Temporal Performance Degradation under NBTI: Estimation and Design for Improved Reliability of Nanoscale Circuits

Bipul C Paul[1], Kunhyuk Kang, Haldun Kufluoglu, Muhammad Ashraful Alam, and Kaushik Roy
[1]Toshiba America Research Inc., San Jose, CA 95131, USA
E-mail: bpaul@tari.toshiba.com
School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907, USA
E-mail: {kang18, kufluogl, alam, kaushik}@ecn.purdue.edu

Abstract

Negative Bias Temperature Instability (NBTI) has become one of the major causes for temporal reliability degradation of nanoscale circuits. In this paper, we analyze the temporal delay degradation of logic circuits due to NBTI. We show that knowing the threshold voltage degradation of a single transistor due to NBTI, one can predict the performance degradation of a circuit with a reasonable degree of accuracy. We also propose a sizing algorithm taking NBTI-affected performance degradation into account to ensure the reliability of nanoscale circuits for a given period of time. Experimental results on several benchmark circuits show that with an average of 8.7% increase in area one can ensure reliable performance of circuits for 10 years.

1. Introduction

With the continuous scaling of transistor dimensions, the reliability degradation of circuits has become an important issue in sub-100nm technologies. Due to increasing electric field across the thin oxide, the generation of interface traps under negative bias ($V_{gs} = -V_{dd}$) at elevated temperature (also known as NBTI) in PMOS transistors has become one of the most critical reliability issues that determine the lifetime of CMOS devices [1, 2]. In conventional MOSFETs, due to crystal mismatch at the Si-SiO_2 interface, traps are present in the form of Si dangling bonds after the growth of gate oxide. Traditionally, these interface traps are passivated in ambient hydrogen to improve the device characteristics. However, due to aggressive oxide thickness scaling and process modifications such as the choice of p-poly gates for PMOS and oxide nitridation to prevent boron penetration from the gate, the Si-H bond breaking is accelerated at the interface (degrading the threshold voltage) over time during device operation. This results in temporal performance degradation in nanoscale circuits. Hence, an early estimation of reliability is necessary in the design phase and it should be considered as one of the design parameters to ensure the reliable operation of circuits for a desired period of time.

Due to NBTI the threshold voltage (V_{th}) of the transistor increases with time resulting in the reduction in drive current [3]. The reduction in drive current in turn results in temporal degradation in the performance of a circuit causing reliability degradation over time. Hence, an efficient design technique to minimize this temporal reliability degradation is very essential for digital circuit operation. Considerable eforts have been put in to estimate the threshold voltage degradation [4, 5] as well as to analyze its impact on the drive current and circuit delay [6, 7]. An analytical model is also recently proposed to estimate the temopral perfomance degradation of digital circuits due to NBTI [8]. However, a good design technique taking the temporal reliability degradation as one of the design constraints is rarely available.

In this paper, we analyze the impact of NBTI on the temporal performance degradation of combinatorial circuits and show that by taking this degradation into account during sizing, a desired performance can be ensured with a reasonable area overhead.

We also propose a sizing algorithm to size the gates in a circuit taking the temporal performance degradation into account. This algorithm estimates the threshold voltage degradation of all individual PMOS transistors in a circuit based on their activities and accordingly sizes the circuit for a desired performance. The proposed sizing tool uses Lagrangian relaxation algorithm [9] for global optimization of gate sizes.

We estimate the performance degradation of sev-

3-9810801-0-6/DATE06 © 2006 EDAA

eral ISCAS benchmark circuits implemented in 70nm Berkeley Predictive Technology model (BPTM) [10] and show that the performance degradation of all the circuits has the same power-law dependency as the threshold voltage on time. We also size the above circuits using the proposed algorithm taking the temporal performance degradation into account and compare the area overhead with that of nominal design.

The rest of the paper is organized as follows. Section 2 briefly explains the analytical model for estimating the performance degradation of circuits due to NBTI. We describe the proposed sizing algorithm in section 3 to size circuits considering the temporal performance degradation into account. Section 4 discusses the experimental results on several ISCAS benchmark circuits followed by the conclusion in section 5.

2. Analysis of Temporal Performance Degradation

In this section we will briefly describe the analytical model to estimate the performance degradation of circuits due to NBTI.

2.1. V_{th} degradation

NBTI is the result of trap generation at the Si/SiO_2 interface in negatively biased PMOS transistors at elevated temperatures. The interaction of inversion layer holes with hydrogen-passivated Si atoms can break the Si-H bonds, creating an interface trap and one H atom that can diffuse away from the interface (through the oxide) or can anneal an existing trap. The interface trap generation is modeled successfully in the Reaction-Diffusion (R-D) framework [4]. In this model, interface trap density (ΔN_{IT}) is expressed as,

$$\Delta N_{IT}(t) = \sqrt{\frac{k_f N_0}{k_r}} \left(D_H t\right)^{0.25} \tag{1}$$

where k_f, k_r are the bond-breaking and hydrogen annealing rates, respectively, N_0 is the maximum available Si-H density and D_H is the diffusion coefficient. The bond-breaking rate depends on the accumulation of holes in the inversion layer and the tunneling of the holes into the oxide to dissociate the Si-H bonds [11]. Thus, k_f depends on the hole density, p, hole capture cross-section, σ_0, tunneling coefficient, T_p, and the bond dissociation coefficient, B and can be expressed as $k_f \propto B\sigma_0 p T_p$, where p $(= C_{ox}(V_g - V_{th}) \propto E_{ox})$ and T_p $(\simeq e^{(E_{ox}/E_0)})$ depend on the electric field (E_{ox}) across the oxide. E_0 is the field acceleration factor. B, σ_0 and k_r are assumed to have weak field dependence [11]. Substituting k_f, Eq. (1) can be simplified to

$$\Delta N_{IT}(E_{ox}, t) \equiv \chi \sqrt{E_{ox} e^{\left(\frac{E_{ox}}{E_0}\right)}} t^{0.25} \tag{2}$$

where χ represents the field independent terms. The change in the threshold voltage due to the increase in interface charge, hence, can be given by,

$$\Delta V_{th}(E_{ox}, t) = \frac{q N_{IT}(E_{ox}, t)}{C_{ox}} \tag{3}$$

where q is the electronic charge. Furthermore, the interface traps increase scattering resulting in mobility degradation. The mobility degradation can be expressed as an additional V_{th} shift [7]. The effective threshold voltage degradation can be expressed as,

$$\Delta V_{th}(E_{ox}, t) = (1 + m) \frac{q N_{IT}(E_{ox}, t)}{C_{ox}} \tag{4}$$

where m is a constant representing the equivalent V_{th} shift due to mobility degradation for a given technology.

2.2. Circuit delay degradation

The drain current of a transistor in the saturation region can be approximately represented as,

$$I_d = \beta (V_g - V_{th})^{\alpha}, \quad \beta = \frac{\mu C_{ox} W_{eff}}{L_{eff}} \tag{5}$$

where V_g and V_{th} are the gate voltage and the threshold voltage of the transistor, respectively. μ is the mobility, C_{ox} is the oxide capacitance and L_{eff} and W_{eff} are respectively, the channel length and width of the transistor. α is a constant, whose value ranges from 1 to 2. The delay of a gate can be approximately expressed as [7],

$$\tau = \frac{C_L V_{dd}}{I_d} = \frac{K_1}{(V_g - V_{th})^{\alpha}}, K_1 = \frac{C_L V_{dd}}{\beta} \tag{6}$$

where C_L is the load capacitance and V_{dd} is the supply voltage. Differentiating Eq. (6) with respect to V_{th} we get,

$$\frac{\delta \tau}{\tau} = \frac{\alpha \delta V_{th}}{(V_g - V_{th})} \tag{7}$$

Integrating Eq. (7) we get,

$$\begin{aligned} \int_{\tau_0}^{\tau_0 + \Delta\tau} \frac{d\tau}{\tau} &= \int_{V_{th0}}^{V_{th0} + \Delta V_{th0}} \frac{\alpha d V_{th}}{(V_g - V_{th})} \\ ln\left(1 + \frac{\Delta\tau}{\tau_0}\right) &= -\alpha \cdot ln\left(1 - \frac{\Delta V_{th}}{V_g - V_{th0}}\right) \end{aligned} \tag{8}$$

where V_{th0} is the threshold voltage at any time instant t and τ_0 is the corresponding gate delay. Using Taylor

series expansion on both sides of Eq. (8) and neglecting the higher order terms we get,

$$\frac{\Delta\tau}{\tau_0} = \alpha\frac{\Delta V_{th}}{V_g - V_{th0}} \tag{9}$$

Substituting ΔV_{th} from Eq. (4) in the form of At^n ($n = 0.25$), Eq. (9) can be rewritten as,

$$log\left(\frac{\Delta\tau}{\tau_0}\right) = n \cdot log(t) + log\left(\frac{\alpha A}{V_g - V_{th0}}\right) \tag{10}$$

$$A = (1+m)\frac{q\chi\sqrt{E_{ox}e^{\left(\frac{E_{ox}}{E_0}\right)}}}{C_{ox}}$$

The second term in the right hand side of Eq. (10) is not constant because V_{th} is a function of time. However, due to the logarithmic dependency this term can be treated as constant for a specific range of time (e.g., 10 years [8]). In this period, $log(\Delta\tau/\tau)$ will linearly change with $log(t)$ with the same slope n as the V_{th} degradation. Therefore, by monitoring the threshold voltage degradation, the change in gate delay can be easily estimated with a high degree of accuracy.

The degradation in the delay of a circuit, τ_{ckt} ($\sum_{i=1}^{N}\tau_{g,i}$) can further be estimated as [8],

$$log(\frac{\Delta\tau_{ckt}}{\tau_{ckt}}) = log\left[\frac{\sum_{i=1}^{N}(S_i\cdot t)^n\frac{\alpha A\tau_{g,i}}{V_g - V_{th,i}}}{\sum_{i=1}^{N}\tau_{g,i}}\right] \tag{11}$$

$$= nlog(t) + log\left[\sum_{i=1}^{N}\frac{\alpha A(S_i)^n\tau_{g,i}}{\tau_{ckt}(V_g - V_{th,i})}\right]$$

$\tau_g, \tau_{ckt}: \ (t = 0)$

where N is the number of gates in the critical path and $\tau_{g,i}$ is the delay of ith gate. $S_i \cdot t$ represents the on-time of a PMOS transistor in a particular gate, i, based on its switching activity. Note that S_i is typically calculated over a period of time and shows negligible change. Hence, the slope (n) of $log(\Delta\tau_{ckt}/\tau_{ckt})$ vs. $log(t)$ will not change.

Fig. 1 shows the degradation of threshold voltage and the corresponding gate (inverter) delay with time. The threshold voltage degradation was obtained using Eq. (4). The inverter delay was measured through HSPICE simulation using 70nm BPTM technology[1] ($V_{dd} = 1V$). It can be observed that both the threshold voltage and the gate delay degradation have the same slope ($= 0.25$) as expected from Eq. (10).

It can also be observed from Fig. 1 that the degradation in delay is less than the degradation in threshold voltage. This can be understood from Eq. (9). Since

1 In SPICE simulation we added an appropriate battery to the PMOS gate to replicate the effect of V_{th} change.

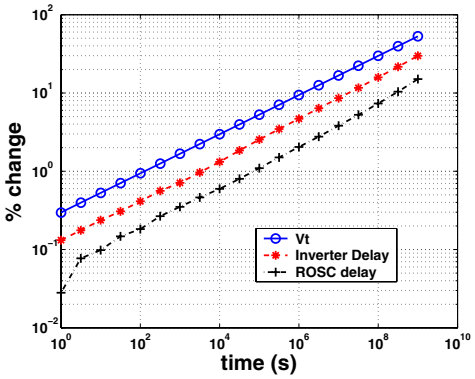

Figure 1. Percentage degradation in inverter and ring oscillator delays with time under NBTI stress. The delays of inverter and the ring oscillator are obtained through HSPICE simulation using BPTM 70nm technology.

$V_g - V_{th0}$ is greater than V_{th0} (within the time range shown in Fig. 1) and α is close to one for short channel transistors, the percentage degradation in delay will be less than that of the threshold voltage and can be quantified as $log[(\alpha/(V_g - V_{th0})]$. For example, with $V_g = 1V$ and $V_{t0} = 0.3V$, a 20% degradation in V_{th} will cause approximately 10% degradation in gate delay (Eq. (9)).

Fig. 1 also shows the performance degradation of a 9-stage ring oscillator with time. The delay was also measured through HSPICE simulation using 70nm BPTM technology with 1V supply. As expected, the circuit performance degradation also has the same slope as V_{th} degradation. However, the percentage degradation is further less than that of a single gate. This is because in a circuit, a low-to-high switching is always followed by a high-to-low switching. While a low-to-high switching is affected due to NBTI induced V_{th} degradation, the high-to-low switching is not affected. Hence, the overall performance degradation is approximately half of that of a single gate.

3. Sizing for reliability improvement

In this section, a gate-sizing algorithm is proposed, which considers the temporal performance degradation of the circuit into account. The algorithm proposed here is based on a technique called Lagrangian relaxation (LR) for solving nonlinear optimization problem. Chen et. al [9] proposed the use of LR for simultaneous sizing of gate and interconnects of a combinational circuit to optimize the total area while maintaining a delay constraint.

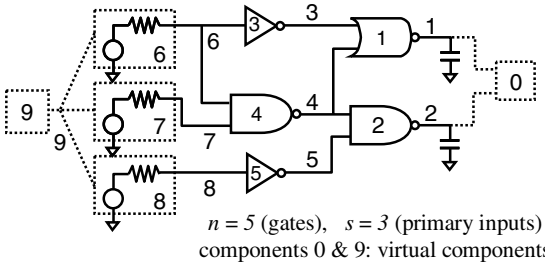

Figure 2. An example circuit for sizing with reliability constraint.

3.1. Problem Formulation

Fig. 2 shows an example circuit representation for sizing with reliability constraint. The circuit consists of n gates, which are to be sized, and s primary inputs. Logic gates and primary inputs are called components. In addition, we add two virtual components, one connecting all primary inputs (component 9 in Fig. 2) and the other connecting primary outputs (component 0). Therefore, for a circuit with n gates and s inputs, there are $n + s + 2$ components. Edge numbers follow their driver gates, i.e., the output of gate i is denoted as edge i. Note that the components and edges are numbered in reverse topological order.

Our objective is to minimize the total area which can be represented by $\sum \alpha_i x_i, i = 1, ..., n$ where x_i is the gate size and α_i is an arbitrary constant multiplier for gate i, which can vary depending on the objective of optimization. Hence, the sizing problem can be formulated as follows.

$$\begin{array}{lll} \text{Minimize} & \sum_{i=1}^{n} \alpha_i x_i & \\ \text{Subject to} & \sum_{i \in p} D_i \leq A_0 & \forall p \in P \\ & L_i \leq x_i \leq U_i & i = 1, ..., n \end{array} \quad (12)$$

where L_i and U_i represent the lower bound and upper bound of the size of gate i, respectively. P is the set of possible paths in a circuit. D_i represents the delay of gate i in a path p. Note that threshold voltages of the transistors in all individual gates will be different due to NBTI depending on their switching activities. The proposed algorithm first calculates the threshold voltage degradation of all transistors in a circuit depending on their switching activities and then sizes the circuit to ensure the performance for a desired period of time (e.g., 10 years).

3.2. V_{th} estimation

This algorithm calculates the switching activities of all gates in a circuit for given signal switching probabilities at all primary inputs [12]. Based on the switching activity, the ON-time ($V_{gs} = -V_{dd}$) of all PMOS transistors are calculated for a desired period of time, T_{max} (e.g., 10 years). Using the ON-time the threshold voltage degradation (ΔV_{th}) of all PMOS transistors are then calculated from Eq. (4).

We calculate the delay of each gate using Sakurai's gate delay model [13]. In [9], Elmore delay model was used for calculating the arrival-time information. However, we use Sakurai's delay model for better accuracy. This model uses precomputed device parameters that are obtained from the I-V characteristics of a particular transistor. However, in our analysis, since the threshold voltage of different transistors are different (hence, the I-V characteristics), we use a look-up table to obtain device parameters corresponding to a particular transistor (threshold voltage).

3.3. Sizing algorithm

In this subsection, we explain how to size the circuit using LR taking the temporal performance degradation into account. Note that the complexity of the problem defined in Eq. (12) is exponentially dependent on the number of components in the circuit ($O(e^n)$). To reduce the complexity to a linear one, the delay constraints on all the paths are transformed into the delay constraints on each gate in the circuit. Therefore, the sizing problem (which is called the primal problem; **PP**) is redefined as,

$$\begin{array}{lll} \text{Minimize} & \sum_{i=1}^{n} \alpha_i x_i & \\ \text{Subject to} & & \\ & a_j \leq A_0 & j \in \text{input(node: 0) /*outputs*/} \\ & a_j + D_i \leq a_i & i = 1, ..., n \quad \forall j \in input(i) \\ & D_i \leq a_i & i = n+1, ..., n+s \quad \text{/*inputs*/} \\ & L_i \leq x_i \leq U_i & i = 1, ..., n. \end{array} \quad (13)$$

a_i represents the signal arrival time at edge i and D_i is the delay associated with gate i considering the temporal performance degradation based on its switching activity. The above problem, **PP** is then solved by Lagrangian relaxation technique [9].

4. Simulation Results

We simulated several ISCAS benchmark circuits to estimate their temporal performance degradation due to NBTI stress and sized them using the proposed algorithm. We considered the following two cases: (1) V_{th} degradation of all PMOS transistors are same (worst case condition) and (2) V_{th} degradation of all individual PMOS transistors are different depending on their

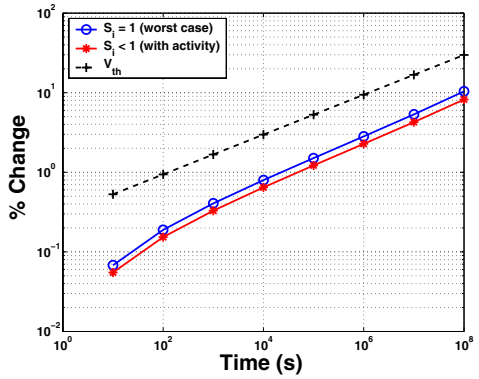

Figure 3. Percentage degradation in the performance of ISCAS C432 benchmark circuit (1) with worst case ($S_i = 1$) NBTI stress and (2) with activity dependent effect.

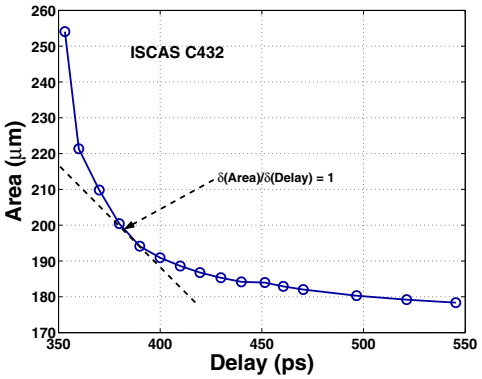

Figure 4. Area vs. delay for ISCAS C432 benchmark circuit. The area was calculated as the sum of transistor width.

on-time ($V_{gs} = -V_{dd}$). The on-time is calculated based on the switching activity of each individual gate. The circuit delay is calculated through static timing analysis. The delay of each gate was calculated using Sakurai's gate delay model [13].

4.1. Performance degradation

Fig. 3 shows the performance degradation of ISCAS C432 benchmark circuit with time. It can be seen that as expected, the slope of $log(\Delta\tau_{ckt})$ vs. $log(t)$ plot both case (1) and (2) are same as that of V_{th} degradation. Also note that the performance degradation considering the switching activity is less than the worst case NBTI stress. The activity factors of all the transistors were calculated by assuming 50% signal switching probability at the primary inputs. Despite a wide variation in ΔV_{th}, the performance degradation is almost comparable to the worst case degradation (Fig. 3). We further calculated the performance degradation of the same circuit (C432) with very low (10%) and high (100%) signal switching probabilities at the primary inputs. It was observed that the performance degradation is a weak function of signal probability. The threshold voltage degradation of PMOS transistor depends on its ON-time, which changes with the signal probability at the input. However, due to the topology of CMOS circuits, when the ON-time of a particular transistor reduces, the ON-time of transistor in the following gate is likely to increase. This makes the overall performance degradation of a circuit insensitive to the signal probability. Hence, calculating the performance degradation with worst case assumption will not result in a significant over estimation. All other benchmark circuits also show the similar delay degradation characteristics with time. Table 1 shows the percentage delay degradation of several ISCAS benchmark circuits with time under both worst case ($S_i = 1$) and activity dependent ($S_i < 1$) NBTI stress. Though the delay degradation depends on the circuit topology, the expected average delay degradation is about 9.2% (worst case) in 10 years, which is approximately four times less than V_{th} degradation of individual transistors.

Circuit	No. of Trans.	Nominal delay (ps)	$\Delta\tau_{ckt}$ (%) (10yr)	
			$S_i = 1$	$S_i < 1$
c432	590	525	8.90	7.32
c499	1816	368	9.20	8.06
c1908	1582	513.5	9.18	8.53
c3540	3638	597.3	9.00	7.86
c74181	372	194.6	9.89	8.68
c74182	92	77.2	10.35	9.63
c74283	188	131.9	7.90	6.83
c74L85	148	115.1	9.50	7.60

Table 1. Delay degradation of ISCAS benchmark circuits under NBTI.

4.2. Sizing

Our proposed sizing algorithm was used to size several ISCAS benchmark circuits taking their temporal performance degradation due to NBTI into account. Fig. 4 shows the area-vs.-delay curve of ISCAS C432 benchmark circuit using conventional LR sizing algorithm. The area (the sum of transistor width) in the plot represents the minimum circuit size for the corresponding delay. Since, the area strongly depends on the

Circuit	Nominal delay (ps)	Nominal area (μm)	% Area overhead	
			$S_i = 1$	$S_i < 1$
c432	385	196.7	14.8	13.6
c499	340	581.47	7.82	6.71
c1908	470	489.67	7.13	6.68
c3540	500	1146.5	3.44	3.31
c74181	180	111.1	9.45	9.0
c74182	80	31.1	11.3	11.2
c74283	125	66.71	10.0	10.0
c74L85	120	42.59	5.85	5.8

Table 2. Sizing ISCAS benchmark circuits considering temporal performance degradation due to NBTI.

delay, the area overhead in sizing considering the reliability into account will also depend on the delay constraint. We chose the delay constraint of a circuit from the above area-vs.-delay plot where the percentage increase in area is equal to the percentage decrease in delay (e.g., 385ps for C432).

Table 2 shows the area overhead in sizing the above benchmark circuits using the proposed algorithm to ensure the reliability for 10 years. We calculated the area overhead for both worst case and activity dependent performance degradation. We chose 50% signal probability for all the benchmark circuits in our analysis. It can be observed from the table that with an average of 8.7% area overhead (worst case) the temporal performance degradation of digital circuits due to NBTI can be avoided. Note that the area overhead in large circuits (e.g., C3540) is even less (3.44%). Hence, though the threshold voltage degradation due to NBTI is estimated to be approximately 35% in 10 years (in 70nm technology), a small area overhead is sufficient to ensure the reliable operation of digital circuits. Also note that the area overhead considering the activity of the circuit is not significantly different from the worst case design. Hence, sizing circuits considering the worst case performance degradation will not result in pesimistic design. This further simplifies the design of digital circuits considering the reliability as one of the design constraints.

5. Conclusions

In this paper, we analyzed the performance degradation of digital circuits due to NBTI stress. It is shown that the overall performance degradation of a circuit is significantly less than the threshold voltage degradation of individual transistors and that it can be handled by employing a sizing technique with an additional reliability constraint.

References

[1] N. Kimizuka, et al, "Impact of bias temperature instability for direct tunneling ultrathin gate oxide on MOSFET scaling," in *Dig. Tech. Papers-Symp. VLSI Technology*, pp. 73-74, 1999.

[2] G. Chen, M. F. Li, C. H. Ang, J. Z. Zheng, and D. L. Kwong, "Dynamic NBTI of p-MOS transistors and its impact on MOSFET scaling," *IEEE Elec. Dev. Lett.*, vol. 23, No. 12, pp. 734-736, Dec. 2002.

[3] D. Schroder, and J. F. Babcock, "Negative bias temperature instability: Road to cross in deep submicron silicon semiconductor manufacturing," *J. Appl. Phys.*, vol. 94, p. 1, 2003.

[4] M. A. Alam, "A critical examination of the mechanics of dynamic NBTI for PMOSFETs," in *Intl. Electron Dev. Meeting*, (IEDM), pp. 346-349, 2003.

[5] S. Zafar, B. H. Lee, and J. Stathis, "Evaluation of NBTI in HfO/sub 2/gate dielectric stacks with tungsten gates," *IEEE Elec. Dev. Lett.*, vol. 25, pp. 153-155, Mar, 2004.

[6] W. Abadeer, and W. Ellis, "Behavior of NBTI under AC dynamic circuit conditions," in *Intl. Reliability Physics Symposium*, (IRPS), pp. 17-22, 2003.

[7] A. T. Krishnan, et al., "NBTI impact on transistor and circuit: models, mechanisms and scaling effects," in *Intl. Electron Dev. Meeting*, pp. 14.5.1-14.5.4, 2003.

[8] B. C. Paul, K. Kang, H. Kufluoglu, M. A. Alam, and K. Roy, "Impact of NBTI on the temporal performance degration of digital circuits," *IEEE Electron Device Letters*, vol. 26, n. 8, pp. 560-562, Aug., 2005.

[9] C. P. Chen, C. C. N. Chu, and D. F. Wong, "Fast and exact simultaneous gate and wire sizing by Lagrangian relaxation," *IEEE Trans. on Computer-Aided Design of Int. Cir. and Syst.*, vol. 18, pp. 1014-1025, 1999.

[10] http://www-device.eecs.berkeley.edu/ ptm.

[11] M. A. Alam, and S. Mahapatra, "A comprehensive model of PMOS NBTI degradation," *Microelectronics Reliability*, vol. 45, pp. 71-81, 2005.

[12] Z. Chen, K. Roy, and T. Chou, "Power sensitivity - A new method to estimate power dissipation considering uncertain specifications of primary inputs", in *Intl. Conf. on Computer Aided Design*, pp. 40-44, 1997.

[13] T. Sakurai, and R. Newton, "Delay analysis of series-connected MOSFET circuits," *IEEE J. of Solid State Circuits*, pp. 122-131, 1991.