Statistical lifetime reliability optimization considering joint effect of process variation and aging

Song Jin\textsuperscript{a,b}, Yinhe Han\textsuperscript{a,b}, Huawei Li\textsuperscript{a,b,*}, Xiaowei Li\textsuperscript{a,b}

\textsuperscript{a} Key Laboratory of Computer System and Architecture, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, PR China
\textsuperscript{b} Graduate University of Chinese Academy of Sciences, Beijing, PR China

\begin{abstract}
Aging effect degrades circuit performance in the runtime, interacts with fabrication-induced device parameter variation, and thus posing significant impact on circuit lifetime reliability. In this work, a statistical circuit optimization flow is proposed to ensure lifetime reliability of the manufactured chip in the presence of process variation and aging effects. It exploits a variation-aware gate-level statistical aging degradation model to characterize circuit lifetime reliability, identifies a set of worst duty cycles on the inputs of statistically critical gates to estimate the worst delay degradations on these gates. Based on the delay degradation information, statistical gate sizing is performed which enables the manufactured chip to satisfy lifetime reliability constraint in term of low area overhead.
\end{abstract}

1. Introduction

With aggressive scaling of the feature size, fabrication-induced process variation (PV) causes circuit performance to deviate from its initial design expectation\cite{1}. Borkar et al. reported that PV effect can cause about 20\% variation in chip leakage and 30\% variation in chip frequency\cite{2}. It is common for circuit optimization procedure at design phase to consider PV-induced uncertainty so that the manufactured chip can meet the specific timing constraint.

However, aging effect, such as negative bias temperature instability—NBTI\cite{3,4,5}, degrades circuit performance continuously in the runtime, results in the chip still suffering from reliability problem during service lifetime. Wang et al.\cite{6} mentioned that path delay can be increased up to 20\% under worst case NBTI-induced degradation. Some safety or task-critical systems, e.g., aerospace application, work in the unmanned situation. Due to the difficulty or the impossibility of repairing or replacing before the end of the device lifetime, both the exceedingly high reliability and the designated lifetime are required. Although some reliability-aware technologies like triple-module redundancy have been used to tolerant hard defect or soft error, they cannot handle aging-induced degradation in the runtime.

Process variation and NBTI effects have strong influence on each other\cite{7,8}. Device parameters, such as threshold voltage and oxide thickness manifest uncertainty after fabrication under PV effect. In such case, NBTI-induced threshold voltage degradation should also be modeled as a statistical process. Similarly, path delay increases under NBTI effect with time, which will change the statistic of circuit timing as operational time evolving. Therefore, lifetime reliability optimization technology cannot treat PV and NBTI effects as two independent issues.

In this work, we propose a statistical circuit optimization flow to improve lifetime reliability of the chip under considering the joint effect of PV and NBTI. Our contributions are as follows.

- A compact model is proposed to characterize the statistical gate delay degradation under the joint effect of PV and NBTI. The newly introduced parameters can reflect PV-induced parameter variation with spatial correlation as well as NBTI-induced temporal degradation. Based on this model, statistical timing analysis can be exploited to characterize lifetime reliability for the whole circuit with linear runtime complexity.
- Considering the practical delay degradations on statistically critical gates, an effective optimization metric is proposed to guide the gate sizing procedure converging quickly with low area overhead.

The proposed statistical circuit optimization flow is illustrated in Fig. 1. As shown in Fig. 1, a statistical gate-level aging model is obtained by incorporating PV-induced uncertainty into the conventional NBTI model (parts inside the dashed box). It can
be used to characterize degradation on the gate delay under PV and aging effects. Extending this gate-level model into the whole circuit, statistical static timing analysis is performed to identify a set of statistically critical gates which affect delay distribution of the circuit (parts inside the dash-dotted box). Then, the worst duty cycles on the inputs of statistically critical gates are identified, which facilitate to obtain the worst delay degradations on these gates in the practical circuit operation. Given the delay degradation information, optimization priority of the statistically critical gate is decided and then used to guide the gate sizing procedure to optimize the circuit timing. The gate sizing procedure terminates until the specific lifetime reliability constraint is satisfied or the maximum area constraint is violated (parts inside the solid box).

The rest of paper is organized as follows. Related work is presented in Section 2. Section 3 introduces the PV-aware gate-level statistical aging degradation model. Identifying statistically critical gates is presented in Section 4. Section 5 introduces the gate sizing procedure. Experiment results are presented in Section 6. We conclude in Section 7.

2. Related work

Previous research commonly treated PV- and aging-induced reliability issues separately. Gate sizing technologies [9–11] based on statistical static timing analysis were proposed to mitigate the PV-induced uncertainty on circuit timing. Teodorescu et al. [12] proposed dynamic fine-grain body biasing to reduce the leakage power under PV effect. However, all of these works only concentrated on mitigating the static PV effect while ignoring the dynamic aging-induced degradation. On the other hand, some proposed NBTI models and NBTI-resistant technologies [3–6,13,14] assumed that the circuit operates at the nominal threshold voltage of PMOS. This underestimated circuit performance degradation induced by the joint effect of aging and PV. On-line aging prediction [15,16] can predict circuit failure in advance under PV and aging effects. However, it can not provide a guide for the optimization procedure at design phase to mitigate the negative impact of PV and aging effects on circuit lifetime reliability.

Recently, Wang et al. [7] thoroughly analyzed the joint effect of PV and aging on single path delay degradation. While they did not consider the spatial correlation between different paths. In [8], Lu et al. proposed an analysis framework to evaluate the circuit lifetime reliability under PV and aging effects. Stochastic collocation method and polynomial chaos expansion were used to model the gate delay under uncertainty.

3. PV-aware statistical gate delay degradation model

In this section, we introduce some new parameters into the conventional NBTI model to formulate the PV-aware NBTI-induced delay degradation model. Extending this model to the whole circuit enables circuit timing analysis to cover NBTI-induced temporal delay degradation and PV-induced spatial delay variation simultaneously.

3.1. Statistical NBTI model

NBTI is a special aging effect on PMOS transistor. A well-known NBTI model is proposed in [6] to predict the long-term threshold voltage degradation of PMOS without considering PV effect:

\[
\Delta V_{th_{,nbiti}} = \left[ \frac{K_t^2 \cdot T_{dik} \cdot \alpha \cdot (1 - \beta_t^{1/2n})}{2n} \right]^{2n}
\]

where

\[
\beta_t = 1 - \left( \frac{2 \cdot \epsilon_t \cdot \epsilon_x + \sqrt{\epsilon_2 \cdot (1 - \alpha) \cdot T_{dik}}}{2 \cdot \epsilon_t + \sqrt{\epsilon_1}} \right)
\]

\[K_t = \frac{q \cdot \epsilon_{ox}}{\epsilon_{ox} \cdot V_{th} \cdot \epsilon_{ox} \cdot (V_{th} - V_{th}) \cdot \sqrt{\epsilon_1}} \cdot \exp(2E_{ox}/E_{ot}) \]

In this model, \( n \) is around 0.16 given diffusion species are \( H_2 \). \( t \) is the totally operational time that circuit experienced. \( V_{th} \) is the nominal threshold voltage of PMOS.

The parameter \( \alpha \) is called duty cycle, which reflects the impact of executed workloads on NBTI-induced degradation. Duty cycle

![Fig. 1. Statistical circuit optimization flow.](image-url)
in NBTI literature is defined as the fraction of time that the PMOS spent in stress state (i.e., input signal is logic ‘0’) for a period of time. It can be recognized as the statistical signal probability [6] (here, the signal probability is defined as the probability that the input signal is logic ’0’). For example, given the signal probability on a gate input for a period of time $t$ is 0.5, we can say that the PMOS connected to this gate input has been stressed for 0.5$t$ during the time $t$. Correspondingly, duty cycle on this gate input is 0.5 as well.

In Eq. (1), $V_{th}$ is assumed as nominal value. However, $V_{th}$ in the manufactured chip will show statistical distribution under PV effect and can be expressed as

$$V_{th} = V_{th,nom} + \Delta V_{th,sys} + \Delta V_{th,ran}$$

(2)

where $V_{th,nom}$ is the nominal threshold voltage. $\Delta V_{th,sys}$ and $\Delta V_{th,ran}$ are the changes of threshold voltage due to systematic and random variations, respectively. In this work, we only concern the intra-die variation. However, the inter-die variation can be easily incorporated into our model by adding a die-level variation value on the intra-die variation.

Substituting Eq. (2) into Eq. (1), a variation-aware statistical NBTI degradation model can be obtained. Moreover, by only concentrating on some important parameters which have strong impacts on NBTI-induced degradation, the statistical model can be further simplified as

$$\Delta V_{th,nbt} = A \cdot (1 - \gamma) \cdot (\Delta V_{th,sys} + \Delta V_{th,ran}) \cdot x^g \cdot t^n$$

(3)

where $A$ is a fitted parameter related to the specific process and operational conditions (such as the supply voltage $V_{dd}$ and the working temperature $T$). $\gamma$ denotes sensitivity of NBTI-induced degradation to PV effect. It can be fitted from HSPICE simulation by using different variation values. $x$ and $t$ are reserved due to their exponential impacts on NBTI-induced degradation [6].

On the presence of spatial correlation, $V_{th}$’s between different gates are no more independent. In this work, the quad-tree partitioning method proposed in [17] is used to model spatial correlation between $V_{th}$’s of different gates. As shown in Fig. 2, area of the die is divided into grids through multi-level quad-tree partition. Each level has total 4$^m$ grids, where $m$ is the level number starting from 0. Each grid in all of the levels is attached an independent Gaussian random variable to represent a component of $V_{th}$ variation. Then the gate $V_{th}$ variation can be modeled as the sum of random variables in one of the grids at different levels. For example, the $V_{th}$ variations of gates 1, 2 and 3 in Fig. 2 can be expressed as

\[
\begin{align*}
\Delta V_{th(1)} &= \Delta V_{th(2)} + \Delta V_{th(1)} + \Delta V_{th(0)} + \Delta V_{th,ran(1)} \\
\Delta V_{th(2)} &= \Delta V_{th(2.4)} + \Delta V_{th(1)} + \Delta V_{th(0)} + \Delta V_{th,ran(2)} \\
\Delta V_{th(3)} &= \Delta V_{th(2.16)} + \Delta V_{th(1.4)} + \Delta V_{th(0.1)} + \Delta V_{th,ran(3)}
\end{align*}
\]

Given this model, it can be seen that the adjacent gates, such as gates 1 and 2 share more common variables, which reflects their stronger correlation. On the contrary, gates 1 and 3 are more weakly correlated as they share only one common variable. $\Delta V_{th,ran}$ is an independent random variable which models the uncorrelated gate $V_{th}$ variation.

Applying the quad-tree partition to model the spatial correlation of $V_{th}$’s between different gates, for PMOS $k$ in a gate, Eq. (3) is reformulated as

$$\Delta V_{th,nbt(k)} = A \cdot (1 - \gamma) \cdot \left( \sum_{l=0}^{m} \Delta V_{th(l)} + \Delta V_{th,ran(k)} \right) \cdot x^g \cdot t^n$$

(4)

Finally, to simplify the notation, Eq. (4) is rewritten as

$$\Delta V_{th,nbt(k)} = A \cdot (1 - \gamma) \cdot \sum_i \Delta V_{th(i)} \cdot x^g \cdot t^n$$

(5)

where $\Delta V_{th(i)}$ corresponds to one of the random variables in the model, such as $\Delta V_{th(0)}$ and $\Delta V_{th,ran(k)}$. The notation $\sum$ takes over all of the random variables in the model. For the random variables which are not associated with PMOS $k$, the coefficient $A \cdot \gamma \cdot x^g \cdot t^n = 0$.

3.2. Statistical gate delay degradation model

Threshold voltage of PMOS $k$ under the joint effect of PV and aging can be expressed as

$$V_{th(k)} = V_{th,nom(k)} + \Delta V_{th,sys(k)} + \Delta V_{th,ran(k)} + \Delta V_{th,nbt(k)}$$

(6)

Substituting $\Delta V_{th,nbt(k)}$ in Eq. (6) with Eq. (5), we get

$$V_{th(k)} = V_{th,nom(k)} + A \cdot x^g \cdot t^n + (1 - A \cdot \gamma \cdot x^g \cdot t^n) \cdot \sum_i \Delta V_{th(i)}$$

(7)

Based on the well-known alpha-power law proposed in [18], the gate propagation delay can be approximately recognized as linear function of the threshold voltage. Therefore, the propagation delay from input node $k$ of the gate to the gate output can be expressed as

$$D_k = F \left( V_{th,nom(k)} + A \cdot x^g \cdot t^n + (1 - A \cdot \gamma \cdot x^g \cdot t^n) \cdot \sum_i \Delta V_{th(i)} \right)$$

(8)

where $F$ denotes the existed linear function relation between the gate propagation delay and the threshold voltage. For simplicity, we use “gate delay” to represent “gate propagation delay” in the rest of paper.

Considering above linear function relation between the gate delay and the threshold voltage, Eq. (8) is transformed as

$$D_k = D_{nom(k)} + \beta_k \cdot x^g \cdot t^n + (1 - A \cdot \gamma \cdot x^g \cdot t^n) \cdot \beta_k \cdot \sum_i \Delta V_{th(i)}$$

(9)

where $D_{nom(k)}$ denotes the nominal gate delay. $\beta_k$ is a fitted coefficient which reflects the increase of gate delay due to NBTI-induced threshold voltage increase under nominal condition. $\beta_k$ is also a fitted coefficient which denotes the impact of PV-induced threshold voltage variation on the gate delay change without considering NBTI effect.

4. Identifying statistically critical gate

In this section we identify a set of statistically critical gates which affect the circuit delay distribution under PV and aging effects.
4.1. Fast statistical static timing analysis method

To obtain the statistically critical gates, we first identify statistically critical paths in the circuit. Statistically critical path denotes that its delay under the joint effect of PV and aging will violate the specific timing constraint. By extending the statistical gate delay degradation model into the whole circuit, statistical static timing analysis (SSTA) can be performed to identify the statistically critical path. In this work we employ a similar SSTA method as in [17], which calculates statistical bound on probability distribution function of the circuit delay. Computational process of the maximum arrival time is linear. It ensures the circuit timing analysis to perform with linear runtime complexity while to preserve correlation information between arrival time and gate delay.

The statistical timing analysis consists of two repeated operations: propagation and merging of the arrival time. The propagation operation propagates the arrival time from gate input to gate output. In this process, gate delay is added to the arrival time. The merging operation merges multiple arrival times converging into gate output from different gate inputs by calculating the maximum of these arrival times.

In this work, the gate delay and the arrival time are both expressed as the form in Eq. (9). The arrival time $A_i$ at gate input node $i$ is

$$A_i = A_{\text{nom}(i)} + \sum_j B_j \cdot x_j^n \cdot t^n + \left( \sum_j \left( 1 - A \cdot \gamma \cdot x_j^n \cdot t^n \right) \cdot \beta_j \right) \cdot \sum_i \Delta V_{\text{th}(i)}$$

where $A_{\text{nom}(i)}$ denotes the nominal arrival time at input node $i$. The notation $\sum_j$ denotes the accumulative process of adding $j$ gate delays onto the arrival time before the arrival time calculation reaches at node $i$.

Given propagation delay $D_i$ of the gate from input node $i$ to the output node $k$:

$$D_i = D_{\text{nom}(i)} + B_i \cdot x_i^n \cdot t^n + (1 - A \cdot \gamma \cdot x_i^n \cdot t^n) \cdot \beta_i \cdot \sum_i \Delta V_{\text{th}(i)}$$

the arrival time $A_k$ at $k$ is

$$A_k = A_{\text{nom}(k)} + \sum_i B_i \cdot x_i^n \cdot t^n + \left( \sum_j \left( 1 - A \cdot \gamma \cdot x_j^n \cdot t^n \right) \cdot \beta_j \right) \cdot \sum_i \Delta V_{\text{th}(i)}$$ (11)

where $A_{\text{nom}(k)} = D_{\text{nom}(i)} + A_{\text{nom}(i)}$.

$$\sum_j B_j \cdot x_j^n \cdot t^n = B_i \cdot x_i^n \cdot t^n + \sum_j B_j \cdot x_j^n \cdot t^n$$,

$$\sum_j \left( 1 - A \cdot \gamma \cdot x_j^n \cdot t^n \right) \cdot \beta_j = (1 - A \cdot \gamma \cdot x_i^n \cdot t^n) \cdot \beta_i + \sum_j (1 - A \cdot \gamma \cdot x_j^n \cdot t^n) \cdot \beta_j$$,

with $l = j + 1$.

Calculation of $A_k$ is exact and preserves the correlation information between the gate delay $D_i$ and the arrival time $A_i$.

Reference [17] proved that for any given numbers $a_1, a_2, \ldots, a_n$ and $x_1, x_2, \ldots, x_n$ the following theorem is valid:

$$\max_{i=1}^{n} a_i + \max_{i=1}^{n} x_i \leq \max_{i=1}^{n} a_i + \sum_{i=1}^{n} x_i$$ (12)

Based on Eq. (12), the maximum of two arrival times can be computed as follows. Given two arrival times $A_1$ and $A_2$:

$$A_1 = A_{\text{nom}(1)} + \sum_m B_m \cdot x_m^n \cdot t^n + \left( \sum_m \left( 1 - A \cdot \gamma \cdot x_m^n \cdot t^n \right) \cdot \beta_m \right) \cdot \sum_i \Delta V_{\text{th}(i)}$$

$$A_2 = A_{\text{nom}(2)} + \sum_p B_p \cdot x_p^n \cdot t^n + \left( \sum_p \left( 1 - A \cdot \gamma \cdot x_p^n \cdot t^n \right) \cdot \beta_p \right) \cdot \sum_i \Delta V_{\text{th}(i)}$$

where $m$ and $p$ are the number of gates visited during the computing processes of $A_1$ and $A_2$, respectively. $A_3$, the maximum of these two arrival times is

$$A_3 = A_{\text{nom}(3)} + \sum_q B_q \cdot x_q^n \cdot t^n + \left( \sum_q \left( 1 - A \cdot \gamma \cdot x_q^n \cdot t^n \right) \cdot \beta_q \right) \cdot \sum_i \Delta V_{\text{th}(i)}$$ (13)

where $A_{\text{nom}(3)} = \max(A_{\text{nom}(1)}, A_{\text{nom}(2)})$.

$$\sum_q B_q \cdot x_q^n \cdot t^n = \max \left( \sum_m B_m \cdot x_m^n \cdot t^n, \sum_p B_p \cdot x_p^n \cdot t^n \right)$$

$$\sum_q \left( 1 - A \cdot \gamma \cdot x_q^n \cdot t^n \right) \cdot \beta_q = \max \left( \sum_m \left( 1 - A \cdot \gamma \cdot x_m^n \cdot t^n \right) \cdot \beta_m, \sum_p \left( 1 - A \cdot \gamma \cdot x_p^n \cdot t^n \right) \cdot \beta_p \right)$$

Here, $q = m + p$. $A_3$ is actually the upper bound of the maximum of $A_1$ and $A_2$ [17]. Using Eqs. (11) and (13), statistical timing analysis can be performed to compute the statistical circuit maximum arrival time or the statistical maximum arrival time at any node.

4.2. Identifying statistically critical gates

To guarantee that all of the statistically critical paths can be covered, SSTA is performed under worst case consideration. That is, a path will be identified as statistically critical path if $\mu + 3\sigma$ of its maximum arrive time is larger than the specific timing constraint.

After the identification of statistically critical paths, all of the gates lying on these paths are identified as statistically critical gates. The subsequent gate sizing procedure optimizes these statistically critical gates to ensure that circuit delay satisfies the timing constraint during the service lifetime even under the joint effect of PV and aging.

5. Statistical gate sizing

It is obvious that the gate with larger delay degradation has larger impact on delay degradation of the path which the gate lies on. Therefore, it is essential for the gate sizing procedure to optimize the statistically critical gate which has largest delay degradation with high priority.

5.1. Identifying worst duty cycles

Delay degradation on the gate is largely decided by the executed workloads in the practical circuit operation. As we mentioned in Section 3, duty cycle can be recognized as the statistical signal probability. Thereby it can be used to reflect the fraction of time that the gate input keeps at logic '0' (i.e., PMOS connected to this input is negative biased) with respect to the executed workloads. For example, if the statistical signal probability on a gate input is 0.5 during time $t$, we can say that the time that the gate input keeps at logic '0' is 0.5$t$, and the duty cycle on this gate input is 0.5 correspondingly.

Based on above observations, we identify the worst duty cycles on the inputs of statistically critical gates. The worst delay degradations on these gates can then be obtained using Eq. (9) given the worst duty cycles.

Similar to the statistical signal probability, duty cycles on the gate inputs in the practical circuit operation are restricted by the category of logic gate and the circuit topology. Thereby calculation and propagation of the duty cycles passing through the circuit can be performed in a similar way that the signal
isn't satisfied, the sizing operation will repeat on the next specific lifetime reliability constraint (i.e., the timing constraint) then performed to obtain the new circuit delay distribution. If the each time of sizing operation, the gate delay is re-profiled. SSTA is and sizes PMOSs and NMOSs lying on the statistically critical consideration. It picks the first

Similar with [8–10], the sizing operation in each iteration can be into a candidate group in descendant order based on their size, e.g., the channel length is 65 at 65 nm technology. The ratios paths passing through the gate.

\[
P_g = \frac{1}{C_1} \sum_{i=1}^{n} \Delta D_i
\]

where \( P_g \) denotes sizing sensitivity of the gate. \( \Delta D_i \) denotes the worst delay degradation on ith PMOS given the worst duty cycle on ith input of the gate. \( n \) is the number of statistically critical paths passing through the gate.

### 5.3. Gate sizing procedure

Initially, all of the gates in the circuit are assigned with base size, e.g., the channel length is 65 at 65 nm technology. The ratios of width and length (W/L) for PMOS and NMOS are set to 10 and 5, respectively. All of the statistically critical gates are sorted into a candidate group in descendant order based on their \( P_g \)'s. Similar with [8–10], the sizing operation in each iteration can be performed on multiple gates at the same time for the runtime consideration. It picks the first \( n \) gates from the candidate group and sizes PMOSs and NMOSs lying on the statistically critical paths simultaneously by increasing a unit width on them. After each time of sizing operation, the gate delay is re-profiled. SSTA is then performed to obtain the new circuit delay distribution. If the specific lifetime reliability constraint (i.e., the timing constraint) isn't satisfied, the sizing operation will repeat on the next \( n \) gates in the candidate group. This iterated sizing operation terminates until the lifetime reliability constraint is satisfied or the maximal area constraint is violated.

### 6. Experimental results and analysis

Experiments are performed on some ISCAS benchmark circuits. Only some primitive gates including INV-, 2-4 inputs NAND- and NOR-gate are used in the netlist synthesis process. Parameters \( A, B, \gamma \) and \( \beta \) in the statistical aging model are fitted by HSPICE simulation. The fitting process uses MOSFET Model Reliability Analysis (MOSRA) in HSPICE under different duty cycles (\( \alpha \)), working temperatures (\( T \)) and variation distributions of \( V_{th} \).

Initially, all of the gates in the circuit are assigned with a base size, i.e., channel length of the transistor is set to 65 nm. The ratios of width and length (W/L) for PMOS and NMOS are set to 10 and 5, respectively. HSPICE simulation is performed to obtain the gate delays with respect to different W/L's. All of the HSPICE simulations are performed under PTM 65 nm technology node [20].

In this work, a 3-level quad-tree partition is used to model the spatial correlation between different gates in the circuit. Each gate is randomly allocated a location on the 4 x 4 grid in the bottom level and then, random variables associated with the gate along the hierarchy are determined. The random variables in the same level have same probability distribution. The total standard variance of 10% is assumed from the threshold voltage. It is divided into systematic variance of 6% and random variance of 8%, respectively. For the 3-level quad-tree model, the systematic variance is further divided into 1.9%, 3.9% and 4.2% and assigned to the level 0 to 2. Assigning larger variance to the lower level denotes the strong correlation between adjacent gates. The statistical circuit optimization flow shown in Fig. 1 is implemented by hybrid programming of C++ and MATLAB. All of the experiments are run on a Intel Xeon eight-core linux server, with working frequency of single core 2.33 GHz and total 16 G memory.

### 6.1. Model verification

We firstly verify the accuracy of proposed variation-aware gate-level delay degradation model (Eq. (9)). Gate delay obtained from Eq. (9) is compared with the one obtained from HSPICE Monte Carlo simulation under the same variation distribution and working conditions. Variation on the threshold voltage is assumed as Gaussian distribution with \( \mu \) of -0.365 V and \( \sigma \) of 10%. Duty cycle (\( \alpha \)) is set to 0.5, operational time (\( t \)) is assumed as 10 years. Moreover, in the HSPICE Monte Carlo simulation, working temperature is also assumed as Gaussian distribution with \( \mu \) of 350 K and \( \sigma \) of 5%. Table 2 lists the error between our model and HSPICE Monte Carlo simulation for the primitive gates. Note that for NAND- and NOR-gate, gate delay is actually the maximum propagation delay in all of the input nodes.

As shown in Table 2, comparing with HSPICE Monte Carlo simulation, the maximum error on \( \mu \) of gate delay in our model is less than 3% while the maximum error on \( \sigma \) is less than 4%. The error mainly derives from the parameter fitting process. Nevertheless, from Table 2 we can see that the proposed variation-aware aging model has acceptable accuracy to be used to model the gate delay degradation under PV and aging effects.

### 6.2. Lifetime reliability analysis

Using the proposed statistical aging degradation model, SSTA is performed to estimate mean and variance of the circuit maximum arrival times under PV and aging effects for 5- and 10-year lifetimes. The analysis is performed under the worst duty cycles on the inputs

### Table 1

<table>
<thead>
<tr>
<th>Gate</th>
<th>Duty cycles on inputs</th>
<th>Calculation and propagation rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>INV</td>
<td>( x )</td>
<td>( 1 - x )</td>
</tr>
<tr>
<td>NAND</td>
<td>( x_1, x_2 )</td>
<td>((1 - x_1)(1 - x_2))</td>
</tr>
<tr>
<td>NOR</td>
<td>( x_1, x_2 )</td>
<td>(1 - x_1 \cdot x_2 )</td>
</tr>
</tbody>
</table>
of statistically critical gates. The circuit operation is divided into active mode and standby mode. Ratio of standby period and active period is assumed as 1:9. Correspondingly, working temperatures during standby period and active period are assumed as 385 and 320 K, respectively. As shown in Fig. 3, with the time increasing, mean of the circuit maximum arrival time increases while variance decreasing. It also means that under the joint effect of PV and aging, more and more paths have larger probability to be critical path. This affects the circuit delay distribution and poses larger challenge on the circuit timing optimization.

6.3. Lifetime reliability optimization

Based on the optimization priority \( P_{k} \)’s of statistically critical gates, gate sizing procedure is then performed to optimize the circuit timing. It terminates until the specific lifetime reliability constraint is satisfied. Here the lifetime reliability constraint is defined as the \( \mu + 3\sigma \) of circuit maximum arrival time is less than the designated timing constraint (e.g., 110% of the circuit delay). Moreover, our sizing procedure is also compared with two other sizing methods: \( S + C \) and \( S + D \). The \( S + C \) sizing method was proposed in [8], which sizes gate by jointly considering sensitivity and criticality of the gate. The \( S + D \) sizing method is derived from the traditional sizing method such as that was used in [9–11]. It considers both sizing sensitivity and practical delay degradation of the gate while without taking criticality of the gate into account. For the three kinds of sizing methods, 50 gates are picked out from the candidate gate group (except c17) in each iteration given their optimizing priorities. Area overhead is obtained by calculating the ratio of total transistor widths after and before the sizing procedure for all of the gates in the circuit. Table 3 illustrates the statistical results for the three kinds of sizing methods under 5- and 10-year lifetimes.

In Table 3, column “area” lists the area overhead consumed by the sizing procedure and column “num” lists the number of iterations in the sizing procedure. We use the iteration number as the metric to represent converging speed of the sizing procedure. This is because that the three kinds of sizing methods size the same number of gates (50) in each iteration as well as they use the same statistical timing analysis procedure. Less number of iterations denotes faster convergence of the sizing procedure. As shown in Table 3, under the goal of satisfying the required lifetime reliability constraint, our sizing method converges quickly with minimum area overhead compared with \( S + C \) and \( S + D \) methods. Area overhead for our sizing method is on average 50% less than the ones for \( S + C \) and \( S + D \) (row AVE in Table 3). This demonstrates the effectiveness of the proposed optimization metric and proves that the practical delay degradation information of the gate should be considered during the sizing procedure. For smaller scale circuit such as c17, area overhead is larger due to less number of the statistically critical gates with respect to larger circuit delay degradation.

Noticed that for some circuits, the area overheads for \( S + D \) are less than the ones for \( S + C \). We think the reason is that some critical gates may have larger delay degradations. Thus during the priority sorting, \( S + D \) method will assign a portion of critical gates with high priorities even though it does not consider criticality of the gate. This also reveals that the practical delay degradation information of the gate should be considered during the gate sizing procedure.

By observing Fig. 3, it can be seen that due to parameter \( n \) is much less than 1 (0.16), increase of the circuit delay becomes

### Table 2

Error of proposed variation-aware aging model.

<table>
<thead>
<tr>
<th>Error [%]</th>
<th>INV</th>
<th>NAND (input)</th>
<th>NOR (input)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>( \mu )</td>
<td>0.97</td>
<td>1.58</td>
<td>1.76</td>
</tr>
</tbody>
</table>
| \( \sigma \)| 1.21| 2.03         | 2.22        | 3.03| 2.81         | 3.12        | 3.55

**Fig. 3.** Joint effect of PV and aging on circuit delay distribution.

### Table 3

Statistical results for three kinds of sizing methods.

<table>
<thead>
<tr>
<th>Circuit</th>
<th>Our 5-year</th>
<th>S + C 5-year</th>
<th>S + D 5-year</th>
<th>Our 10-year</th>
<th>S + C 10-year</th>
<th>S + D 10-year</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Num</td>
<td>Area (%)</td>
<td>Num</td>
<td>Area (%)</td>
<td>Num</td>
<td>Area (%)</td>
</tr>
<tr>
<td>c17</td>
<td>2</td>
<td>18.5</td>
<td>3</td>
<td>27.7</td>
<td>2</td>
<td>23</td>
</tr>
<tr>
<td>c380</td>
<td>3</td>
<td>5.2</td>
<td>5</td>
<td>10.4</td>
<td>4</td>
<td>11.3</td>
</tr>
<tr>
<td>c1908</td>
<td>2</td>
<td>3.8</td>
<td>5</td>
<td>7.3</td>
<td>8</td>
<td>17.1</td>
</tr>
<tr>
<td>c2070</td>
<td>2</td>
<td>2.2</td>
<td>4</td>
<td>5.3</td>
<td>4</td>
<td>8.4</td>
</tr>
<tr>
<td>c3540</td>
<td>8</td>
<td>7.9</td>
<td>11</td>
<td>15.9</td>
<td>16</td>
<td>21.8</td>
</tr>
<tr>
<td>c5315</td>
<td>5</td>
<td>5.5</td>
<td>9</td>
<td>11.0</td>
<td>15</td>
<td>14.6</td>
</tr>
<tr>
<td>c6288</td>
<td>7</td>
<td>8.9</td>
<td>12</td>
<td>16.6</td>
<td>13</td>
<td>18.6</td>
</tr>
<tr>
<td>c7552</td>
<td>7</td>
<td>7.7</td>
<td>10</td>
<td>14.2</td>
<td>10</td>
<td>13.8</td>
</tr>
<tr>
<td>c298</td>
<td>2</td>
<td>8.2</td>
<td>5</td>
<td>23.0</td>
<td>4</td>
<td>22.6</td>
</tr>
<tr>
<td>s820</td>
<td>3</td>
<td>6.7</td>
<td>6</td>
<td>17.2</td>
<td>4</td>
<td>14.2</td>
</tr>
<tr>
<td>s1196</td>
<td>5</td>
<td>8.0</td>
<td>9</td>
<td>17.4</td>
<td>7</td>
<td>17.2</td>
</tr>
<tr>
<td>s1238</td>
<td>5</td>
<td>7.4</td>
<td>7</td>
<td>11.8</td>
<td>6</td>
<td>14.1</td>
</tr>
<tr>
<td>AVE</td>
<td>7.5</td>
<td>14.82</td>
<td></td>
<td>16.39</td>
<td></td>
<td>16.21</td>
</tr>
</tbody>
</table>
slower as operational time increasing. This observation is verified by the data in Table 3. The area overhead to achieve 10-year lifetime reliability constraint only increases a little compared to the one for 5-year. Obviously we can ensure the circuit reliability for longer lifetime with only small amount of additional area overhead.

7. Conclusion

This work proposes a statistical circuit optimization flow to optimize circuit timing under the joint effect of PV and aging. A variation-aware gate-level aging degradation model is used to characterize lifetime reliability of the gate. By considering the practical delay degradation information of statistically critical gates, gate sizing procedure converges quickly in term of low area overhead.

Acknowledgments

The work is supported in part by national Basic Research Program of China (973) under Grant no. 2011CB302503; in part by National Natural Science Foundation of China (NSFC) under Grant nos. (60806014, 61076037, 60906018, 60921002, 60831105026).

References


Song Jin is now a Ph.D. candidate in Key Laboratory of Computer System and Architecture, Institute of Computing Technology, Chinese Academy of Sciences. His research interests include circuit reliability design, testing and verification.

Yinhe Han received the B. E. degree from Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2001, and the M. E. and Ph.D. degree in computer science from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, in 2003 and 2006, respectively. He is currently an Associate Professor at the Institute of Computing Technology, Chinese Academy of Sciences. His research interests include VLSI Design and Test, Reliable System and Architecture. Dr. Han was a recipient of the Test Technology Technical Council Best Paper Award at the Asian Test Symposium 2003. He is also a member of ACM/IEEE/IEICE.

Huawei Li received the B.S. degree in computer science from Xiangtan University, Hunan, China, in 1996, and the M.S. and Ph.D. degrees from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, in 1999 and 2001, respectively. She is currently a Professor at the Institute of Computing Technology, Chinese Academy of Sciences. Her research interests include VLSI/SOC design verification, test generation, delay test and design for reliability. She is a senior member of IEEE. She was a Technical Program Chair of IEEE Asian Test Symposium (ATS) in 2007, and Workshop of RTL and High Level Testing (WRTLT) in 2003.

Xiaowei Li received the B.E. and M.E. degrees in computer science from Hefei University of Technology (China) in 1985 and 1988, respectively, and his Ph.D. degree in computer science from the Institute of Computing Technology, Chinese Academy of Sciences in 1991. Dr. Li joined Peking University as a Postdoctoral Research Associate in 1991, and was promoted to Associate Professor in 1993, all with the Department of Computer Science and Technology. From 1997 to 1998, he was a Visiting Research Fellow in the Department of Electrical and Electronic Engineering at the University of Hong Kong. In 1999 and 2000, he was a Visiting Professor in the Graduate School of Information Science, Nara Institute of Science and Technology, Japan. He joined the Institute of Computing Technology, Chinese Academy of Sciences as a professor in 2000. His research interests include VLSI/SOC design verification and test generation, design for testability, low-power design, dependable computing. Dr. Li received the Natural Science Award from the Chinese Academy of Sciences in 1992, the Certificate of Appreciation from IEEE Computer Society in 2001. He is a senior member of IEEE. He is an area editor of the Journal of Computer Science and Technology and an associate editor-in-chief of the Journal of Computer-Aided Design & Computer Graphics (in Chinese).