# Computational Modeling of Negative Bias Temperature Instability (NBTI) for Reliability-aware VLSI Design

Haldun Kufluoglu, Bipul C. Paul, Kun-Hyuk Kang, Kaushik Roy, Muhammad A. Alam
School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907

*Abstract*—**Modeling of temporal degradation in CMOS circuits due to NBTI and a design methodology to overcome its reliability limits are presented. Device-level NBTI is modeled and calibrated to experimental data. Implications of the model for future-generation devices mentioned in ITRS are discussed. The framework for performance degradation of circuits is established. The tools are then applied on benchmark circuits of different technology generations. Time-dependent parametric variations are obtained for the circuits and a straight-forward sizing algorithm is proposed. The results show that NBTI-reliable circuits can be achieved with small area overhead and this technique is effective for upcoming technology generations.**

## I. Introduction

Negative Bias Temperature Instability (NBTI) poses significant parametric reliability concerns for analog and digital CMOS circuits [1-6]. In PMOSFETs employed by current technology generations, the threshold voltages of the transistors increase over time and with the reduction in the drive current, the circuit performance degrades considerably and decreases the lifetime of applications. In order to maintain a high yield of reliable circuits, understanding the parametric shift at the device-level to estimate the impact of NBTI on circuit behavior is of utmost importance. Further solutions to NBTI would be possible once a solid evaluation technique is established.

In this paper, we develop a computational model to connect the parametric degradation of MOSFETs and circuit-level implications of NBTI. We calibrate this framework with experimental data, and then apply to several benchmark circuits among different generations. After the amount of degradation is extracted, we use a sizing algorithm that considers the switching activity of individual transistors to ensure reliability of circuits for NBTI.

## II. NBTI Theory And Modeling

In CMOS circuits, NBTI affects the PMOSFETs. At the inverter level, when the input signal is *low* (i.e. 0V, see Fig. 1a), the PMOSFET is *ON* ($V_{GS} < 0$V) and NBTI degrades the transistor current. Previously, this condition of the input being *low* was not considered as a threat for circuit reliability since NBTI was negligible in older technologies with thick oxides and buried channel devices and the dominant degradation mechanism was in NMOSFETs due to hot carriers effects. Looking at the operation of the inverter, the PMOSFETs spend approximately 50% of the time under NBTI stress, thus poor NBTI reliability can significantly affect circuit performance.

The degradation due to NBTI stems from the traps created at the Si/oxide interface of PMOSFETs where crystalline Si meets amorphous $SiO_2$. Dangling Si bonds exist at the interface due to the mismatch between these dissimilar materials. During the manufacture, the interface traps are passivated by hydrogen annealing, forming stable Si-H bonds. Hydrogen treatment offered a viable solution to this problem for decades. However, aggressive scaling trends and processing modifications lead to increasing stress on the Si/oxide interface and accelerate the Si-H bond breaking (see Fig. 1b). These unsatisfied, Si-, bonds then shift the threshold voltage of the transistors and increase carrier scattering. The drive current decreases over time and translates into reduced speed at the circuit level.

### A. Reaction-Diffusion (R-D) Model

The interface trap generation is modeled successfully in the Reaction-Diffusion (R-D) framework [7-11]. In this model, the Si-H bond breaking at the interface is represented as the chemical reaction given by,

$$Si - H + h^{+} \rightleftharpoons Si\bullet + H^{0} \ , \quad (1)$$

where $h^{+}$ is a hole in the PMOS inversion layer. The rate of increase in interface trap density ($N_{IT}$) is a competition between bond-breaking and self-annealing processes, i.e.,

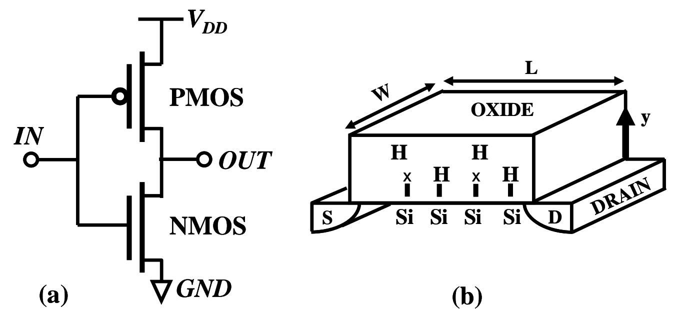


Fig. 1. (a) NBTI occurs when the input of a CMOS inverter is *low*. A PMOSFET is under NBTI stress nearly half of the circuit lifetime assuming equal probability of *low* and *high* input signal. (b) High oxide field combined with the inversion layer holes in a PMOSFET accelerate Si-H bond breaking and generate interface traps (represented by 'x' at the Si/oxide interface).

$$\frac{dN_{IT}}{dt} = k_f\left(N_0 - N_{IT}\right) - k_r N_{IT} N_H^{(0)}, \tag{2}$$

where $k_f$, $k_r$, $N_0$ and $N_H^{(0)}$ are bond-breaking rate, H annealing rate, maximum available Si-H density before stress and hydrogen density at the Si/oxide interface, respectively.

Initially, the interface trap generation is reaction-dominated as shown in region *i* of Fig. 2. After sufficient $N_H^{(0)}$ builds up, the reaction in (2) reaches equilibrium as in region *ii* of Fig. 2. With further build-up at the interface, the hydrogen begins to diffuse into the oxide such that,

$$\frac{dN_H}{dt} = D_H \frac{d^2 N_H}{dy^2}. \tag{3}$$

The diffusion process is slower compared to the chemical reaction and therefore limits the trap generation rate. The hydrogen profile moves as $(D_H t)^{0.5}$ during the diffusion and since every H atom corresponds to one interface trap, i.e.,

$$N_{IT}(t) = \int_0^{\sqrt{D_H t}} N_H(y,t)\,dy = N_H^{(0)} \cdot \sqrt{D_H t}. \tag{4}$$

During the diffusion-dominated regime (*iii* in Fig. 2), $dN_{IT}/dt$ term is negligible compared to the other terms in (2), so

$$k_f N_0 = k_r N_{IT} N_H^{(0)}. \tag{5}$$

From (4) and (5), the time-dependence of the interface trap density is obtained as,

$$N_{IT}(t) = \sqrt{\frac{k_f N_0}{k_r}} \cdot \left(D_H t\right)^{0.25}, \tag{6}$$

which gives the time-behavior observed experimentally. The bond-breaking rate necessitates the accumulation of holes in inversion layer, then tunneling into the oxide and dissociation of the Si-H bonds [2]. Thus $k_f$ includes the dependence on oxide electric field ($\varepsilon_{OX}$) in the hole density $p=C_{OX}(V_G-V_T)$ α $\varepsilon_{OX}$, and in the tunneling terms, $T_p \sim exp(\varepsilon_{ox}/\varepsilon_0)$, as in (7).

$$N_{IT}(\varepsilon_{ox},t) \equiv C \cdot \sqrt{\varepsilon_{ox} e^{\left(\varepsilon_{ox}/\varepsilon_0\right)}} \cdot (t)^{0.25}, \tag{7}$$

where $\varepsilon_0$ is a technology dependent constant and $C$ contains the field-independent components. The interface charge alter the threshold voltage of the PMOSFETs according to (8),

$$\Delta V_T(\varepsilon_{ox},t) = \frac{q\Delta N_{IT}(\varepsilon_{ox},t)}{C_{OX}}, \tag{8}$$

where $q$ is the unit charge and $C_{OX}$ is the oxide capacitance. Additionally, the interface traps increase scattering and reduce the mobility. The mobility degradation is encapsulated as an effective $V_T$ shift in [6,12,13]. Combining these effects, the complete degradation becomes

$$\Delta V_T(\varepsilon_{ox},t) = (m+1) \cdot \frac{q\Delta N_{IT}(\varepsilon_{ox},t)}{C_{OX}}, \tag{9}$$

with $m$ constant for a given technology.

The R-D simulations compare well with experimental NBTI data for DC and AC stress, as shown in Fig. 3. We only consider DC degradation in this paper, however AC condition can be easily accommodated since the degradation in DC and AC conditions are proportional in NBTI [8,9,14].

### B. *Implications for future generation MOSFETs*

The interface trap generation rate in the R-D model strongly depends on the diffusion of hydrogen. For future generation devices considered in ITRS, the geometry of the transistor would directly affect the diffusion process ($\nabla^2$ replaces $d^2/dx^2$ in (3) to include multi-dimensional diffusion) and therefore the NBTI degradation. Numerical R-D simulations that take the device geometry into account are performed for narrow-width, tri-gate and nanowire FETs [10,11,15]. Analytical formulation that removes the need for time-consuming simulations to capture the effect of device geometry on NBTI is also developed [15]. Here, we are only presenting tri-gate results in Fig. 4. As the device width is scaled, the NBTI degradation rate, characterized by $n$ in Fig. 2, increases significantly [10]. This implies that possible advantages of the future generation devices should be re-evaluated in the context of reliability-performance trade-off.

## III. Circuit Analysis

### A. *Gate delay degradation model*

In order to understand the impact of NBTI on circuits, we choose circuit delay as the performance metric and connect

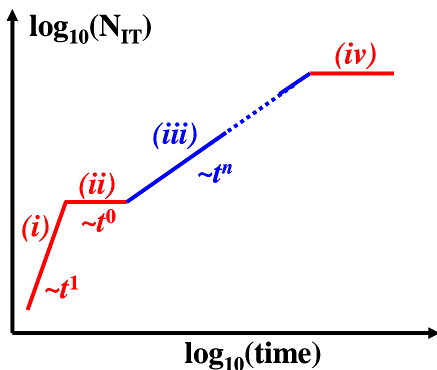


Fig. 2. The time-dependence obtained from the R-D model gives four distinct regions: (i) reaction-dominated regime, governed by (2), (ii) equilibrium stage when $N_H^{(0)}$ builds-up, (iii) diffusion-dominated regime where traditional NBTI time-exponent ($n$~0.25) is observed, and (iv) saturation of trap generation when all the Si-H bonds are broken. Regions *i* and *ii* evolve quickly, therefore are not observed in experiments.

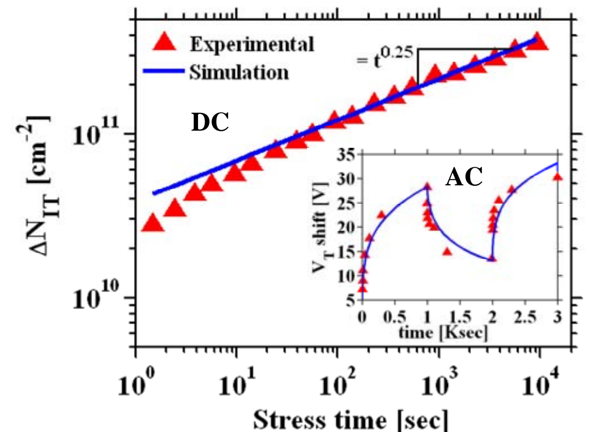


Fig. 3. R-D simulations match the experimental NBTI data well for both DC and AC (inset) bias. Notice the linear scale in the inset. The degradation decreases during AC stress due to partial annealing of the traps by hydrogen (see the annealing term in (2)) when $V_{GS}$=0V. Data from [16,17].

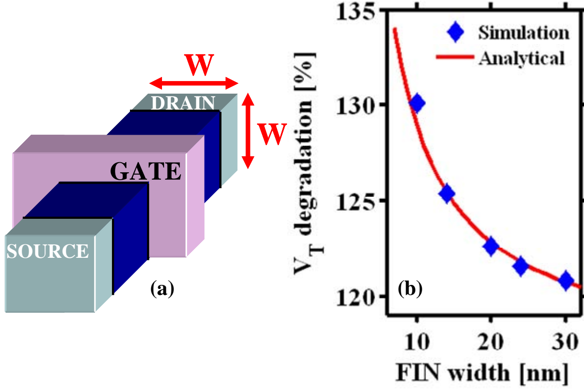


Fig. 4. R-D model is extended to study non-planar transistor geometries such as (a) the tri-gate MOSFET. (b) NBTI degradation rate increases as $W$, an important parameter for scaling, is reduced. Analytical derivations capture the same trend without time-consuming simulations. The model is calibrated to experimental $W$=30nm device from [18].

delay to device-level $\Delta V_T$ from (9). The saturation current of a MOSFET having channel width $W$ and gate length $L$ can be written as

$$I_D = \frac{\mu C_{OX} W}{L}(V_G - V_T)^{\alpha}, \tag{10}$$

where $\alpha$ (~ 1-2) is a constant. When operated at $V_{DD}$ with a load capacitance of $C_L$, the gate delay is

$$\tau = \frac{C_L V_{DD}}{I_D} = \frac{K}{(V_G - V_T)^{\alpha}}. \tag{11}$$

For $V_T$ changing over time, we can obtain the relative shift in gate delay,

$$\frac{\Delta\tau}{\tau} = \frac{\alpha \cdot \Delta V_T}{(V_G - V_T)}. \tag{12}$$

Substituting $\Delta V_T$ from (6) and (8) in the form of $At^n$, the time dependence of delay can be written as

$$\log(\Delta\tau) = n \cdot \log(t) + \log\left(\frac{\alpha\tau A}{V_G - V_T}\right), \tag{13}$$

where second term on the right side of (13) can be treated as a constant over the lifetime of the MOSFET since $(\alpha\tau A) << (V_G - V_T)$. Hence, $\Delta\tau$ follows $\Delta V_T$ with the same time slope in NBTI and by monitoring the degradation in $V_T$, the delay degradation can be estimated accurately.

Another important aspect of the delay degradation is that, percentage-wise, it is always significantly smaller than the change in $V_T$ because of the rightmost term in (13). This implies improved lifetime for the circuits if delay or speed is the criterion, compared to projection from device-level experiments where $\Delta V_T$ limits the reliability.

The performance of degradation of any circuit (based on the activities of individual MOSFETs) can be estimated as follows. The delay of a circuit ($\tau_{ckt}$) is the sum of all individual gate delays ($\tau_{gate}$) in a critical path, and using (13), the performance degradation of a circuit having $N$ gates is

$$\log(\Delta\tau_{ckt}) = N \cdot \log\left(\frac{\alpha\tau A}{V_G - V_T}\right) + \sum_{i=1}^{N} n \cdot \log(S_i \cdot t), \tag{14}$$

$S_i t$ being the *ON*-time of a PMOS (gate $i$) based on its switching activity.

### B. Sizing algorithm

A gate-sizing algorithm based on nonlinear optimization of circuit area while maintaining the delay constraints is used to implement a Lagrangian Relaxation (LR) technique [19,20]. In the context of NBTI, the delay constraints on all the paths in a circuit are transformed into the delay constraints on each gate to reduce the complexity. The optimization problem is translated into a mathematical notation using a Lagrangian multiplier, $\lambda$, for each constraint on arrival time as follows:

$$\begin{aligned} \text{Minimize } L_\lambda(w,a) = & \sum_{i=1}^{n} \alpha_i w_i + \sum_{j \in input(0)}^{n} \lambda_{j0}(a_j - A_0) \\ & + \sum_{i=1}^{n} \sum_{j \in input(i)}^{n} \lambda_{ji}(a_j + D_i - a_i) \\ & + \sum_{i=n+1}^{n+s} \lambda_{mi}(D_i - a_i), \end{aligned} \tag{15}$$

where $\lambda_{ji}$ corresponds to the input edge $j$ and output edge $i$ of gate $i$, $\alpha_i$ represents the signal arrival time at edge $i$, $w_i$ stands for the channel width for gate $i$, $s$ is the number of primary inputs, $m$=$n$+$s$+1, $D_i$ is the delay associated with gate $i$ considering the temporal performance degradation based on its switching activity and $A_0$ is the delay constraint. This formulation optimizes the gate sizes and it is not a path-based algorithm. Additionally, $A_0$ is on the circuit delay and not on any specific path. Optimizing $L_\lambda$ provides the minimum size of a circuit while ensuring the performance for a desired period of time under NBTI stress.

### C. Simulation results

First, the device-level NBTI model is calibrated with experimental data and methodology presented in [2]. Notice that NBTI does not depend on gate length and is primarily affected by oxide thickness for a particular technology. The calibration is projected to the ISCAS benchmark circuits used in the simulations to estimate their temporal performance degradation due to NBTI. We considered two cases: (*i*) $V_T$ degradation of all PMOSFETs are the same (worst case condition), (*ii*) $V_T$ degradation of PMOSFETs are different depending on their *ON*-time ($V_{GS}$=-$V_{DD}$) which is calculated from their switching activity, $S_i$. The activity factors were computed by assuming 50% signal switching probability at the primary inputs. The circuit delay is calculated through conventional static timing analysis [21]. Fig. 5 shows the temporal performance degradation of ISCAS C432 benchmark circuit using 70-nm Berkeley Predictive Technology Model (BPTM) [22]. As expected, the $\Delta\tau_{ckt}$ and $\Delta V_T$ are parallel over time for both cases considered. This behavior was observed experimentally for a ring oscillator in [6]. Additionally, the delay degradation is much less compared to $\Delta V_T$ confirming (13). Furthermore, despite the wide variation in $\Delta V_T$ (not shown), the performance degradation is almost comparable to the worst case degradation. Therefore, estimating performance

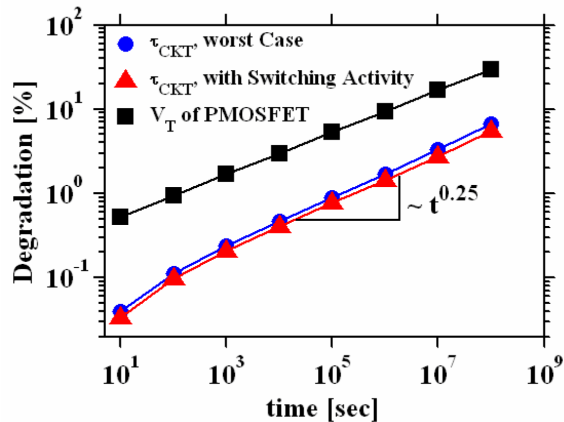


Fig. 5. Worst case and switching-activity dependent degradation for ISCAS C432 benchmark circuit. The delay of this circuit degrades with the same time-exponent. Worst case condition does not exceed the switching-activity dependent shift significantly. $S_i$ is calculated over a period of time and shows negligible change.

degradation with the worst-case assumption will not result in a significant overestimation [23]. We further considered the degradation of the same circuit with low (10%) and high (100%) signal switching probabilities at the inputs [24]. It was observed that the signal probability affects the degradation weakly due to the topology of the CMOS circuitry. When the ON-time of a particular transistor increases, that of another MOSFET likely decreases, so the overall degradation reflects an average behavior. The delay degradation results of several ISCAS benchmark circuits under NBTI stress with worst-case ($S_i$=1) and activity-dependent ($S_i$<1) conditions were also studied. The average delay degradation expected for the circuits is about 9.2% in 10 years (i.e., practical lifetime of the circuit), approximately four times less compared to 35% of $\Delta V_T$. The delay constraint for the sizing algorithm is obtained from the characteristic area-delay relationship. The nominal delay and area for a particular circuit is chosen at the point where the increase in relative area equals the drop in relative delay. The percentage area overhead is calculated using the sizing algorithm to ensure NBTI reliability for 10 years. It was observed that with an average overhead of 8.7%, the temporal degradation of these digital circuits in 70-nm-technology can

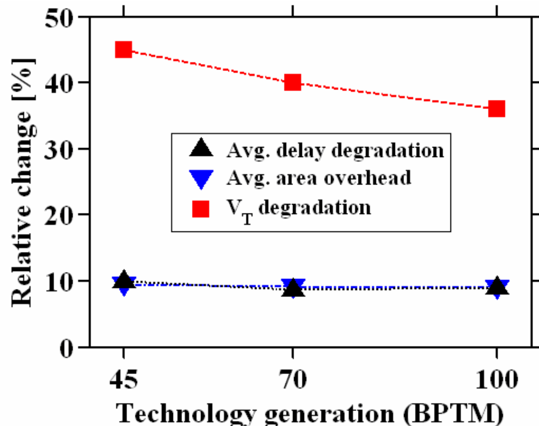


Fig. 6. Degradation and area overhead for the circuits of different BPTM technology generations (i.e., 45, 70 and 100-nm) assuming 10 years of operation. The $\Delta V_T$ increases with scaling, but the delay degradation for fixed circuit area or area overhead for fixed delay do not change appreciably among

be completely avoided. For large circuits, the area overhead is generally lesser. Lastly, we compare the impact of NBTI on different BPTM technology generations as shown in Fig. 6. The $\Delta V_T$ increases with technology scaling, as expected from (13), however the corresponding circuit delay degradation and area overhead are comparable. Although, at the device level, $\Delta V_T$ is predicted to be about 40% in 10 years for these technologies, the circuit delay degradation is much smaller and the desired performance can be maintained with an acceptable increase in the area [25].

## Conclusion

We propose a straight-forward modeling to evaluate the NBTI degradation in circuits. Simulations show that the impact of NBTI on circuit delay is less than the threshold voltage shift at the PMOSFET level. Sizing the circuits at the initial design stage can alleviate the effects of NBTI with small area overhead. The sizing solution seems to be applicable to upcoming technology generations and can ensure NBTI-reliable circuits.

## Acknowledgment

The authors gratefully acknowledge the resources provided by Network for Computational Nanotechnology (NCN) at Purdue University and funding from NSF/SRC contract No. 2004-HJ-1238.

## References


[1] D. Schroder et al., JAP, 94, p. 1, 2003.
[2] M. A. Alam et al., Microelectronics Reliability, 45(1), p. 71, 2005.
[3] V. Reddy et al., IEEE IRPS Proc. p.248, 2002.
[4] W. Abadeer et al., IEEE IRPS Proc. p. 17, 2003.
[5] V. Reddy, et al., IEEE Int'l Test Conf. Proc. p. 148, 2004.
[6] A. T. Krishnan et al., IEEE IEDM Proc. p. 349, 2003.
[7] K. Jeppson et al., JAP, 48, p.2004, 1977.
[8] M. A. Alam, IEEE IEDM Proc. p. 346, 2003.
[9] S. Chakravarthi et al., IEEE IRPS Proc. p. 273, 2004.
[10] H. Kufluoglu et al., IEEE IEDM Proc. p. 113, 2004.
[11] H. Kufluoglu et al., To appear in Journal of Computational Elec., 2005.
[12] S. C. Sun et al., IEEE J. Solid-State Cir., 15(4), p. 562, 1980
[13] J. E. Chung et al., IEEE TED 38(6), p. 1362, 1991.
[14] H. Kufluoglu et al., unpublished results.
[15] H. Kufluoglu et al., IEEE TED, in review, 2005.
[16] V. Huard et al., IEEE IRPS Proc. p. 40, 2004.
[17] G. Chen et al., IEEE IRPS Proc. p. 196, 2003.
[18] S. Maeda et al., IEEE IRPS Proc. p.8, 2004.
[19] C. P. Chen et al., IEEE Tr. Com. Aid. Des. Int. Cir. Sys., p. 1014, 1999.
[20] M. S. Bazaraa et al., Nonlinear Prog.:Theory and Alg., Wiley, 1993.
[21] T. Sakurai et al., IEEE. Journal of Solid State Cir., p. 122, 1991.
[22] BPTM, available at http://www-device.eecs.berkeley.edu/ptm
[23] B. C. Paul et al., IEEE Elec. Dev. Lett. 26(8), p. 560, 2005.
[24] Z. Chen et al., IEEE ICCAD, p. 40, 1997.
[25] M. A. Alam et al., IEEE ICICDT., 2005.