and β are the parameters of the Weibull distribution. A common representation of a Weibull distribution is on the so-called *Weibull scale*, under the transform

$$W = \ln(-\ln(1 - \Pr)) = \beta \ln(\gamma_i t / \alpha) + \ln(a_i) \quad (7)$$

In other words, if we plot W as a function of $\ln(t)$, the result is a straight line with slope β and intercept $\ln(a_i)$.

Under variations, for transistor i , the Weibull slope β is a linear function of oxide thickness [1], [12]:

$$\beta_i = \beta_{i0} + c \Delta T_{ox}^{(i)} = \beta_{i0} + c \mathbf{k}_{T_i}^T \mathbf{e} \quad (8)$$

where β_{i0} denotes the nominal value. The T_{BD} distribution of i^{th} NMOS transistor under process variation has the same form as Equation (6), with β replaced by β_i . Its area, $a_i = W_i L_i$, is a product of two correlated Gaussians.

B. Cell-Level Analysis

The above analysis determines the probability that a transistor will experience an HBD in the presence of process variations. However, not every HBD results in circuit failure. The intuitive idea is that breakdown results in a leakage current through a device that is supposedly off, so that a resistive divider is created with the potential of changing the voltage at a gate output. Depending on the location of the breakdown and the breakdown resistance, some incidences may result in a logic failure while others may not. In fact, transistor sizing may be used to make a circuit more robust to such events. We begin by reprising the approach in [9] for the nominal case, and then present our solution for the case that accounts for process variations.

Consider a cell n that contains a transistor with oxide breakdown. Let k be the pin of cell n connected to the gate of this transistor, and let m be the logic cell that drives pin k of cell n . Then for any broken down NMOS transistor i , we can find the corresponding case index (m, n, k) . Figure 3(a) shows an example of such a breakdown case using a NAND2 as cell m , a NOR2 as cell n , and $k = 1$. Note that we assume a single breakdown event in these two cells, since the probability of two breakdowns is miniscule.

Thus, any breakdown of an NMOS transistor i can be mapped to a case indexed by the cell m that drives it, the cell n that contains it, and the input pin k that it connects to, as shown in Fig. 3. The input combination is denoted as \mathbf{V} . The circuit failure is judged by hard thresholds over the output voltages of cell m and n , which are also functions of x_{BD} , as the example shown in Fig. 4. An example simulation is shown in Fig. 4, showing the output voltage of cell m and cell n . Given a threshold V_H for the cell with a logic 1 output and V_L for a logic 0 output, it is seen that breakdown events close to the source or drain could cause a failure that violates the threshold.

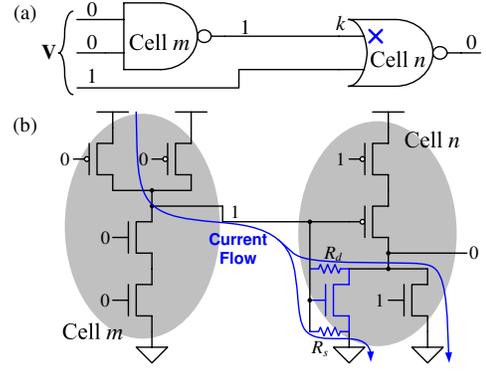


Fig. 3. Cell-level analysis of the breakdown case.

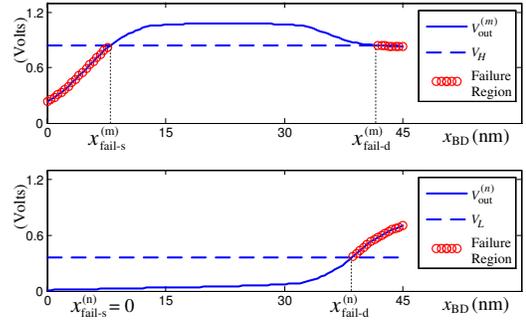


Fig. 4. Calculation of failure probability.

Based on the model in Equation (3), the breakdown case can be analyzed using a SPICE DC sweep over x_{BD} , enumerating all input combinations. This precharacterizes the (m, n, k, \mathbf{V}) case to determine the points $x_{fail-s}^{(m)}$, $x_{fail-s}^{(n)}$, $x_{fail-d}^{(m)}$, and $x_{fail-d}^{(n)}$, which refer to the breakdown locations where the corresponding cell output voltages cross the threshold, as illustrated in Fig. 4².

Given the uniform distribution of x_{BD} [3], the failure probability is equal to the proportion of x_{BD} where the threshold is violated. The source-side and drain-side failure probabilities under certain input vector were calculated separately as:

$$\begin{aligned} \Pr_{(fail-s|BD)}^{(m,n,k,\mathbf{V})} &= \max(p_s^{(m)}, p_s^{(n)}) \\ \Pr_{(fail-d|BD)}^{(m,n,k,\mathbf{V})} &= \max(p_d^{(m)}, p_d^{(n)}) \end{aligned} \quad (9)$$

where, for a given breakdown case (m, n, k, \mathbf{V}) at gate i ,

$$\begin{aligned} p_s^{(m)} &= \frac{x_{fail-s}^{(m)}}{L}, & p_d^{(m)} &= 1 - \frac{x_{fail-d}^{(m)}}{L} \\ p_s^{(n)} &= \frac{x_{fail-s}^{(n)}}{L}, & p_d^{(n)} &= 1 - \frac{x_{fail-d}^{(n)}}{L} \end{aligned} \quad (10)$$

The local failure probability caused by i^{th} NMOS HBD is calculated as the two-sided sum of the worst-case probabilities over all possible input vectors, and is given by

$$\Pr^{(i)} = \Pr_{(fail|BD)}^{(m,n,k)} = \max_{\mathbf{V}} \Pr_{(fail-s|BD)}^{(m,n,k,\mathbf{V})} + \max_{\mathbf{V}} \Pr_{(fail-d|BD)}^{(m,n,k,\mathbf{V})} \quad (11)$$

Under process variations, these failure probabilities depend on the breakdown resistor R_{BD} and parameters of all transistors in involved

²If no crossing point exists, the value of the parameter is set to zero at the source end or L at the drain end.

cell m and n . Using a first-order Taylor expansion,

$$p = p_0 + d_r^0 \lambda_r \epsilon_r + \sum_j d_{W_j}^0 \Delta W_j + \sum_j d_{L_j}^0 \Delta L_j + \sum_j d_{T_j}^0 \Delta T_j,$$

where $p \in \{p_s^{(m)}, p_s^{(n)}, p_d^{(m)}, p_d^{(n)}\}$. Here, d_x^0 is the first-order Taylor coefficients on parameter x . These coefficients is obtained using SPICE sensitivity analysis, and ΔW_j , ΔL_j and ΔT_j are random variables that can be expressed in the form in Equation (2). Since the case failure probability p is a linear combination of these process parameters and ϵ_r , it can also be expressed with vector \mathbf{e} ,

$$\begin{aligned} p &= p_0 + \mathbf{k}_p^T \mathbf{e} + d_r^0 \lambda_r \epsilon_r, \\ p &\in \{p_s^{(m)}, p_s^{(n)}, p_d^{(m)}, p_d^{(n)}\} \end{aligned} \quad (12)$$

Note that ϵ_r is the Gaussian representing the R_{BD} randomness from Equation (4), and is independent of the elements in \mathbf{e} .

From Equation (9), (11), and (12) we can obtain the source-side and drain-side failure probabilities using analytical methods. This involves applying the max operation on correlated Gaussians. The work in [13] provided a solution for this max function and approximated the result as a Gaussian in the same random space \mathbf{e} . Using such an approach, the final failure probability for case (m, n, k) is calculated using Equation (11) as the sum of maximum over all input vectors, and is approximated as a Gaussian of the form

$$\Pr^{(i)} = \Pr_{(\text{fail}|BD)}^{(m,n,k)} = \Pr_0^{(i)} + \mathbf{k}_{\Pr_i}^T \mathbf{e} + d_i \epsilon_{r_i} \quad (13)$$

The first part of the cell-level analysis, which is the case failure sensitivity analysis for (m, n, k, \mathbf{V}) , only depends on the cell library and can be performed as library characterization, and therefore its complexity is not included in discussions about the runtime or the computational complexity. The second part, including Equation (12) to (13), requires the specified data of process variations, therefore can only be implemented on a circuit-by-circuit basis.

C. Circuit-Level Analysis

The failure probability of a large digital circuit as a function of time is calculated using the weakest-link property, since any local failure caused by a NMOS HBD implies circuit failure.

$$\begin{aligned} \Pr_{\text{fail}}^{(\text{ckt})}(t) &= 1 - \prod_{i \in \text{NMOS}} \left(1 - \Pr_{\text{fail}}^{(i)}(t)\right) \\ &= 1 - \prod_{i \in \text{NMOS}} \left(1 - \Pr^{(i)} \Pr_{BD}^{(i)}(t)\right), \end{aligned} \quad (14)$$

i.e., a circuit is failure-free if every NMOS device is failure-free, or if any device failures are addressed by inherent resiliency.

By substituting Equation (6) and using the first-order Taylor approximation, the circuit failure probability was derived in [9] as

$$\Pr_{\text{fail}}^{(\text{ckt})}(t) = 1 - \exp\left(-\left(\frac{t}{\alpha}\right)^\beta \sum_{i \in \text{NMOS}} \Pr^{(i)} \gamma_i^\beta a_i\right). \quad (15)$$

which can also be expressed in the Weibull scale [1] as

$$\begin{aligned} W &= \ln\left(-\ln\left(1 - \Pr_{\text{fail}}^{(\text{ckt})}(t)\right)\right) = \ln \sum_{i \in \text{NMOS}} \left(\frac{\gamma_i t}{\alpha}\right)^\beta \Pr^{(i)} a_i \\ &= \beta \ln\left(\frac{t}{\alpha}\right) + \ln \sum_{i \in \text{NMOS}} \Pr^{(i)} \gamma_i^\beta a_i. \end{aligned} \quad (16)$$

Therefore, the failure probability of a large digital circuit also follows the Weibull distribution, and has the same slope β as a single device.

Note that the Weibull scale circuit failure probability using area scaling is $W = \beta \ln(t/\alpha) + \ln \sum a_i$. Since $\Pr^{(i)} \leq 1$ and $\gamma_i \leq 1$,

the area scaling based techniques always yield pessimistic result that is much larger than Equation (16).

Under a statistical model, we derive the following:

$$\exp(W) = \sum_{i \in \text{NMOS}} \left(\frac{\gamma_i t}{\alpha}\right)^{\beta_i} \Pr^{(i)} a_i = \sum_i \exp(y_i) \quad (17)$$

$$\text{where } y_i = \beta_i \ln\left(\frac{\gamma_i t}{\alpha}\right) + \ln\left(\Pr^{(i)} a_i\right) \quad (18)$$

$$= \beta_i \ln\left(\frac{\gamma_i t}{\alpha}\right) + \ln \Pr^{(i)} + \ln W_i + \ln L_i \quad (19)$$

Under process variations, for the i^{th} NMOS transistor, β_i is a Gaussian in random space \mathbf{e} as shown in Equation (8); $\Pr^{(i)}$ is a Gaussian in space $\mathbf{e} \cup \epsilon_{r_i}$ as in Equation (13); W_i and L_i are also Gaussians in space \mathbf{e} as assumed in Section II. Their logarithms are approximated Gaussians using moment-matching (see Appendix). As shown in our experimental results section, that approximation does not hurt the final result. Since $\Pr^{(i)}$ contains an additional random basis ϵ_{r_i} for R_{BD} variation, the sum of the logarithms S_i will contain both \mathbf{e} and ϵ_{r_i} . Denoting \mathbf{k}_{S_i} and q_i as the coefficients for these two parts, and μ_{S_i} as the mean of the sum,

$$S_i = \ln \Pr^{(i)} + \ln W_i + \ln L_i = \mu_{S_i} + \mathbf{k}_{S_i}^T \mathbf{e} + q_i \epsilon_{r_i} \quad (20)$$

Therefore y_i can be expressed as a Gaussian using \mathbf{e} and ϵ_{r_i} . Denoting $F_i = \ln(\gamma_i t/\alpha)$ and substituting Equation (19) with (20),

$$\begin{aligned} y_i &= \beta_i \ln\left(\frac{\gamma_i t}{\alpha}\right) + S_i \\ &= \beta_i F_i + \mu_{S_i} + (c F_i \mathbf{k}_{T_i} + \mathbf{k}_{S_i})^T \mathbf{e} + q_i \epsilon_{r_i} \end{aligned} \quad (21)$$

which means that y_i is also a Gaussian expressed in terms of \mathbf{e} and ϵ_{r_i} , and $\exp(y_i)$ will have a lognormal distribution. Note that y_i is the Weibull-scale failure probability corresponding to the HBD of i^{th} NMOS transistor. Under process variations, the speed of HBD generation is affected by ΔT_{ox} through β_i , and W_i and L_i through a_i , the probability of failure after HBD is affected by variations of all related transistors in corresponding cells through $\Pr^{(i)}$ and a_i . From Equation (17), $\exp(W)$ is the sum of correlated lognormal RVs. It can be approximated to a lognormal using Wilkinson's method [14], and its first two moments, u_1 and u_2 , are³

$$\begin{aligned} u_1 &= \sum_i \exp(\mu_{y_i} + \sigma_{y_i}^2/2) \\ u_2 &= \sum_i \exp(2\mu_{y_i} + 2\sigma_{y_i}^2) + \\ &\quad 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^N e^{\mu_{y_i} + \mu_{y_j}} e^{\frac{1}{2}(\sigma_{y_i}^2 + \sigma_{y_j}^2 + 2r_{ij} \sigma_{y_i} \sigma_{y_j})} \end{aligned} \quad (22)$$

When $\exp(W)$ is small enough, using a first-order Taylor expansion, we find from Equation (16) that

$$\Pr_{\text{fail}}^{(\text{ckt})} = 1 - \exp(-\exp(W)) \approx 1 - (1 - \exp(W)) = e^W \quad (23)$$

This result indicates that, when the circuit failure probability $\Pr_{\text{fail}}^{(\text{ckt})}$ is small (which is actually the case we are interested in, since a circuit with a very large number of breakdowns is unlikely to be functional),

³The calculation of u_2 requires the covariance of y_i and y_j . When the HBD case for NMOS i also involves NMOS j (i.e. j belongs to cell m or n) or vice versa, the random parts ϵ of y_i and y_j are actually correlated since they contain process parameters from the same transistor(s). This kind of case is fairly rare (about $2/N$ for a circuit with N logic cells), hence the correlations of the random parts are omitted to simplify the computation.

it can be approximated with $\exp(W)$, which has lognormal distribution with the first two moments given in Equation (22). When $\Pr_{\text{fail}}^{(\text{ckt})}$ is large, its distribution is unknown, but the mean and variance still can be calculated using a numerical method based on Equation (23). With this information of the circuit failure distribution, it is possible to predict the circuit failure probability at given time t with any specific confidence (e.g. 99%) using the distribution function.

The result also shows that the circuit-level failure probability under process variation is no longer a strict Weibull distribution along time, since the $\sigma_{y_i}^2$ in Equation (22) brings second order term $\ln^2 t$. Although this observation is based on approximations, it is confirmed by simulation results. The second-order and cross terms from $\sigma_{y_i}^2$ also prevent a general closed-form expression.

Due to the process variations, the mean value of circuit failure probability is increased by the $\sigma_{y_i}^2$ in Equation (22). The variance ($u_2 - u_1^2$) also increases with larger $\sigma_{y_i}^2$. This verifies that process variations exaggerate the likelihood of failure. Moreover, u_2 contains the term r_{ij} which depends positively on the spatial correlation. This means higher spatial correlation will increase the variance of failure probability, thus elevating the reliability issue.

D. Computational Complexity of the Analysis Method

For the circuit-level analysis in Section III.C, the calculation of y_i has constant complexity due to the limited number of involved devices, principal components, and elements in the random part \mathbf{e}_r (when using sparse computation). Using the recursive technique proposed in [15], a sum operation over N lognormal variables can be computed as $N - 1$ sum operations on two lognormal variables, keeping the computational complexity at $O(N)$. As stated earlier, the cell-level analysis in Section III.B is not included in this cost since it is a characterization that is carried out once for a library.

IV. EXPERIMENTAL RESULTS

The proposed method for circuit oxide reliability with process variations and spatial correlation was tested on the ISCAS85 and ITC99 benchmark circuits. The circuits were synthesized by ABC [16] using the Nangate 45nm open cell library [17], and then placement was carried out using a simulated annealing algorithm. The library characterization of cell-level failure probability in Section III.B was performed using HSPICE simulation and 45nm PTM model [18]. The method was implemented in C++ and tested on a Linux PC with 3GHz CPU and 2GB RAM.

The parameters for the Weibull distribution are $\alpha = 10000$ (arbitrary unit) and $\beta = 1.2$ [1]. The process variation of T_{ox} is chosen so that its 3σ point is 4% of its mean [8], and is split into 20% of global variation, 20% of spatially correlated variation and 60% of random variation. The variation of W and L sets the 3σ point to 12% of the mean [19], and is split to 40% of global variation, 40% of spatially correlated variation and 20% of random variation. The correlation matrix is uses the distance based method in [11]. The number of grids grows with the circuit size.

For each benchmark circuit, the mean and standard deviation of the failure probability are calculated at the time when the nominal circuit has a failure probability of 1%, using the proposed method and Monte Carlo (MC) simulation, separately. The MC simulation randomly generates 5000 circuit instances with different process parameters according to their distribution and correlation models: for each sample, we evaluate the failure probability by using the

random value of the process parameters as inputs to the approach in [9].

Table I presents the statistics of the circuit failure probability using the proposed method. The first three columns represent the circuit name and its characteristics. Information about the mean and standard deviation of the failure probability using our approach are presented in the next two columns, and the corresponding errors relative to MC in the following two. It can be seen that our approach closely matches MC, with average errors of 0.8% for the mean and 1.8% for the standard deviation. The value of the mean is very close to the nominal failure probability of 1%, but the standard deviation is considerable. The last two columns compare the circuit lifetime at time $\mu+3\sigma$ for our approach with the nominal approach in [9] and the area-scaling method under variations, respectively. We see that the circuit lifetime decrease 19–23% due to process variation, and the proposed approach shows 4.7–5.9 \times lifetime relaxation against the pessimistic area-scaling method.

Fig. 5 plots the probability density function (PDF) and cumulative density function (CDF) of benchmark c7552 at the nominal failure probability of 1%. The dotted curves show results of MC simulation, while the solid curves show lognormal distribution obtained using proposed method. The nearly perfect match of these two methods validates the approximations made during the analysis, and proves that the circuit failure probability has a *lognormal distribution* in the region of interest, rather than a Weibull distribution. Consistent with this observation, if we plot this on the Weibull scale, it can be seen that this distribution is not a straight line with a constant Weibull slope.

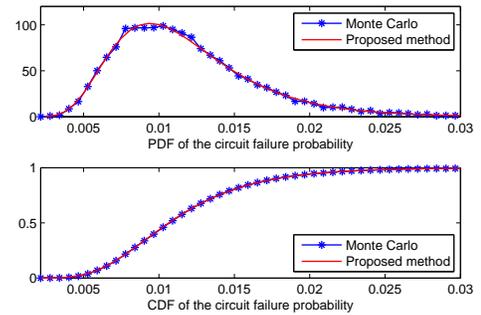


Fig. 5. Comparison of the PDF and CDF of circuit failure.

The proposed method is also tested with other process parameter variance and correlation data besides the condition assumed above. Table II shows the $\mu+3\sigma$ value of circuit failure when nominal circuit failure probability is 1%, and its error against MC simulation for benchmark c7552, under several process variation and spatial correlation conditions:

TABLE II
CIRCUIT FAILURE OF C7552 UNDER DIFFERENT TEST CONDITIONS.

Process Variation	Less correlation $g/s/r=10/10/80\%$		Medium correlation $g/s/r=30/40/30\%$		More correlation $g/s/r=50/40/10\%$	
	$\mu+3\sigma$	Error	$\mu+3\sigma$	Error	$\mu+3\sigma$	Error
$\sigma/\mu=1\%$	1.13%	0.20%	1.23%	0.24%	1.27%	0.15%
$\sigma/\mu=2\%$	1.27%	0.04%	1.47%	-0.06%	1.56%	-0.27%
$\sigma/\mu=5\%$	1.85%	0.81%	2.45%	0.44%	2.72%	2.15%
$\sigma/\mu=10\%$	3.99%	0.62%	6.15%	2.16%	7.29%	5.01%

The labels g, s, r in the table stand for the global part, the spatially correlated part and the random part of the parameter variations. The results indicate that the relative error to MC simulation is small under all the test conditions, indicating the proposed method is accurate and robust to different conditions of process variations. Moreover, we

TABLE I
COMPARISONS OF THE MEAN μ AND σ OF CIRCUIT FAILURE.

circuit name	Size		Failure probability		Error to MC		Runtime		3σ lifetime	
	#Cells	#Grids	μ	$\frac{\sigma}{\mu}$	μ	$\frac{\sigma}{\mu}$	Proposed	MC	Nominal	Area scaling
c432	221	4	1.02%	8.87%	0.89%	-0.45%	1.06s	24.8s	-18.6%	5.2×
c880	384	9	1.02%	8.99%	1.08%	2.81%	1.50s	38.4s	-18.8%	5.1×
c1355	596	9	1.02%	9.20%	0.73%	2.52%	1.88s	41.1s	-19.1%	5.1×
c2670	759	16	1.02%	9.53%	0.63%	1.16%	4.84s	126s	-19.7%	5.9×
c3540	1033	16	1.02%	9.56%	0.62%	2.05%	6.72s	191s	-19.8%	5.3×
c5315	1699	25	1.02%	9.63%	0.86%	5.57%	6.30s	164s	-20.0%	5.2×
c6288	3560	64	1.02%	10.4%	1.12%	5.53%	17.2s	434s	-21.3%	5.4×
c7552	2316	36	1.03%	9.82%	0.55%	1.00%	9.34s	275s	-20.2%	5.1×
b14	4996	81	1.02%	10.2%	0.87%	1.96%	53.5s	1064s	-21.0%	5.2×
b15	6548	100	1.02%	10.3%	0.80%	0.59%	61.6s	1247s	-21.1%	5.0×
b17	20407	361	1.03%	11.3%	0.90%	-0.53%	233s	4195s	-22.8%	4.7×
b20	11033	169	1.03%	10.7%	0.97%	1.51%	120s	2156s	-21.8%	5.0×
b21	10873	169	1.03%	10.7%	0.79%	1.01%	69.1s	2489s	-21.7%	4.9×
b22	14974	225	1.03%	10.9%	0.75%	0.78%	107s	3321s	-22.1%	4.9×

observe that as the $\mu+3\sigma$ value of the failure probability increases when the spatial component of the process variation increases, or when the correlation increases. This verifies again that the process variations and spatial correlation elevate the reliability issues due to oxide breakdown.

V. CONCLUSION

The paper has focused on the reliability issues caused by gate oxide breakdown in CMOS digital circuits, with the consideration of process variations and the inherent resilience in a circuit that prevents every breakdown from causing circuit failure. The proposed approach takes account for the effective stressing for HBD generation and the probability of circuit failure after HBD occurrences. The circuit failure probability at specified time instant is derived to be a lognormal distribution due the process variations. Experimental results show this approach is accurate compared with Monte Carlo simulation, and gives 4.7–5.9× better lifetime prediction compared with the pessimistic area-scaling method.

REFERENCES

- [1] E. Y. Wu et al. CMOS scaling beyond the 100-nm node with silicon-dioxide-based gate dielectrics. *IBM J. Res. Dev.*, 46(2/3):287–298, March/May 2002.
- [2] F. Crupi et al. A comparative study of the oxide breakdown in short-channel nMOSFETs and pMOSFETs stressed in inversion and in accumulation regimes. *IEEE Trans. Device and Mater. Rel.*, 3(1):8–13, March 2003.
- [3] R. Degraeve et al. Relation between breakdown mode and location in short-channel nMOSFETs and its impact on reliability specifications. *IEEE Trans. Device and Mater. Rel.*, 1(3):163–169, September 2001.
- [4] J. H. Stathis. Physical and predictive models of ultrathin oxide reliability in CMOS devices and circuits. *IEEE Trans. Device and Mater. Rel.*, 1(1):43–59, March 2001.
- [5] H. Wang et al. Impact of random soft oxide breakdown on SRAM energy/delay drift. *IEEE Trans. Device and Mater. Rel.*, 7(4):581–591, December 2007.
- [6] B. Kaczer et al. Impact of MOSFET gate oxide breakdown on digital circuit operation and reliability. *IEEE Trans. Electron Devices*, 49(3):500–506, March 2002.
- [7] K. Chopra et al. A statistical approach for full-chip gate-oxide reliability analysis. In *Proc. ICCAD*, pages 698–705, November 2008.
- [8] Z. Cheng, D. Blaauw, and D. Sylvester. Post-fabrication measurement-driven oxide breakdown reliability prediction and management. In *Proc. ICCAD*, pages 441–448, November 2009.
- [9] J. Fang and S. S. Sapatnekar. Scalable methods for the analysis and optimization of gate oxide breakdown. In *Proceedings of the IEEE International Symposium on Quality Electronic Design*, pages 638–645, March 2010.

- [10] H. Chang and S. S. Sapatnekar. Statistical timing analysis considering spatial correlations using a single PERT-like traversal. In *Proc. ICCAD*, pages 621–625, November 2003.
- [11] J. Xiong, V. Zolotov, and L. He. Robust extraction of spatial correlation. *IEEE Trans. Comput.-Aided Des.*, 26(4):619–631, April 2007.
- [12] R. Degraeve, G. Groeseneken, R. Bellens, M. Depas, and H. E. Maes. A consistent model for the thickness dependence of intrinsic breakdown in ultra-thin oxides. In *Proceedings of the IEEE International Electronic Devices Meeting*, pages 863–866, December 1995.
- [13] C. Clark. The greatest of a finite set of random variables. *Operations Research*, 9:85–91, 1961.
- [14] A. A. Abu-Dayya and N. C. Beaulieu. Comparison of methods of computing correlated lognormal sum distributions and outages for digital wireless applications. In *IEEE 44th Vehicular Technology Conference*, volume 1, pages 175–179, June 1994.
- [15] A. Srivastava, S. Shah, K. Agarwal, D. Sylvester, D. Blaauw, and S. Director. Accurate and efficient gate-level parametric yield estimation considering correlated variations in leakage power and performance. In *Proc. DAC*, pages 535–540, 2005.
- [16] Berkeley Logic Synthesis and Verification Group. Abc: A system for sequential synthesis and verification, release 70930. <http://www.eecs.berkeley.edu/~alanmi/abc/>.
- [17] Nangate 45nm Open Cell Library. <http://www.nangate.com/>.
- [18] Predictive Technology Model. <http://www.eas.asu.edu/~ptm/>.
- [19] International technology roadmap for semiconductors, 2008 update, process integration, devices and structures.

APPENDIX

Logarithm of a Gaussian RV: For $x \sim N(\mu_x, \sigma_x^2)$, given $\mu_x \gg \sigma_x > 0$ so that $x > 0$ is always true, its logarithm $y = \ln x$ can be approximated linearly as $y = c + kx$. In order to get better accuracy, the following moments matching method is used instead of first-order Taylor expansion.

For $y = \ln x$, we want to approximate it as $y' \sim N(\mu_y, \sigma_y^2)$. Therefore $x' = \exp(y')$ has a lognormal distribution with first two moments

$$\begin{aligned} u_1 &= \exp(\mu_y + \sigma_y^2/2) \\ u_2 &= \exp(2\mu_y + 2\sigma_y^2) \end{aligned} \quad (24)$$

By matching the first two moments of x' and x : $u_1 = \mu_x$, $u_2 = \sigma_x^2 + \mu_x^2$, we can get the distribution of y as

$$\begin{aligned} \mu_y &= 2 \ln \mu_x - \frac{1}{2} \ln(\sigma_x^2 + \mu_x^2) \\ \sigma_y^2 &= \ln(\sigma_x^2 + \mu_x^2) - 2 \ln \mu_x \end{aligned} \quad (25)$$

Therefore the coefficients for the linear form $y = c + kx$ are $k = \sigma_y/\sigma_x$ and $c = \mu_y - \mu_x\sigma_y/\sigma_x$. \square