A Novel Gate-level NBTI Delay Degradation Model with Stacking Effect

No Author Given

No Institute Given

Abstract. In this paper, we propose a gate-level NBTI delay degradation model, where the stress voltage variability due to PMOS transistors' stacking effect is considered for the first time. Experimental results show that our gate-level NBTI delay degradation model results in a tighten upper bound for circuit performance analysis. The traditional circuit degradation analysis leads to on average 59.3% overestimation. The pin reordering technique can mitigate on average 6.4% performance degradation in our benchmark circuits.

1 Introduction

As technology scales, accelerated aging effect [1] for nanoscale devices poses as a key challenge for designers to find countermeasures that effectively mitigate the degradation and prolong system's lifetime. Negative bias temperature instability (NBTI), which has deleterious effect on the threshold voltage and the drive current of PMOS transistors, is emerging as one of the major reliability concerns [2].

Due to NBTI effect, the threshold voltage of PMOS transistor is shifted, carrier mobility and drain current are reduced [3], and the performance degradation occurs [4–6]. The NBTI phenomena can be classified as *static NBTI* and *dynamic NBTI*. Static NBTI is under the DC stress condition, and the detailed physical mechanism was described in [7]. The impact of electric and environment parameters (such as electric field across the oxide and temperature) on the interface trap generation was studied in [8, 9]. Dynamic NBTI under the AC stress condition leads to a less severe parameter's shift over long time because of the recovery phenomenon [4, 9–11].

Many analytical NBTI models have been proposed recently. The impact of NBTI on the worst case performance degradation of digital circuits was analyzed in [12]. An analytical model for multi-cycle dynamic NBTI was proposed in [13], where a recursion process was used to evaluate the NBTI effect. A predictive NBTI model was proposed in [14, 15], the effect of various process and design parameters was described. An accurate and fast close-form analytical model was proposed in [16], where temperature-aware NBTI modeling was also considered.

Most of these previous proposed NBTI models may suffer from inaccuracy or high computational complexity, and gate-level NBTI modeling is still in its infancy. In this paper, based on an accurate and fast close-form analytical model [16], we propose a gate-level NBTI delay degradation model considering stacking effect. Our contribution in this paper distinguishes itself in the following aspects:

- A single transistor analytical NBTI model is extended to a novel gate-level model, which for the first time considers the variability of the stress voltage due to stacking effect;
- A novel accurate gate-level delay model for V_{th} degradation is first proposed. A tightened upper bound for circuit performance degradation can be achieved with our new gate-level delay model.

The rest of the paper is organized as follows. In Section 2, we first review previous NBTI models, then our model considering the variability of the stress voltage due to stacking effect is described. In Section 3, the new gate-level delay model is presented based on traditional delay analysis. The simulation results of the ALU benchmark circuits are shown and analyzed in Section 4. Finally, Section 5 concludes the paper.

Note that the simulation results in the following sections are based on a standard cell library constructed using the PTM 90nm bulk CMOS model [17]. $V_{dd} = 1.2$V, $|V_{th}| = 200$mV are set for all the transistors in the circuits. The operation time is set to 3×10^8s (about 10yr).

2 NBTI Model

2.1 Previous NBTI models

A threshold voltage degradation ΔV_{th} is caused by the interface trap generation due to PMOS NBTI effect, which is described by [18]

$$\Delta V_{th} = -(1+m)\frac{q_e N_{it}(t)}{C_{ox}} \tag{1}$$

where m represents equivalent V_{th} shifts due to mobility degradation, q_e is the electronic charge, C_{ox} is the gate oxide capacitance, and $N_{it}(t)$ is the interface trap generation due to PMOS NBTI effect.

The interface trap generation is often described by the reaction-diffusion (R-D) model [19]. An analytical solution exists under the DC stress condition, which is regarded as static NBTI model,

$$N_{it}(t) = 1.16\sqrt{\frac{k_f N_0}{k_r}}(D_H t)^{1/4} \tag{2}$$

where N_0 is the concentration of initial interface defects; k_f is dissociation rate which depends on electric field across the gate oxide, and k_r is constant self-annealing rate; and D_H is the corresponding diffusion coefficient [19].

In the multi-cycle dynamic NBTI model proposed by Kumar et al. [13], the interface trap generation can be evaluated by a recursion formula,

$$N_{it}[(n+p_s)T] = N_{it}^0 \left[p_s + \left(\frac{N_{it}(nT)}{N_{it}^0} \right)^4 \right]^{1/4} \tag{3}$$

where $N_{it}^0 = AT^{1/4}$, and $\beta = \sqrt{\frac{1-p_s}{2}}$; T and p_s are the period and the duty cycle of the stress waveform, respectively. Actually, Eq. (3) describes a tight upper bound of all the relaxation phases.
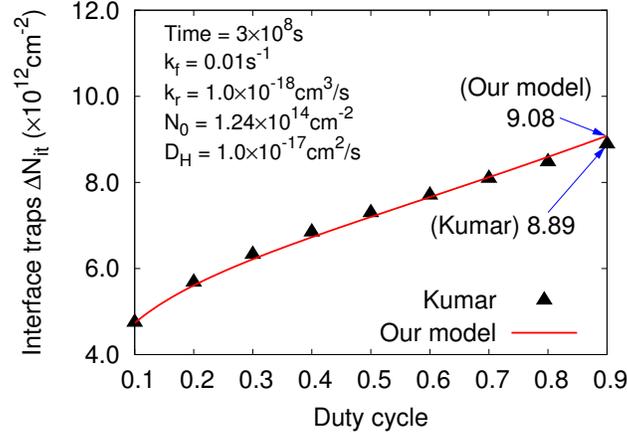
Fig. 1. Comparison between Kumar's and our models

A close-form equation was proposed in [16] using the fitting approach. The model in [16] can describe the dynamic NBTI effect with the same accuracy but faster.

In this paper, we use the same method in [16] to construct our dynamic NBTI model. Eq. (3) in Kumar's model [13] is used as the fitting target function. Hence, the interface trap generation can be described as

$$N_{it}(t) = 1.16 \cdot \xi(p_s) \cdot \sqrt{\frac{k_f N_0}{k_r}} (D_H t)^{1/4} \tag{4}$$

where $\xi(p_s) = p_s^{0.27 p_s + 0.28}$.

The comparison between Kumar's and our model is shown in Fig. 1, and the *Maximum Error* of $N_{it}(t)$ is 2.14% ($9.08 \times 10^{12} \text{cm}^{-2}$ from our model, and $8.89 \times 10^{12} \text{cm}^{-2}$ from Kumar's model in Fig. 1).

2.2 Our novel gate-level NBTI model with stacking effect

Traditionally, the estimation method of V_{th} degradation in a logic gate due to NBTI is as follows: the PMOS transistors and their corresponding inputs are first assumed to be mutually independent; then the V_{th} degradation of each PMOS transistor is analyzed independently based on Eq. (4); finally, the maximum value is chosen to be the V_{th} degradation of the gate and can be used to calculate the delay degradation of the gate. Obviously, the above method is not accurate in the gate-level NBTI analysis, because the stacking effect is not considered. In [14], V_{th} variability due to the body effect in the transistor stack was considered, but only the static NBTI effect was analyzed. In this section, a novel gate-level NBTI model with stacking effect is proposed based on the transistor-level NBTI model given in Section 2.1.

In this paper, the stress voltage variability due to stacking effect in the logic gate is considered. Because of the resistance of the transistors, the internal nodes are biased
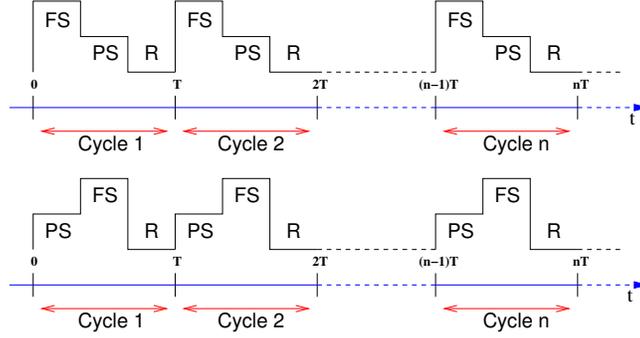
Fig. 2. Stress waveform applied to the gate of the PMOS transistor (R is "relaxation", and FS, PS are "full stress", "partial stress", respectively).

at a middle voltage, which leads to different V_{gs} of the PMOS transistors. Therefore, when the PMOS transistor is under stress, it is not always $-V_{dd}$ biased. We denote the stress condition under $-V_{dd}$ as "full stress (FS)", and the stress under a lower voltage as "partial stress (PS)". Before the new V_{th} degradation model with stacking effect is proposed, the interface trap generation due to dynamic PMOS NBTI effect mixed with "full stress" and "partial stress" should be analyzed.

The random aperiodic signal can be converted to deterministic periodic waveform, based on the signal probability (SP). With the same SP, the NBTI effect will be the same [6]. Hence, we use the waveform shown in Fig. 2 as the input of PMOS transistor. In the first waveform, the "full stress" phase is ahead of the "partial stress" phase in one cycle; and the second waveform shows the reversed condition. From the numeric simulation based on the reaction-diffusion model [19], we find that the order of these phases have negligible impact on the final generation of interface traps. Fig. 3 shows the comparison between the stress waveforms in Fig. 2 and the error is 0.18% ($5.50 \times 10^{12} \text{cm}^{-2}$ vs. $5.51 \times 10^{12} \text{cm}^{-2}$).

In the following part of this paper, we assume that "full stress" is always ahead of "partial stress" in a cycle. Fig. 4 shows the numeric simulated interface trap generation due to dynamic NBTI under different time ratio of "full stress" phase to "partial stress" phase. We find that the mixed effect of these two stress phases can be derived by weighted averaging the "full stress" and "partial stress" effect, which is described as

$$N_{it,mixed} = \frac{p_{FS}}{p_{FS} + p_{PS}} N_{it,FS} + \frac{p_{PS}}{p_{FS} + p_{PS}} N_{it,PS} \tag{5}$$

where $N_{it,FS}$ is the interface trap generation if all stress phases are "full stress"; and $N_{it,PS}$ is the interface trap generation if all stress phases are "partial stress". The parameters p_{FS} and p_{PS} are signal probabilities of "full stress" and "partial stress", respectively. By calculating, we find the maximum error occurs at "30% FS, 30% PS". The simulated trap generation is $5.50 \times 10^{12} \text{cm}^{-2}$ as shown in Fig. 4, and by Eq. (5), the estimated interface trap generation is $5.38 \times 10^{12} \text{cm}^{-2}$. Therefore, the maximum error is 2.18%.

However, in a transistor stack, the PMOS transistors can be biased at various voltages, so there exists more than one "partial stress" condition. Therefore the law de-
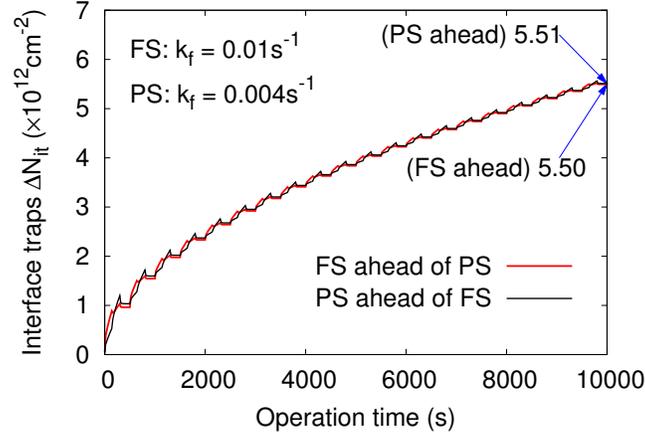
Fig. 3. The impact of the stress phases' order on interface trap generation

scribed above is extended to more than two different stress conditions. First, we number these stress conditions as $S_0, S_1, S_2, \ldots$, and S_0 is always the "full stress" condition. The signal probabilities of these stress conditions are $p_0, p_1, p_2, \ldots$, respectively, and the signal probability of relaxation condition is denoted as r; so the duty cycle p_s of all the stress conditions is

$$p_s = \sum_i p_i = 1 - r \tag{6}$$

where the number of i's is related to the number of PMOS transistors in the stack.

As the threshold voltage degradation is proportional to the interface trap generation, the final V_{th} degradation due to PMOS NBTI effect with more than one stress condition is modeled according to Eq. (5) as

$$\Delta V_{th} = \sum_i \Delta V_{th,i} \frac{p_i}{p_s} \tag{7}$$

where $\Delta V_{th,i}$ is the corresponding threshold voltage degradation if all the stress phases are S_i. According to Eq. (1) and (4), $\Delta V_{th,i}$ is expressed as

$$\Delta V_{th,i} = \eta_i \cdot p_s^{0.27 p_s + 0.28} \cdot t^{1/4} \tag{8}$$

and the parameter η_i is decided by the predictive model proposed in [14],

$$\eta_i = A \cdot T_{ox} \sqrt{C_{ox}(V_{gs,i} - V_{th})} \cdot \exp(\frac{E_{ox,i}}{E_0}) \cdot \exp(\frac{-E_a}{k_b T}) \tag{9}$$

where $V_{gs,i}$ is the stress voltage corresponding to different stress phase due to stacking effect first described in this paper, and other parameters are the same as in [14].

If only one "full stress" condition is considered, that is $p_s = p_0$, Eq. (7) can be simplified as

$$\Delta V_{th} = \Delta V_{th,0} = \eta_0 \cdot p_s^{0.27 p_s + 0.28} \cdot t^{1/4} \tag{10}$$

which consists with the NBTI model in section 2.1.
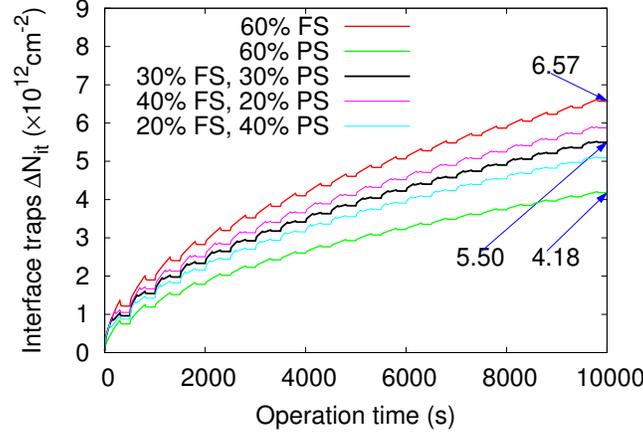
Fig. 4. The analysis of mixed NBTI effect of "full stress" and "partial stress"

3 Gate-level delay degradation analysis

3.1 Traditional gate delay model

Previously, the propagation delay of a gate can be approximately expressed as [18]

$$t_{pd} = \frac{C_L V_{dd}}{I_d} = \frac{C_L V_{dd} L_{eff}}{\mu C_{ox} W_{eff} (V_{gs} - V_{th})^\alpha} \tag{11}$$

where α is the velocity saturation index, and C_L contains the parasitic capacitance. The shift in the transistor threshold voltage ΔV_t can be derived using Eq. (10). Hence, with the Taylor series expansion, the delay degradation Δt_{pd} for the gate is derived as

$$\Delta t_{pd} = \frac{\alpha \Delta V_{th}}{V_{gs} - V_{th}} \cdot t_{pd0} \tag{12}$$

where t_{pd0} is the original delay of the gate without any V_{th} degradation, and can be extracted from third-party time analysis tools.

3.2 Our novel gate-level delay model

The proposed NBTI model with stacking effect described in Section 2.2 leads to different V_{th} degradation for each PMOS transistor in a logic gate, but the gate delay model described in Section 3.1 is incapable to handle this situation. So a novel gate-level delay model is proposed in this paper.

An NOR4 gate is used to illustrate our derivation, and the schematic is shown in Fig. 5, where C_L is the external load capacitance. If the V_{th} degradation of these PMOS transistors are small, the gate delay can be considered linear with ΔV_{th} and C_L,

$$t_{pd} = t_{pd0} + \Delta t_{pd} = t_{pd0} + \sum_i [(\alpha_i \cdot C_L + \beta_i)\Delta V_{th,Mi}], \quad i = 0, 1, 2, 3 \tag{13}$$
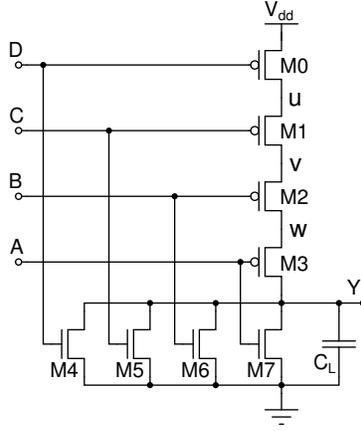
Fig. 5. The schematic of NOR4 gate

Table 1. Threshold voltage degradation in NOR4 gate

Transistor	M0	M1	M2	M3
ΔV_{th}	25.7mV	20.6mV	14.5mV	5.5mV

where the parameters α_i and β_i describe the effect of charging external load capacitance C_L and internal parasitic capacitance respectively, and they only depend on the gate type.

In order to use the existing results extracted from the timing analysis tools directly, the term C_L in Eq. (13) should be eliminated. From Eq. (11), we can derive another linear equation, that the original propagation delay is linear with external load capacitance C_L,

$$t_{pd0} = P \cdot C_L + Q \tag{14}$$

where P is the load delay factor, and Q describes the intrinsic delay. From Eq. (13) and (14), t_{pd} can be derived as

$$t_{pd} = t_{pd0} + \sum_i \left[\left(\frac{t_{pd0} - Q}{P} \alpha_i + \beta_i \right) \Delta V_{th,Mi} \right]$$

$$= t_{pd0} + t_{pd0} \cdot \sum_i (g_i \Delta V_{th,Mi}) + \sum_i (h_i \Delta V_{th,Mi}) \tag{15}$$

$$g_i = \frac{\alpha_i}{P}, \qquad h_i = \left(\beta_i - \frac{Q}{P} \alpha_i \right) \tag{16}$$

where g_i and h_i only depend on the gate type. In the standard-cell design, the parameters g_i and h_i of all the gates in the cell library can be calculated in advance, and then a lookup table is created.

We demonstrate the impact of stress voltage variability due to stacking effect on NBTI analysis. The signal probabilities of all the input patterns are equal. The V_{th} degradation of all the PMOS transistors in NOR4 gate are shown in Table 1. We can see that

Table 2. Comparison between the traditional gate delay analysis and our gate-level delay model

Gate	Original delay	Hspice	Our model		Traditional model	
Type	t_{pd0}	Δt_{pd}	Δt_{pd}	Estimation error	Δt_{pd}	Overestimation
NOR4	168.1ps	6.9ps	7.0ps	1.4%	10.5ps	50.0%
NOR3	142.5ps	5.7ps	5.8ps	1.8%	8.4ps	44.8%
NOR2	111.2ps	4.7ps	4.7ps	0.0%	6.1ps	29.8%
INV	83.5ps	3.6ps	3.6ps	0.0%	3.6ps	0.0%

transistor M0, which is closest to the power supply as shown in Fig. 5, has the largest threshold voltage degradation; while M3 has the smallest threshold voltage degradation. Therefore, the gate-level delay analysis is necessary for accurate estimation of NBTI effect.

The comparison between traditional gate delay analysis and our novel gate-level delay model is shown in Table 2. The third column of Table 2 is the gate delay degradation with stacking effect simulated by Hspice, and the fourth column is calculated by our delay model, while the estimation error of our model is shown in the fifth column. These data demonstrate that our gate-level delay model is accurate enough for delay analysis. If the traditional approach is used to analyze the gate delay degradation, the worst case $\Delta V_{th,M0} = 25.7$mV is set as ΔV_{th} in Eq. (12), and the results are shown in the sixth column. We can see that the traditional gate delay analysis overestimates the delay degradation, and these overestimations compared to our model are shown in the seventh column. We can see that more transistors in PMOS stack lead more overestimation: 29.8% overestimation in NOR2 gate, while 50.0% in NOR4 gate.

From Table 2, we can also see that in the gate with no stacking effect, as INV and AND gate, gate-level delay analysis leads to the same result with traditional analysis. Only the result for INV gate is listed in Table 2 for brevity.

4 Experimental Results

In this section, some ALU circuits and c6288 circuit in ISCAS85 are used as the benchmarks to investigate the effect on the circuit performance degradation using our NBTI delay degradation model. As stacking effect leads to different V_{th} degradation of the PMOS transistors, the gate delay can be minimized using pin reordering technique just as that in leakage minimization [20]. The enumeration searching pin reordering technique is used in our experiment just to estimate the upper bound of our gate-level model in mitigating circuit performance degradation due to NBTI.

The results are shown in Table 3. The circuits array4 and array8 are 4x4 and 8x8 array multipliers; bk16 and bk32 are 16-bit and 32-bit Brent Kung adders; booth9 is 9x9 booth multiplier; ks16 and ks32 are 16-bit and 32-bit Kogge Stone adders; log16 and log32 are 16-bit and 32-bit log shifter; and pm8 and pm16 are 8x8 and 16x16 parallel multipliers. R_{stack} is the ratio of gates with PMOS transistor stack. The original delay t_{pd0} is extracted from an STA tool. The delay degradation with no stacking effect $\Delta t_{pd,ns}$ is evaluated using transistor-level NBTI model Eq. (4) and gate delay model Eq. (12).

Table 3. Delay degradation of benchmark circuits

Circuits	R_{stack}	Original delay t_{pd0} (ns)	No stacking effect $\Delta t_{pd,ns}$ (ps)	Stacking effect $\Delta t_{pd,ws}$ (ps)	Pin reordering $\Delta t_{pd,pr}$ (ps)
array4	61/125	4.11	204.9	157.6	153.4
array8	347/663	4.89	212.8	188.5	179.2
bk16	47/177	2.31	171.7	74.6	67.0
bk32	124/384	3.15	240.1	119.9	77.8
booth9	277/603	2.91	147.3	138.3	134.9
ks16	31/99	2.71	181.0	100.8	99.3
ks32	138/375	3.87	289.1	126.0	120.0
log16	135/160	1.56	79.5	45.2	45.2
log32	268/457	2.20	122.9	60.5	60.5
pm8	278/613	2.71	106.9	95.5	93.1
pm16	1713/3042	4.48	227.9	210.1	205.8
c6288	2128/2447	8.54	564.0	457.0	410.6
Avg.	N/A	N/A	59.3%	0.0%	-6.4%

The delay degradation with stacking effect $\Delta t_{pd,ws}$ is evaluated using our novel gate-level NBTI and delay model Eq. (7) and (15).

In Table 3, we use the fifth column ($\Delta t_{pd,ws}$) as the standard data, which the fourth and sixth columns are compared to. We can see that from Table 3, the traditional method brings on average 59.3% overestimation of the circuit delay degradation. The pin reordering technique leads to on average 6.4% improvement of circuit performance. The overestimation of the circuit delay degradation and the improvement of circuit performance by pin reordering technique depend on not only R_{stack}, but also the contribution of gates with PMOS stack to the critical paths in the circuit. For example, ks32 leads to 129.4% overestimation of delay degradation, much larger than pm16, although R_{stack} of ks32 is less than that of pm16. bk32 and ks32 have almost the same R_{stack}, and the overestimations of delay degradation are both large, but bk32 has a larger improvement of circuit performance by pin reordering. Almost all the gates in c6288 circuit are NOR2, the overestimation of circuit delay degradation is 23.4%, very close to the overestimation of a single NOR2 gate: 29.8%.

5 Conclusion

Negative bias temperature instability is emerging as one of the major circuit performance degradation concerns. Fast and accurate analysis of NBTI-induced circuit degradation is important for circuit designers to find mitigation solutions. In this paper, we use a simple close-form analytical V_{th} degradation model for PMOS to develop a novel gate-level NBTI and delay model. The stress voltage variability due to PMOS transistors' stacking effect is for the first time considered in gate-level NBTI modeling.

The traditional analysis of gate delay degradation due to NBTI results in 50.0% overestimation for an NOR4 gate, while in the circuit performance degradation analysis, the maximum overestimation is 130.2% in 16-bit Brent Kung adder (bk16) circuit. The

mitigation of performance degradation by pin reordering technique can reach up to 35.1% in 32-bit Brent Kung adder (bk32) circuit.

References

1. Borkar, S.: Designing reliable systems from unreliable components: The challenges of transistor variability and degradation. Micro, IEEE **25**(6) (2005) 10–16 0272-1732.
2. Huard, V., Denais, M., Parthasarathy, C.: NBTI degradation: From physical mechanisms to modelling. Microelectron. Reliab. **46**(1) (2006) 1–23
3. Kufluoglu, H., Alam, M.: Theory of interface-trap-induced NBTI degradation for reduced cross section MOSFETs. IEEE Trans. Electron Devices **53**(5) (2006) 1120–1130
4. Wittmann, R., Puchner, H., Hinh, L., Ceric, H., Gehring, A., Selberherr, S.: Impact of NBTI-driven parameter degradation on lifetime of a 90nm p-MOSFET. In: Proc. Intl. Integrated Reliab. Workshop Final Report. (2005) 99–102
5. Reddy, V., Krishnan, A.T., Marshall, A., Rodriguez, J., Natarajan, S., Rost, T., Krishnan, S.: Impact of negative bias temperature instability on digital circuit reliability. Microelectron. Reliab. **45**(1) (2005) 31–38
6. Kumar, S., Kim, C., Sapatnekar, S.: Impact of NBTI on SRAM Read Stability and Design for Reliability. In: Proc. ISQED. (2006) 210–218
7. Ogawa, S., Shiono, N.: Generalized diffusion-reaction model for the low-field charge-buildup instability at the Si-SiO$_2$ interface. Physical Review B **51**(7) (1995) 4218–4230
8. Alam, M., Mahapatra, S.: A comprehensive model of PMOS NBTI degradation. Microelectron. Reliab. **45**(1) (2005) 71–81
9. Mahapatra, S., Saha, D., Varghese, D., Kumar, P.: On the generation and recovery of interface traps in MOSFETs subjected to NBTI, FN, and HCI stress. IEEE Trans. Electron Device **53**(7) (2006) 1583–1592
10. Chen, G., Li, M., Ang, C., Zheng, J., Kwong, D.: Dynamic NBTI of p-MOS transistors and its impact on MOSFET scaling. IEEE Electron Device Lett. **23**(12) (2002) 734–736
11. Mahapatra, S., Bharath Kumar, P., Dalei, T., Sana, D., Alam, M.: Mechanism of negative bias temperature instability in CMOS devices: degradation, recovery and impact of nitrogen. In: IEDM Tech. Dig. (2004) 105–108
12. Paul, B., Kang, K., Kufluoglu, H., Alam, M., Roy, K.: Impact of NBTI on the temporal performance degradation of digital circuits. IEEE Electron Device Lett. **26**(8) (2005) 560–562
13. Kumar, S., Kim, C., Sapatnekar, S.: An Analytical Model for Negative Bias Temperature Instability. In: Proc. IEEE/ACM ICCAD. (2006) 493–496
14. Vattikonda, R., Wang, W., Cao, Y.: Modeling and Minimization of PMOS NBTI Effect for Robust Nanometer Design. In: Proc. DAC. (2006) 1047–1052
15. Bhardwaj, S., Wang, W., Vattikonda, R., Cao, Y., Vrudhula, S.: Predictive Modeling of the NBTI Effect for Reliable Design. In: Proc. CICC. (2006) 189–192
16. Luo, H., Wang, Y., He, K., Luo, R., Yang, H., Xie, Y.: Modeling of PMOS NBTI Effect Considering Temperature Variation. In: Proc. ISQED. (2007)
17. Nanoscale Integration and Modeling Group, ASU: Predictive Technology Model (PTM)
18. Paul, B., Kang, K., Kufluoglu, H., Alam, M., Roy, K.: Temporal Performance Degradation under NBTI: Estimation and Design for Improved Reliability of Nanoscale Circuits. In: Proc. DATE. Volume 1. (2006) 1–6
19. Stathis, J., Zafar, S.: The negative bias temperature instability in MOS devices: A review. Microelectron. Reliab. **46**(2-4) (2006) 270–286
20. Sultania, A., Sylvester, D., Sapatnekar, S.: Transistor and pin reordering for gate oxide leakage reduction in dual Tox circuits. In: Proc. ICCD. (2004) 228–233